

Univerzita Karlova v Praze

1. lékařská fakulta

Studijní program: Biomedicínská informatika



Mgr. Petra Přečková

Jazyk lékařských zpráv a jeho informačně lexikální analýza

The language of medical reports and its information-lexical analysis

Disertační práce

Vedoucí závěrečné práce/Školitel: Prof. RNDr. Jana Zvárová, DrSc.

Praha, 2011

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem řádně uvedla a citovala všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 27. 6. 2011

Petra Přečková

Podpis

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.

Identifikační záznam:

PŘEČKOVÁ, Petra. *Jazyk lékařských zpráv a jeho informačně lexikální analýza.* [The language of medical reports and its information-lexical analysis]. Praha, 2011, 105 s. Disertační práce. Univerzita Karlova v Praze, 1. lékařská fakulta, Oddělení medicínské informatiky ÚI AV ČR, v. v. i. 2011. Vedoucí závěrečné práce/Školitel Zvárová, Jana.

Abstrakt:

Cílem disertační práce byla informačně lexikální analýza českých lékařských zpráv a využitelnost mezinárodních klasifikačních systémů v českém zdravotnickém prostředí. Analýza lékařských zpráv byla založena na attributech 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 metod pro měření diverzity lékařských zpráv psaných v jakémkoli jazyce. Tato nová aplikace je založena na obecných konceptech diverzity 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 vývojářům 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 a pacienty. 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 from 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*

OBSAH

1. ÚVOD	4
2. CÍLE PRÁCE	6
3. MATERIÁL A METODIKA	8
3. 1. Kódovací a klasifikační systémy	8
3. 1. 1. MKN – Mezinárodní klasifikace nemocí a přidružených zdravotních problémů	9
3. 1. 1. 2 Historie	10
3. 1. 1. 3. Obsah a uspořádání MKN-10	11
3. 1. 1. 4. Kategorie MKN-10	12
3. 1. 2. SNOMED	17
3. 1. 3. SNOMED CT	18
3. 1. 3. 1. Využití terminologie SNOMED CT	18
3. 1. 3. 2. Základní komponenty terminologie SNOMED CT	19
Koncepty	19
Popisy (druhy, označení)	21
Druhy popisů	22
3. 1. 4. MeSH	30
3. 1. 5. LOINC	33
3. 1. 6. ICD-O	35
3. 1. 7. TNM	36
3. 1. 8. DSM	36
3. 1. 9. Další klasifikační systémy	36
3. 2. Konverzní nástroje	37

3. 2. 1. Zdroje UMLS	37
Metathesaurus UMLS	38
Semantic Network (Sémantická síť) UMLS	42
SPECIALIST lexicon (ODBORNÝ lexikon) UMLS	43
3. 2. 2. MetamorphoSys ULMS	45
3. 3. Minimální datový model pro kardiologii	47
4. VÝSLEDKY	52
4. 1. Využití klasifikačních systémů pro sdílenou zdravotní péči	52
4. 1. 1. Mapování terminologie a tvorba archetypů	54
4. 1. 2. Standardizace klinického obsahu	55
4. 2. Jazyk lékařských zpráv a využití mezinárodních klasifikačních systémů v minimálním datovém modelu pro kardiologii	58
4. 2. 1. Český jazyk	58
4. 2. 2. Jazykový rozbor textových lékařských zpráv	59
4. 2. 3. Analýza znaků Minimálního datového modelu pro kardiologii v textových lékařských zprávách	62
4. 2. 4. Analýza znaků Minimálního datového modelu pro kardiologii v softwarové aplikaci ADAMEK	69
4. 2. 5. Atributy Minimálního datového modelu pro kardiologii zakódované pomocí SNOMED CT	74
4. 2. 6. Atributy Minimálního datového modelu pro kardiologii zakódované pomocí MKN-10	77
4. 3. Výpočet míry diverzity vybraných atributů a jejich kategorií v textových a strukturovaných lékařských zprávách	81
4. 3. 1. Tradiční míry diverzity	81

4. 3. 2. f-diverzita a relativní f-diverzita	82
4. 3. 3. Míry vzácnosti, vlastní a marginální f-diverzita	84
4. 3. 4. Diverzita vybraných atributů a jejich kategorií v textových a strukturovaných lékařských zprávách	86
5. DISKUSE	93
6. ZÁVĚR	95
LITERATURA	98
PŘÍLOHA	106

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. Chápání přesnějšího vymezení klinické jednotky (příznak, diagnóza) je v řadě oborů u jednotlivých lékařských škol rozdílné i v národním měřítku a mezinárodně uznávané konvence dosud nejsou příliš časté. Větší řád je např. v rámci botaniky a zoologie. V těchto oborech je zákonem autorská priorita, tzn. že pojmenování je platné jen podle autora, jenž popsal druh jako první. To vede k zamezení opakovaného popisu téhož druhu s různým názvem a tím i synonymem.

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. Problematikou kódovacích a klasifikačních systémů se zabývá kapitola 3. 1.

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 nejvhodnějších 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 člancích [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. O tom pojednává detailněji kapitola 3. 3.

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 diverzity 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. Tomuto tématu se věnuje kapitola 4. 3. Výpočet míry diverzity vybraných atributů a jejich kategorií v textových a strukturovaných lékařských zprávách.

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.

Klasifikační systémy (klasifikace) jsou takové kódovací systémy, které jsou založeny na principu vytváření tříd. Třídy tvoří agregované pojmy, které se shodují v alespoň jednom klasifikačním atributu. Třídy klasifikace musí pokrývat úplně vymezenou oblast a nesmí se překrývat. Tvorba klasifikačních systémů a nomenklatur byla motivována především jejich praktickým využitím v evidenci, třídění a statistickém zpracování lékařské informace. Prvotním zájmem bylo evidovat výskyt nemocí a příčiny smrti.

Struktura klasifikačního systému může být rozdílná. *Hierarchická klasifikace* vyjadřuje, že podřízený pojem nebo třída je buď specifikací nebo součástí nadřizeného pojmu. *Monohierarchická klasifikace* má právě jednu nadřizenou třídu pro každou třídu (kromě kořene hierarchie). *Polyhierarchická klasifikace* umožňuje, že daná třída může být podřízená více než jedné třídě. *Multidimenzionální klasifikace* sestávají ze dvou či více dílčích klasifikací. Zde je zapotřebí klasifikační atribut pro každou dílčí klasifikaci, která popisují objekt v různých sémantických dimenzích. Objekt je potom klasifikován v každé dílčí klasifikaci odděleně. Typické dimenze v multidimenzionální klasifikaci jsou etiologie, topografie a patologie.

Nomenklatura je systematicky vytvářená množina autorizovaných termínů (deskriptorů) pro danou oblast. Na rozdíl od klasifikace, cílem nomenklatury není zařadit objekt do kategorie, ale popsat jej jednoznačně a přesně, aby se dal dále zpracovávat (například vybrat pacienty s danou kombinací atributů, zpracovat informaci o pacientovi s upozorněním na kontraindikace).

3. 1. 1. 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.

MKN se musí přizpůsobovat vývoji požadavků současné lékařské vědy, aby mohla poskytovat odpovídající informace. Na druhé straně se však od ní požaduje, aby byla stabilizovaná v dostatečně dlouhém časovém období, aby byla jednotná pro celý svět, protože jen tak může sloužit jako základ pro srovnávání nemocnosti populačních skupin i geneticky

odlišných, žijících v různých podmínkách a poskytovat i informace o dlouhodobých trendech vývoje. Kompromisem mezi těmito protichůdnými požadavky bylo přijetí zásady revizí. V současné době se využívá již Desáté revize.

3. 1. 1. 2 Historie

Jak již bylo zmíněno výše, předchůdcem MKN byl Mezinárodní seznam příčin úmrtí (International List of Causes of Death), který v roce 1893 prosadil francouzský lékař Jacques Bertillon na konferenci Mezinárodního statistického institutu (International Statistical Institute) v Chicagu v USA. Tento statistický systém začalo využívat mnoho států a v roce 1898 ho Americká asociace veřejného zdraví (American Public Health Association) (APHA) doporučila k oficiálnímu používání matrikářům v Kanadě, Mexiku a Spojených státech amerických. Zároveň tato asociace doporučila, aby docházelo k pravidelným revizím vždy po deseti letech.

V roce 1900 svolala francouzská vláda první mezinárodní konferenci, jejímž cílem byla revize Klasifikace příčin úmrtí. V této době se jednalo o jednu, ne příliš objemnou knihu, která byla doplněna abecedním rejstříkem. Další konference byly svolány v roce 1910, 1920, 1929 a 1938. Až do páté revize byly prováděny pouze dílčí změny v obsahu, bez zásadního zásahu do struktury. Po smrti Bertillona v roce 1922 byla ustanovena „Smíšená komise“, která byla složena ze zástupců Mezinárodního statistického institutu a Zdravotní organizace Společnosti národů (Health Organization of the League of Nations), která připravovala podklady a návrhy k jednání konferencí.

V průběhu let vzniklo v jednotlivých zemích mnoho doplňků a rozšíření, z nichž některé rozšiřovaly klasifikaci příčin úmrtí i o klasifikaci nefatálních nemocí, ale do mezinárodní verze nebyly dlouho přijaty. V roce 1938 ale mezinárodní konference přijala

rezoluci, která obsahovala doporučení, aby byly různé národní seznamy v maximální možné míře zapracovány do Mezinárodní klasifikace příčin úmrtí.

V roce 1948 převzala za klasifikaci zodpovědnost Světová zdravotnická organizace a šestou revizí, o níž jednala mezinárodní konference v Paříži, započala přeměna systému v univerzální seznam diagnóz. Název byl změněn na „Manual of International Statistical Classification of Diseases, Injuries and Causes of Death“ v českém překladu „Mezinárodní statistická klasifikace nemocí, úrazů a příčin úmrtí“ (MKN). Klasifikace byla vydána ve dvou dílech a obsahovala už i klasifikaci duševních poruch. Další konference se konaly v letech 1955, 1965 a 1975. Od sedmé revize zaujaly nefatální nemoci v tomto seznamu rovnocenné místo a MKN zahrnuje i kódy dalších okolností, které ovlivňují kontakt se zdravotnickými službami.

V současné době se využívá desátá revize MKN (MKN-10). V České republice je tato klasifikace v platnosti od roku 1994. Ukázalo se ale, že stanovený desetiletý interval mezi revizemi byl příliš krátký. Práce na revizním procesu musely být zahájeny dříve, než byla platná verze MKN používána tak dlouho, aby mohla být důkladně zhodnocena. Potřeba konzultovat s mnoha zeměmi a organizacemi činí tento proces velmi zdlouhavým. Vydání jedenácté revize MKN (MKN-11) je tedy očekáváno až kolem roku 2015.

3. 1. 1. 3. Obsah a uspořádání MKN-10

MKN má podobu číselníku. Ve verzi MKN-9 byly kódy diagnóz trojčíferná čísla. Jednotlivé výseky číselné řady odpovídaly skupinám nemocí a stavů. Rozšířená verze, ICD-9-CM obsahovala navíc E-kódy vyjadřující vnější příčiny úrazů a jejich čísla byla ze stejné části číselné řady jako kódy pro úrazy a V-kódy označovaly další faktory ovlivňující zdravotní stav nebo kontakt se zdravotnickými službami. Tyto kódy odpovídají Z-kódům v MKN-10.

Jádrem klasifikace MKN-10 je třímístný kód, který je povinnou úrovní kódování pro mezinárodní hlášení o úmrtnosti pro databázi Světové zdravotnické organizace a pro všeobecné mezinárodní srovnávání. V MKN-10 je prvním znakem zleva vždy velké písmeno latinské abecedy, které udává hlavní kategorii. Znaky na druhém a třetím místě určují hlavní skupinu diagnóz. Za tečkou na čtvrtém, případně i dalším místě, následuje podrobnější členění. Výsledkem je více než zdvojnásobení kódovacích možností ve srovnání s devátou revizí. Z 26 možných písmen bylo použito 25. Písmeno U bylo ponecháno volné pro doplňky a změny a pro možné prozatímní klasifikace k vyřešení potíží, které mohou vzniknout mezi revizemi. Kódy U00-U49 se mohou používat pro prozatímní přidělení novým nemocem nejisté etiologie. Kódy U50-U99 mohou být použity ve výzkumech, například zkouší-li se možnosti alternativního podtřídění pro zvláštní projekt.

3. 1. 1. 4. Kategorie MKN-10

Mezinárodní klasifikace se člení do těchto kategorií:

- Infekční a parazitární nemoci (A, B),
 - *např. A84.1 – československá encefalitida přenášená klišťaty,*
 - *B17.1– hepatitida typu C,*
- novotvary (C),
 - *C15.5 – zhoubný novotvar dolní třetiny jícnu,*
- novotvary, nemoci krve a imunity (D),
 - *D52.1 – anémie z nedostatku kyseliny listové, vyvolaná léky,*
- nemoci endokrinní a metabolické (E),
 - *E66.1 – obezita způsobená léky,*
- nemoci duševní a poruchy chování (F),
 - *F20.0 – paranoidní schizofrenie,*

- nemoci nervové soustavy (G),
 - *G47.1 – poruchy nadměrné spavosti,*
- nemoci oka a očních adnex, nemoci ucha (H),
 - *H11.2 – jizvy spojivky,*
- nemoci oběhové soustavy (I),
 - *I13.0 – hypertenzní nemoc srdce a ledvin s (městnavým) selháním srdce,*
- nemoci dýchací soustavy (J),
 - *J37.0 – chronická laryngitida,*
- nemoci trávicí soustavy (K),
 - *K70.4 – alkoholická cirhóza jater,*
- nemoci kůže a podkožního vaziva (L),
 - *L70.0 – acne vulgaris,*
- nemoci svalové a kosterní soustavy (M),
 - *M24.2 – poruchy vazů,*
- nemoci močové a pohlavní soustavy (N),
 - *N21.1 – kámen v močové trubici,*
- těhotenství, porod, šestinedělí, perinatální stavy, vrozené vady, deformace (O, P, Q),
 - *O30.2 – těhotenství čtyřčtetné,*
 - *P05.1 – malý plod vzhledem k délce těhotenství,*
 - *Q12.0 – vrozený základ,*
- příznaky, znaky a nálezy nezařazené jinde (R),
 - *R78.2 – nález kokainu v krvi*

- poranění, otravy, následky působení vnějších příčin (S, T),
 - *S42.0 – zlomenina klíční kosti,*
 - *T18.2 – cizí těleso v žaludku,*
- zevní příčiny nemocí a úmrtí (V, W, X, Y),
 - *V86.0 – řidič zcela terénního nebo jiného mimosilničního motorového vozidla, zraněný při provozní (silniční) nehodě,*
 - *X34 – pád ze skály - W15; oběť zemětřesení,*
 - *Y06.1 – zanedbání a opuštění rodičem,*
- faktory ovlivňující zdravotní stav (Z),
 - *Z54.2 – rekonvalescence po chemoterapii.*

ICD-10: - Mozilla Firefox

http://apps.who.int/classifications/apps/icd/icd10online/

World Health Organization ICD Version 2007

List of Chapters
ICD-10 Homepage

International Statistical Classification of Diseases and Related Health Problems 10th Revision Version for 2007

Tabular List of inclusions and four-character subcategories

Chapter List

Chapter	Blocks	Title
I	A00-B99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00-E90	Endocrine, nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	H60-H95	Diseases of the ear and mastoid process
IX	I00-I99	Diseases of the circulatory system
X	J00-J99	Diseases of the respiratory system
XI	K00-K93	Diseases of the digestive system
XII	L00-L99	Diseases of the skin and subcutaneous tissue
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIV	N00-N99	Diseases of the genitourinary system
XV	O00-O99	Pregnancy, childbirth and the puerperium
XVI	P00-P96	Certain conditions originating in the perinatal period
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes
XX	V01-Y98	External causes of morbidity and mortality
XXI	Z00-Z99	Factors influencing health status and contact with health services
XXII	U00-U99	Codes for special purposes

Chapter I:

**Certain infectious and parasitic diseases
(A00-B99)**

[A00-A09](#) Intestinal infectious diseases
[A15-A19](#) Tuberculosis
[A20-A28](#) Certain zoonotic bacterial diseases
[A30-A49](#) Other bacterial diseases
[A50-A64](#) Infections with a predominantly sexual mode of transmission
[A65-A69](#) Other spirochaetal diseases

Search ICD-10
Full search
OK Help
Move to ICD code: OK

Application prepared by: WHO & DIMDI (German Institute of Medical Documentation and Information)

Obr. 1. Ukázka z MKN 10.

ICD-10: - Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje nápověda

ICD-10: http://apps.who.int/classifications/apps/icd/icd10online/

World Health Organization ICD Version 2007

Chapter IX
Diseases of the circulatory system (I00-I99)
Diseases of arteries, arterioles and capillaries (I70-I79)

I70 Atherosclerosis
Includes: arteriolosclerosis
arteriosclerosis
arteriosclerotic vascular disease
atheroma
degeneration:
- arterial
- arteriovascular
- vascular
endarteritis deformans or obliterans
senile:
- arteritis
- endarteritis
Excludes: cerebral (I67.2)
coronary (I25.1)
mesenteric (K55.1)
pulmonary (I27.0)

I70.0 Atherosclerosis of aorta
I70.1 Atherosclerosis of renal artery
Goldblatt's kidney
Excludes: atherosclerosis of renal arterioles (I12.-)
I70.2 Atherosclerosis of arteries of extremities
Atherosclerotic gangrene
Mönckeberg's (medial) sclerosis
I70.8 Atherosclerosis of other arteries
I70.9 Generalized and unspecified atherosclerosis

I71 Aortic aneurysm and dissection
I71.0 Dissection of aorta [any part]
Dissecting aneurysm of aorta (ruptured) [any part]
I71.1 Thoracic aortic aneurysm, ruptured
I71.2 Thoracic aortic aneurysm, without mention of rupture
I71.3 Abdominal aortic aneurysm, ruptured
I71.4 Abdominal aortic aneurysm, without mention of rupture

Search ICD-10
atherosclerosis
Full search
OK Help

Move to ICD code:
 OK

Application prepared by: WHO & DIMDI (German Institute of Medical Documentation and Information)

Obr. 2. Ateroskleróza v MNK-10.

3. 1. 2. SNOMED

Akronym SNOMED vznikl ze spojení **S**ystematized **N**omenclature of **M**edicine. SNOMED byl prvně publikován v roce 1965. Jedná se o detailní klinickou referenční terminologii založenou na kódování. Skládá se z pojmů vztahujících se ke zdravotnictví a umožňuje využívat zdravotnické informace kdykoli a kdekoli je to potřeba. SNOMED poskytuje „společný jazyk“, který umožňuje konzistentní způsob získávání, sdílení a shromažďování zdravotnických dat od různých klinických skupin, mezi které patří ošetřovatelství, medicína, laboratoře, lékárny i veterinární medicína. Tento klasifikační systém je používán ve více než 40 státech. SNOMED umožňuje popis jakékoli situace v medicíně pomocí 11 úrovní – dimenzí:

1. Topografie;
2. Morfologie;
3. Funkce;
4. Živé organismy;
5. Fyzičtí činitelé, aktivity a síly;
6. Chemikálie, léky a biologické produkty;
7. Procedury;
8. Zaměstnání;
9. Sociální kontext;
10. Nemoci/diagnózy;
11. Modifikátory.

Jednotlivé pojmy jsou označovány zkratkou dimenze a pětímístným zvláštním kódem, kde je využíváno čísel 0-9 a navíc písmen A-F. Jednotlivá místa kódu směrem doprava stále zpřesňují obsah popisovaného pojmu.

3. 1. 3. 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 **S**ystematized **N**omenclature of **M**edicine **R**eference **T**erminology, 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.

3. 1. 3. 1. Využití terminologie SNOMED CT

Zdravotnické softwarové aplikace se zaměřují na sběr klinických dat, na propojení klinických znalostních databází, získávání informací a také na shromažďování a výměnu dat. Informace ale mohou být zaznamenány různými způsoby v různou dobu a na různých místech.

Standardizované informace zlepšují analýzu. SNOMED CT poskytuje standard pro klinické informace. Softwarové aplikace mohou využívat koncepty, hierarchie a vztahy jako společný referenční bod pro analýzu dat. SNOMED CT slouží jako základ, na kterém mohou zdravotnické organizace vyvinout efektivní aplikace, aby mohly provádět výzkum ze závěrů, hodnotit kvalitu péče a náklady na ní a aby mohly navrhnout efektivní lékařská doporučení pro léčbu.

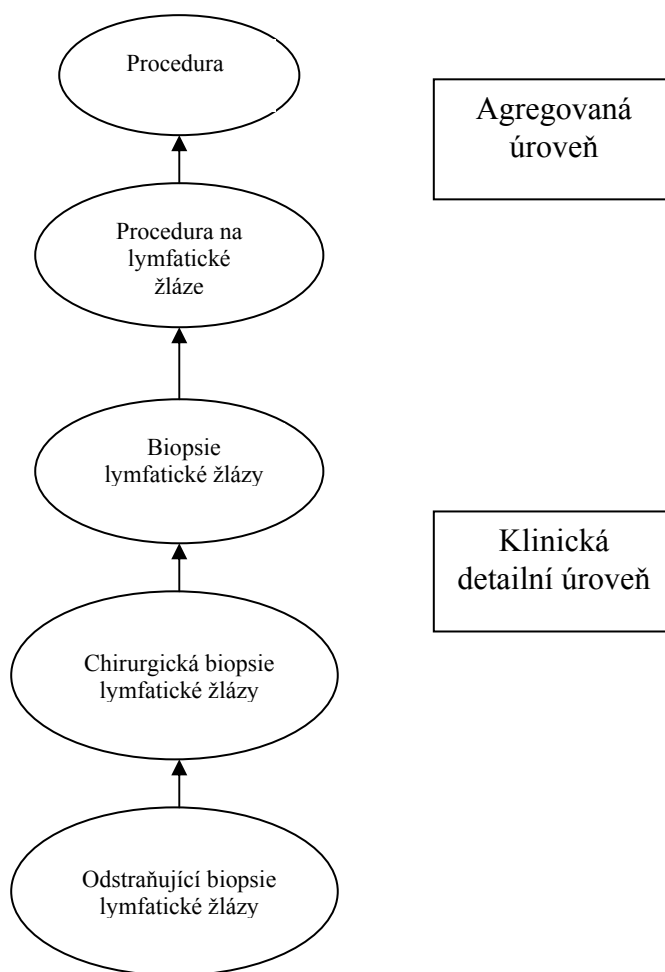
Standardizovaná terminologie může přinést výhody lékařům, pacientům, administrátorům, softwarovým vývojářům a plátcům. Klinická terminologie může pomoci poskytovatelům lékařské péče tak, že jim poskytne snáze dostupné a kompletní informace, které náleží k procesu zdravotnické péče (chorobopis pacienta, nemoci, léčby, laboratorní výsledky, atd.) a proto vyústí v lepší výsledky v péči o pacienta. Klinická terminologie může umožnit poskytovateli lékařské péče identifikovat pacienty podle zakódované informace v jejich záznamech a tím usnadnit další vyšetřování a léčbu.

3. 1. 3. 2. Základní komponenty terminologie SNOMED CT

Koncepty

V rámci SNOMED CT znamená „koncept“ klinický význam, který je identifikován jedinečným numerickým identifikátorem (ConceptID), který se nikdy nemění. Koncepty jsou reprezentovány jedinečným, pro člověka čitelným „Úplně specifickým názvem“ (Fully Specified Name) (FSN). Koncepty jsou formálně definovány ve vztazích k dalším konceptům. Tyto „logické definice“ poskytují explicitní význam, který může počítač zpracovat a dotazovat se na něj. Každý koncept má také skupinu termínů, které pojmenovávají koncept způsobem čitelným pro člověka.

Koncepty představují různé stupně klinického detailu. Koncepty mohou být velice obecné nebo mohou představovat zvyšující se specifické úrovně detailu, kterým se také říká zvyšující se granularita. Zvyšující úrovně granularity zlepšují schopnost kódovat klinická data v náležité úrovni detailu.



Obr. 3. Úrovně granularity.

Koncepty ve SNOMED CT mají jedinečné numerické identifikátory, které se nazývají ConceptID. ConceptID neobsahuje hierarchické nebo implicitní významy. Numerický identifikátor neukazuje žádnou informaci o povaze konceptu.

Příklad:

367416001 je ConceptID pro koncept *angina pectoris (disorder)*.

Popisy (druhy, označení)

Popisy konceptu (concept descriptions) jsou termíny nebo názvy, které jsou přiděleny konceptu ve SNOMED CT. „Termín“ v tomto kontextu znamená frázi, která je použita k pojmenování konceptu. Jedinečné DescriptionID identifikuje popis. Násobné popisy mohou být spojeny s konceptem, který je identifikován svým ConceptID.

Příklad:

Několik popisů spojených s ConceptID 22298006:

- Úplně specifický název: *Myocardial infarction (disorder)*

DescriptionID 751689013

- Preferovaný termín: Myocardial infarction

DescriptionID 37436014

- Synonymum: Cardiac infarction

DescriptionID 37442013

- Synonymum: Heart attack

DescriptionID 37443015

- Synonymum: Infarction of heart

DescriptionID 37441018

Každý z výše zmíněných popisů má jedinečné DescriptionID a všechny tyto popisy jsou spojeny s jedním konceptem (a jedním ConceptID 22298006).

Druhy popisů

Fully Specified Name (FSN) (Úplně specifický název)

Každý koncept má jeden jedinečný FSN, který má poskytnout jednoznačný způsob, jak pojmenovat koncept. Účelem FSN je jednoznačně identifikovat koncept a objasnit jeho význam. Neznamená to nutně, že představuje nejčastěji používanou nebo přirozenou frázi konceptu. Každý FSN je ukončen „sémantickým přívlastkem“, který je v závorce na konci konceptu. „Sémantický přívlastek“ označuje sémantickou kategorii, do které koncept patří (např. Disorder (choroba), Organism (organismus), Person (osoba), atd.). Například *Hematom (morfologická abnormalita)* je FSN, které představuje popis toho, co patologové vidí na úrovni tkáně, zatímco *Hematom (choroba)* je FSN, který označuje koncept, který by použili praktičtí lékaři pro kódování klinické diagnózy hematomu.

Preferred Term (Preferovaný termín)

Každý koncept má jeden preferovaný název, který zachycuje obvyklé slovo nebo frázi, kterou pojmenovávají koncept kliničtí lékaři. Například koncept *54987000 Repair of common bile duct (procedure) (obnova žlučovodu (procedura))* má preferovaný termín *Choledochoplasty (plastika žlučovodu)*, který představuje obvyklý název, který používají kliničtí lékaři k popisu této procedury.

Na rozdíl od FSN nemusí být preferované termíny jedinečné. Občas se může stát, že preferovaný termín pro jeden koncept může být synonymem nebo preferovaným termínem pro jiný koncept.

Příklad:

Cold sensation quality (qualifier value) (druh pocitu nachlazení (hodnota kvalifikátoru)) má preferovaný termín *Cold (nachlazení)*.

Common cold (disorder) (běžné nachlazení (choroba)) má synonymum *Cold (nachlazení)*.

V obou případech představuje *cold (nachlazení)* obvyklou klinickou frází, která se používá k zachycení významu FSN.

Synonyma

Synonyma představují další doplňkové termíny, které představují stejný koncept jako FSN. Synonyma, stejně jako preferované termíny, nemusí být jedinečné.

Příklad:

Některá synonyma, která jsou spojena s ConceptID 22298006, který má FNS *Myocardial infarction (disorder)* jsou:

- Synonymum: Cardiac infarction
DescriptionID 37442013
- Synonymum: Heart attack
DescriptionID 37443015
- Synonymum: Infarction of heart
DescriptionID 37441018

Vztahy

Koncepty ve SNOMED CT jsou propojovány pomocí vztahů. Existují čtyři druhy vztahů, které mohou být ve SNOMED CT přiřazeny konceptům:

- defining (definující),
- qualifying (vymezující),
- historical (historické),
- additional (doplňkové).

Každý koncept ve SNOMED CT je logicky definovaný svými vztahy k jiným konceptům.

Hierarchie

Koncepty SNOMED CT jsou organizovány do hierarchií. Koncept klasifikace SNOMED CT je „Root concept“ (kořenový koncept). Koncept zahrnuje koncept nejvyšší úrovně (supertyp) a všechny koncepty pod ním (jeho subtypy). Protože jsou hierarchie klesající, tak koncepty uvnitř nich se stávají více a více specifickými (nebo-li granulovanými).

„Subtypy“ (nebo-li „potomci“) jsou potomci „supertypu“ (nebo-li „rodiče“).

Příklad:

Streptococcal arthritis (disorder) (streptokoková artritida (choroba)) je subtypem konceptu *Bacterial arthritis (disorder)* (bakteriální artritida (choroba)).

„Supertypy“ jsou předky „subtypu“.

Příklad:

Bacterial arthritis (disorder) je supertyp *Streptococcal arthritis (disorder)*.

Mezi nejvyšší hierarchie patří:

- clinical finding (klinický nález),
- procedure (procedura),
- observable entity (pozorovatelná entita),
- body structure (struktura těla),
- organism (organismus),
- substance (substance),
- pharmaceutical/biologic produkt (farmaceutický/biologický produkt),
- specimen (vzorek),
- special concept (speciální koncept),
- physical object (fyzický předmět),
- physical force (fyzikální síla),
- event (událost),
- environments/geographical locations (prostředí/geografická místa),
- social context (sociální kontext),
- situation with explicit context (situace s explicitním kontextem),
- staging and scales (fáze a měřítka),
- linkage concept (spojovací koncept),
- qualifier value (hodnota kvalifikátoru) a
- record artifact (artefakt záznamu).

Hierarchie ***Klinický nález*** obsahuje sub-hierarchii *Disease* (nemoc). Koncepty, které jsou potomci *Disease* (nebo disorders (choroby, zdravotní potíže)), jsou vždy abnormální klinické stavy.

Koncepty **Procedura** představují aktivity, které jsou prováděny při péči o zdraví. Tato hierarchie představuje širokou škálu aktivit, včetně, ale ne pouze, invazních procedur (*odstranění nitrolebeční tepny (procedura)*), podávání léků (*očkování proti černému kašli (procedura)*), zobrazovací procedury (*ultrasonografie prsu (procedura)*), vzdělávací procedury (*osvěta o dietě s nízkým obsahem soli (procedura)*) a administrativní procedury (*přenos lékařských záznamů (procedura)*).

Situace s explicitním kontextem byla až do července 2006 nazývaná kategorií závislou na kontextu. Tato hierarchie byla potom přejmenována, aby lépe popsala význam konceptů v této hierarchii. Koncepty v hierarchii *Procedura* a *Klinické nálezy* mohou v klinickém záznamu představovat podmínky a procedury, které ještě neproběhly (např. *plánovaná endoskopie (situace)*); podmínky a procedury, které se vztahují k někomu jinému než k pacientovi (např. *rodinná anamnéza: diabetes mellitus (situace)*) nebo podmínky a procedury, které se objevily v jiné době než v přítomnosti (např. *záznamy o dřívější splenektomii (situace)*). Ve všech těchto případech je klinický kontext upřesněný. Druhý příklad, ve kterém je důraz konceptu kladen na jinou osobu než na pacienta, může být vyjádřen ve zdravotním záznamu kombinací záznamu v „rodinné anamnéze“ s hodnotou „diabetes“. Specifický kontext (v tomto případě rodinná anamnéza) by byl vyjádřen strukturou záznamu. V tomto případě kontextově závislý koncept *Rodinná anamnéza: diabetes mellitus (situace)* by se nepoužil, protože informační model už aspekt diabetu mellitu v rodinné anamnéze zachytil.

Na koncepty v hierarchii **Pozorovatelná entita** můžeme pomýšlet jako na ty, které zastupují otázku nebo proceduru, které mohou podat odpověď nebo výsledek. Například *levý ventrikulární koncový diastolický tlak (pozorovatelná entita)* by mohl být interpretován jako otázka: „Co je to levý ventrikulární koncový diastolický tlak?“ nebo „Co je to měřený levý ventrikulární koncový diastolický tlak?“. Pozorovatelné veličiny jsou elementy, které mohou

být použity k zakódování elementů na kontrolním seznamu nebo jakýkoli element, kterému může být přidělena hodnota. *Barva nehtu (pozorovatelná entita)* je pozorovatelná veličina. *Šedé nehty (nález)* je nález. Jedno z využití *pozorovatelných entit* v klinickém záznamu jsou záhlaví v šabloně. *Pohlaví (pozorovatelná entita)* může být využito k zakódování sekce šablony „pohlaví“, kde by si uživatel vybral „muž“ nebo „žena“. „Ženský rod“ by potom znamenal nález.

Koncepty **Struktura těla** zahrnují normální i abnormální anatomické struktury. Normální anatomické struktury mohou být použity ke specifikaci místa na těle, které se týká nemoci nebo procedury, např. *struktura mitrální chlopně (struktura těla)*. Morfologické změny normálních struktur těla jsou vyjádřeny sub-hierarchií *Struktura těla, změněná od své původní anatomické struktury (morfologická abnormalita)*. Příklad může být *polyp (morfologická abnormalita)*.

Hierarchie **Organismus** zahrnuje důležité organismy v lidské a zvířecí medicíně. Organismy se ve SNOMED CT používají také při modelování příčin nemocí. Jsou důležité ve veřejném zdravotnictví pro podmínky podléhající ohlašovací povinnosti a pro protokoly o nakažlivých nemocech v klinických systémech podpory rozhodování. Sub-hierarchie organismu zahrnují například *zvíře (organismus)*, *mikroorganismus (organismus)*, *rostlina (organismus)*. Příklad konceptu Organismus je *lišejník (rostlina) (organismus)*.

Hierarchie **Substance** zahrnuje koncepty, které se používají pro zaznamenávání aktivních chemických složek léků, potravin a chemických alergenů, nepříznivých účinků, toxicity nebo informací o otravě a pokynů lékařů a sester. Koncepty z této hierarchie představují obecné „substance“ a chemické složky *Farmaceutického/biologického produktu (produkt)*, který je v separátní hierarchii. Nicméně, sub-hierarchie *Substance* také zahrnují například *substanci těla (substance)* (koncepty, které vyjadřují substance těla); *dietní substanci (substance)* a *diagnostickou substanci (substance)*. Příkladem je *insulin (substance)*.

Hierarchie **Farmaceutický/biologický produkt** stojí odděleně od hierarchie *Substance*. Tato hierarchie má jasně rozlišovat léčiva (produkty) od jejich chemických složek (substance). Například *Diazepam (produkt)*.

Hierarchie **Vzorek** zahrnuje koncepty, které představují entity, které jsou získány (většinou od pacienta) během vyšetření nebo analýzy. *Vzorky* mohou být definovány atributy, které specifikují: normální nebo abnormální struktura těla, ze které jsou získány; procedura, která se používá ke sběru vzorků; zdroj, ze kterého byly sebrány a substance, ze které se skládají. Příkladem je *vzorek z prostaty získaný jehlovou biopsií (vzorek)*.

Koncepty v hierarchii **Fyzický předmět** zahrnují přírodní a umělé předměty. Jedním z použití těchto konceptů je modelování procedur, které používají různá zařízení (např. katetrizace). Příkladem konceptu v této hierarchii je *filtr duté žíly (fyzický předmět)*.

Koncepty v hierarchii **Fyzická síla** jsou zaměřeny zejména na vyjádření fyzických sil, které mohou hrát roli jako mechanismus zranění, například *střídavý proud (fyzická síla)*.

Hierarchie **Událost** zahrnuje koncepty, které zastupují výskyty (vyjma procedur a zásahů). Příkladem těchto konceptů je *bioterroristický útok (událost)* nebo *zemětřesení (událost)*.

Hierarchie **Prostředí a geografická místa** obsahuje různé druhy prostředí a také názvy míst jako jsou země, státy a regiony, například *Kanárské ostrovy (geografické místo)*, *rehabilitační oddělení (prostředí)* nebo *jednotka intenzivní péče (prostředí)*.

Hierarchie **Sociální kontext** obsahuje sociální podmínky a okolnosti, které jsou důležité pro zdravotnictví. Patří sem rodinný stav, ekonomický stav, etnické a náboženské dědictví, životní styl a povolání. Tyto koncepty představují sociální aspekty, které ovlivňují zdraví a léčbu pacienta. Mezi sub-hierarchie *Sociálního kontextu* patří: etnická skupina, povolání, osoba, náboženství/filosofie a ekonomický status.

Hierarchie **Fáze a měřítka** je rozdělena na sub-hierarchie jako jsou *hodnotící škála* a *fáze nádoru*.

Hierarchie **Spojovací koncept** obsahuje koncepty, které se používají pro vazby. Dělí se na sub-hierarchie *uplatnění vztahu* a *atribut*. Sub-hierarchie uplatnění vztahu umožňuje použití konceptů klasifikace SNOMED CT ve výkazech HL7, které prokazují vztahy mezi výkazy. Příkladem konceptu uplatnění vztahu je *má vysvětlení (uplatnění vztahu)*. Koncepty, které se odvozují od této sub-hierarchie jsou používány ke stavbě vztahů mezi dvěma koncepty klasifikace SNOMED CT, jelikož ukazují druh vztahu mezi těmito koncepty. Některé atributy mohou být použity k logické definici konceptu (definující atributy). Tato sub-hierarchie také zahrnuje nedefinující atributy (jako ty, které se používají ke sledování historických vztahů mezi koncepty) nebo atributy, které mohou být užitečné k modelování definic konceptů, ale které nebyly ještě použity v modelování dřívějších konceptů ve SNOMED CT.

Hierarchie **Hodnota kvalifikátoru** zahrnuje některé koncepty, které se používají jako hodnoty pro atributy SNOMED CT, které nejsou zahrnuty nikde jinde ve SNOMED CT. Nicméně tyto hodnoty pro atributy nejsou omezeny pouze na tuto hierarchii a mohou být nalezeny i v jiné hierarchii. Příkladem konceptu této hierarchie je *levý (hodnota kvalifikátoru)* nebo *jednostranný (hodnota kvalifikátoru)*.

Jednou ze sub-hierarchií **Speciálního konceptu** je *Nečinný koncept*, který je supertypem pro všechny koncepty, které byly ukončeny a ukazují na aktivní koncept v terminologii.

Artefakt záznamu je entita, která je vytvořena osobou nebo osobami, aby poskytla dalším lidem informace o událostech a stavech různých záležitostí. Většinou je záznam nezávislý na svých jednotlivých fyzických doložených příkladech a skládá se z jednotlivých částí informací (většinou slov, slovních spojení a vět, ale také z čísel, grafů a další elementů

informací). *Artefakty záznamu* nemusí být kompletní zprávy nebo kompletní záznamy. Mohou být částí větších *artefaktů záznamu*. Například celkový zdravotní záznam je *artefakt záznamu*, který také může obsahovat další *artefakty záznamu* ve formě jednotlivých dokumentů nebo zpráv, které na druhou stranu mohou obsahovat jemněji granulované *artefakty záznamů* jako jsou sekce nebo dokonce záhlaví sekcí.

3 . 1. 4. MeSH

Medical Subject Headings [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.

MeSH Browser - Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje nápověda

Medical Subject Headings - Home Page MeSH Browser

http://www.nlm.nih.gov/mesh/MBrowser.html

NATIONAL LIBRARY OF MEDICINE **MEDICAL SUBJECT HEADINGS** **MeSH**

MeSH Home | Contact NLM | Site Index | Search Our Web Site | NLM Home

Health Information | Library Services | Research Programs | New & Noteworthy | General Information

MeSH Browser (2011 MeSH):
The files are updated every week on Sunday.
[Go to 2010 MeSH](#)

Enter term or the beginning of any root fragments: or

Search for these record types:

Main Headings

Qualifiers

Supplementary Concepts

All of the Above

Search as MeSH Unique ID

Search as text words in Annotation & Scope Note

Search in these fields of chemicals:


Heading Mapped To (HM) (Supplementary List)

Indexing Information (II) (Supplementary List)

Pharmacological Action (PA)

CAS Registry/EC Number (RN)

Related CAS Registry Number (RR)

 MeSH vocabulary suggestions

[About MeSH Browser](#) | [MeSH Home Page](#) | [Questions or Comments](#)

[NLM Classification, the scheme used to categorize and organize books, audiovisuals, and similar materials.](#)

U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894
National Institutes of Health
Department of Health & Human Services
[Copyright and Privacy Policy](#)
Last updated: 26 August 2010

Obr. 4. Ukázka z MeSH.

Atherosclerosis - Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje Nápověda

Medical Subject Headings - Home Page x Atherosclerosis x +

http://www.nlm.nih.gov/cgi/mesh/2011/MB_cgi

National Library of Medicine - Medical Subject Headings

2011 MeSH

MeSH Descriptor Data

[Return to Entry Page](#)

Standard View [Go to Concept View](#), [Go to Expanded Concept View](#)

MeSH Heading	Atherosclerosis
Tree Number	C14.907.137.126.307
Scope Note	A thickening and loss of elasticity of the walls of ARTERIES that occurs with formation of ATHEROSCLEROTIC PLAQUES within the ARTERIAL INTIMA
Entry Term	Atherogenesis
See Also	Chylomicron Remnants
See Also	Plaque, Atherosclerotic
Allowable Qualifiers	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH UR US VE VI
Previous Indexing	Arteriosclerosis (1965-2005)
History Note	2006; use ARTERIOSCLEROSIS 1963-2005
Date of Entry	20050630
Unique ID	D050197

MeSH Tree Structures

[Cardiovascular Diseases \[C14\]](#)

[Vascular Diseases \[C14.907\]](#)

[Arterial Occlusive Diseases \[C14.907.137\]](#)

[Arteriosclerosis \[C14.907.137.126\]](#)

[Arteriolosclerosis \[C14.907.137.126.056\]](#)

[Arteriosclerosis Obliterans \[C14.907.137.126.114\]](#)

▶ [Atherosclerosis \[C14.907.137.126.307\]](#)

[Peripheral Arterial Disease \[C14.907.137.126.307.500\]](#)

[Coronary Artery Disease \[C14.907.137.126.339\]](#)

[Intracranial Arteriosclerosis \[C14.907.137.126.372\] +](#)

Obr. 5. Ateroskleróza v MeSH.

3. 1. 5. LOINC

Klasifikační systém **Logical Observations Identifiers Names and Codes (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.

Při mapování lokálních kódů různých testů na kódy LOINC napomáhá mapovací program Regenstrief LOINC Mapping Assistant (RELMA™).

search LOINC - Mozilla Firefox


Soubor Úpravy Zobrazení Historie Záložky Nástroje nápověda

search LOINC

http://search.loinc.org/search.zul;jsessionid=FE0FEDA587C886CE88804273721C0E25

Search LOINC

Options Help



atherosclerosis Search

LOINC	Component	Property	System	Timing	Scale	Method	Class	Type	Status	Short name
58334-4	Multi-Ethnic Study of Atherosclerosis (MESA) classification of underlying cause of death	Type	^Patient	Pt	Nom		PHENX	unknown	TRIAL	MESA class of death

Search generated 1 hits in 0.004 secs.

Copyright © 2010 Regenstrief Institute Inc.

Obr. 6. Ukázka z LOINC.

3. 1. 6. 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 diferenciaci. Dimenze jsou určeny pro třídění morfologických typů nádorů. V současné době existuje její třetí verze.

The screenshot shows a PDF document titled "ICD-0-3 SEER SITE/HISTOLOGY VALIDATION LIST December 4, 2009" on page 50. The document lists various cancer types for the stomach (STOMACH) with their corresponding ICD-O codes and descriptions. The codes are organized into categories such as NEOPLASM, CARCINOMA, NOS, CARCINOMA, UNDIFF., NOS, GIANT & SPINDLE CELL CARCINOMA, SMALL CELL CARCINOMA, NOS, PAPILLARY CARCINOMA, NOS, and SQUAMOUS CELL CARCINOMA, NOS.

Category	Code	Description
NEOPLASM	800	Neoplasm, malignant
	8001/3	Tumor cells, malignant
	8002/3	Malignant tumor, small cell type
	8003/3	Malignant tumor, giant cell type
	8004/3	Malignant tumor, spindle cell type
CARCINOMA, NOS	801	Carcinoma in situ, NOS
	8010/3	Carcinoma, NOS
	8011/3	Epithelioma, malignant
	8012/3	Large cell carcinoma, NOS
	8013/3	Large cell neuroendocrine carcinoma
CARCINOMA, UNDIFF., NOS	802	Carcinoma, undifferentiated type, NOS
	8021/3	Carcinoma, anaplastic type, NOS
	8022/3	Pleomorphic carcinoma
GIANT & SPINDLE CELL CARCINOMA	803	Giant cell and spindle cell carcinoma
	8031/3	Giant cell carcinoma
	8032/3	Spindle cell carcinoma
	8033/3	Pseudosarcomatous carcinoma
	8034/3	Polygonal cell carcinoma
SMALL CELL CARCINOMA, NOS	804	Small cell carcinoma, NOS
	8043/3	Small cell carcinoma, fusiform cell
PAPILLARY CARCINOMA, NOS	805	Papillary carcinoma in situ
	8050/3	Papillary carcinoma, NOS
	8051/3	Verrucous carcinoma, NOS
	8052/3	Papillary squamous cell carcinoma, non-invasive
SQUAMOUS CELL CARCINOMA, NOS	807	Squamous cell carcinoma in situ, NOS
	8070/3	Squamous cell carcinoma, NOS
	8071/3	Sq. cell carcinoma, keratinizing, NOS
	8072/3	Sq. cell carcinoma, lg. cell, non-ker.

Obr. 7. Ukázka z ICD-O.

3. 1. 7. 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.

3. 1. 8. 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. 1. 9. Další klasifikační systémy

V současné době existuje v medicíně více než 100 různých klasifikačních systémů. Mezi ně patří i *AI/RHEUM*; *Alternative Billing Concepts (ABC)*; *Alcohol and Other Drug Thesaurus (AOD)*; *Beth Israel Vocabulary*; *Canonical Clinical Problem Statement System (CCPSS)*; *Current Dental Terminology (CDT)*; *Medical Entities Dictionary (MED)*; *Current Procedural Terminology (CPT)*; *International Classification of Primary Care (ICPC)*; *McMaster University Epidemiology Terms*; *CRISP Thesaurus*; *Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART)*; *Diseases Database*; *DXplain*; *Gene Ontology (GO)*; *Healthcare Common Procedure Coding System (HCPCS)*; *Home Health Care Classification (HHCC)*; *Health Level Seven Vocabulary (HL7)*; *Master Drug Data Base (MDDB)*; *Medical Dictionary for Regulatory Activities Terminology (MedDRA)*; *Multum MediSource Lexicon (MMSL)*; *NANDA nursing diagnoses*; *NCBI Taxonomy* a mnoho dalších.

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.

3. 2. 1. Zdroje 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.

Metathesaurus UMLS

Metathesaurus je velice rozsáhlá, víceúčelová a mnohajazyčná databáze slovní zásoby, která obsahuje informace o biomedicínských konceptech a konceptech vztahujících se ke zdraví. Obsahuje jejich různé názvy a vztahy mezi nimi. Metathesaurus byl vytvořen systémovými vývojáři a je sestaven z mnoha elektronických verzí mnoha různých tezurů, klasifikací, skupin kódů a seznamů řízených termínů, které se používají v péči o pacienta, v účtování zdravotnických služeb, ve statistikách veřejného zdraví, při indexování a katalogizování biomedicínské literatury a k základnímu a klinickému výzkumu a výzkumu zdravotnických služeb. Odkazuje se na ně jako na „zdrojové slovníky“ Metathesauru. Termín Metathesaurus vychází z třetí definice předpony „Meta“ ve Webster’s dictionary, tj. „komplexnější, přesahující“. V tomto smyslu přesahuje Metathesaurus specifické tezaury, slovníky a klasifikace, které obsahuje.

Metathesaurus odráží a zachovává významy, názvy konceptů a vztahy ze svých zdrojových slovníků. Jestliže dva různé zdrojové slovníky používají stejný název pro různé koncepty, zobrazuje Metathesaurus oba dva významy a ukazuje, v kterém zdrojovém slovníku je daný výraz obsažen. Jestliže se stejný koncept objevuje v různých hierarchických kontextech v různých zdrojových slovnících, jsou v Metathesauru zahrnuty všechny hierarchie. Jestliže se objeví v různých zdrojových slovnících odporující si vztahy mezi dvěma koncepty, jsou v Metathesauru zahrnuty oba pohledy.

Jinými slovy, Metathesaurus nepředstavuje komplexní ontologii biomedicíny vytvořenou Národní lékařskou knihovnou nebo jeden stálý pohled na svět (až na vysokou úroveň sémantických druhů přiřazených všem svým konceptům). Metathesaurus zachovává mnoho pohledů na svět, které jsou obsaženy v jeho zdrojových slovnících, protože tyto různé pohledy mohou být užitečné pro různé úkoly.

Obsah Metathesauru je určen kombinací obsahu jeho zdrojových slovníků. Mnoho vztahů (zejména synonyma), atributy konceptů a některé názvy konceptů doplňuje Národní lékařská knihovna během vytváření a údržby Metathesauru, ale v podstatě všechny koncepty pocházejí z jednoho nebo více zdrojových slovníků. Až na pár výjimek platí, že jestliže žádný ze zdrojových slovníků neobsahuje nějaký koncept, neobjeví se tento koncept ani v Metathesauru.

Jelikož se jedná o víceúčelový zdroj, který zahrnuje koncepty a termíny z mnoha různých zdrojových slovníků, které byly vytvořeny pro různé účely, musí být Metathesaurus upraven, aby mohl být efektivně používán ve většině specifických aplikacích. Rozhodnutí, co zahrnout do přizpůsobené podmnožiny Metathesauru, bude mít důležitý dopad na jeho přínos v uživatelském systému. Zdrojové slovníky, které jsou pro některé účely nezbytné, např. LOINC pro standardní výměnu laboratorních dat, mohou být v neprospěch jiným účelům, jako je například zpracování přirozeného jazyka. Vyloučení podmnožiny názvů konceptů, která se nachází ve zdrojovém slovníku a která je jinak užitečná, může být důležité, např. nestandardní zkratky nebo zkrácené formy, kterým chybí oprávněnost nebo vyvolávají nepravé výsledky při zpracování přirozeného jazyka.

Metathesaurus poskytuje informace, které mohou počítačové programy využít k vytvoření standardních dat, k interpretaci uživatelských dotazů, k interakci s uživateli k vylepšení jejich otázek a ke konverzi uživatelských termínů na slovní zásobu, která se používá v příslušných informačních zdrojích. Metathesaurus se používá ve velké škále aplikací, mezi něž patří:

- propojování mezi různými klinickými nebo biomedicínskými slovníky;
- získávání informací z databází a z informačních zdrojů volného textu;
- propojování záznamů o pacientech se souvisejícími informacemi v bibliografických, volnotextových nebo faktografických databázích;

- zpracování přirozeného jazyka;
- výzkum automatické indexace a strukturované vkládání dat.

V mnoha případech je užitečnost Metathesauru zvýšena, jestliže je používán společně s ODBORNÝM lexikonem, lexikálními programy a Sémantickou sítí UMLS. K získání souvislých, srovnatelných výsledků při vytváření datových aplikací, jako je vkládání dat o pacientech, je nezbytné definovat, které koncepty a výrazy Metathesauru mohou být obsaženy ve vytvářených záznamech. Toho lze dosáhnout výběrem jednoho z mnoha zdrojových slovníků Metathesauru, které poskytují nejvhodnější koncepty a výrazy pro vytvářená specifická data. Další koncepty a výrazy Metathesauru potom poskytnou synonyma a související výrazy, které mohou navést uživatele na slovníky, které jsou vybrány pro vytváření aplikace specifických dat.

Metathesaurus (a další produkty UMLS) jsou volně k dispozici uživatelům nejen v USA, ale i v zahraničí. Uživatel musí vyplnit online webovou Licenční smlouvu pro používání Metathesauru UMLS. Uživatelé licence jsou zodpovědní za dodržování omezení týkajících se používání obsahu Metathesauru UMLS, která jsou v licenci podrobně popsána. Přestože většina obsahu Metathesauru může být používána s minimálními omezeními, některá využití některých zdrojových slovníků Metathesauru vyžadují samostatné smlouvy s jednotlivými výrobci slovníků, které mohou zahrnovat poplatky.

Metathesaurus UMLS je pro držitele licence volně ke stažení, případně mohu využít webového či aplikačního programového rozhraní (API) z UMLS Knowledge Source Serveru. Metathesaurus UMLS je držitelům licence na vyžádání k dispozici také na DVD.

UMLS Terminology Browser -- Metathesaurus -- Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje nápověda

UMLS Terminology Browser -- Metathesaurus

https://uts.nlm.nih.gov/metathesaurus.html#C0004153;0;1;CUI;2010AB;EXACT_MATCH;SNOMEDCT;

A service of the U.S. National Library of Medicine | National Institutes of Health

My Profile | Sign Out | Contact

Welcome back, preckova

UMLS Terminology Services
Metathesaurus Browser

UTS Home Applications SNOMED CT Resources Downloads Documentation UMLS Home

Search Tree Recent Searches

Term CUI Code

atherosclerosis Go

Release: 2010AB

Search Type: Word

Source: SNOMEDCT

SPN
SRC
TKMT
ULT

Search Results (24)

C0004153 Atherosclerosis
C0002725 Cerebral Atherosclerosis
C0010054 Coronary Atherosclerosis
C0017327 Generalized atherosclerosis
C0155733 Atherosclerosis of aorta
C0155734 Atherosclerosis of renal artery
C0455407 FH. Atherosclerosis
C0577631 Carotid Atherosclerosis
C0729733 Aortiliac atherosclerosis
C1579923 Precerebral atherosclerosis
C2711237 Atherosclerosis of artery
C2733226 Atherosclerosis of coronary artery
C0004155 Atherosclerosis of other specified arteries
C0155735 Atherosclerosis of arteries of the extremit
C0348648 Atherosclerosis of other arteries
C0398356 Abdominal aortic atherosclerosis
C1541100 Atherosclerosis: [precerebral] or [cerebral
C2585376 History of aortiliac atherosclerosis
C2711885 Atherosclerosis of bypass graft of limb
C1579936 (Aortic atherosclerosis) or (aorto-iliac disc
C2711922 Atherosclerosis of nonautologous bypass

Basic View Report View Raw View

Concept: [C0004153] Atherosclerosis

Semantic Types

Disease or Syndrome [T047]

Definitions

CSP/PT | age, lifestyle, diet, and gene related degeneration of arteries due to deposition of lipid plaques (atheromas) on inner arterial walls; main cause of coronary artery disease, a leading cause of death.

MSH/MH | A thickening and loss of elasticity of the walls of ARTERIES that occurs with formation of ATHEROSCLEROTIC PLAQUES within the ARTERIAL INTIMA.

NCI/PT | Build-up of fatty material and calcium deposition in the arterial wall resulting in partial or complete occlusion of the arterial lumen.

Atoms (149) string [AUI / RSAB / TTY / Code]

- atherosclerosis [A0474128/AOD/DE/000005410]
- ATHEROSCLEROSIS [A0387068/CCPSS/PT/0061599]
- VASCULAR DISEASE ATHEROSCLEROTIC [A1612437/CCPSS/PT/0061608]
- ATHEROSCLEROSIS [A0387067/COSTAR/PT/082]
- ATHEROSCLEROTIC VASCULAR DISEASE [A0530700/COSTAR/PT/U000111]
- atherosclerosis [A0474129/CSP/PT/0571-2508]
- ATHEROSCLEROSIS [A0387068/CST/GT/ARTERIOSCLEROSIS]
- Atherosklerose [A1427104/DMCID10/HT/170]
- ATHEROSCLEROSIS [A0387069/DXP/FI/U000277]
- ATHEROSCLEROSIS [A0387070/DXP/SY/NOCODE]
- Atherosclerosis [A0027478/ICD10/HT/170]
- Atherosclerosis [A0027479/ICD10AM/HT/170]
- Atherosclerosis [A17787957/ICD10CM/HT/170]
- Atherosclerosis [A6763909/ICD10DUT/HT/170]
- Atherosclerosis [A8340125/ICD9CM/HT/440]
- Atherosclerosis [A0027481/CP2P/PT/K92024]

Copyright | Privacy | Accessibility | Freedom of Information Act | National Institutes of Health | Health & Human Services

Obr. 8. Ukázka z Metathesauru.

Semantic Network (Sémantická síť) UMLS

Sémantická síť se skládá

1. z množiny obecných předmětových kategorií neboli sémantických druhů, které poskytují konzistentní kategorizaci všech konceptů zastoupených v Metathesauru UMLS a
2. z množiny užitečných a důležitých vztahů neboli sémantických vztahů, které existují mezi sémantickými druhy. Tato část dokumentace poskytuje přehled o Sémantické síti a popisuje soubory Sémantické sítě. Ukázky záznamů ilustrují strukturu a obsah těchto souborů.

Rozsah Sémantické sítě UMLS je široký a umožňuje sémantickou kategorizaci široké škály terminologie v mnohonásobných doménách. Hlavní seskupení sémantických druhů zahrnuje organismy, anatomické struktury, biologické funkce, chemikálie, události, fyzické předměty a koncepty nebo pojmy. Spojení mezi sémantickými druhy obstarávají strukturu sítě a reprezentují důležité vztahy v oblasti biomedicíny. Hlavní propojení mezi sémantickými druhy je propojení „isa“. Toto propojení zakládá hierarchii druhů uvnitř Sítě a používá se k rozhodnutí o nejspecifičtějším sémantickém dostupném druhu pro stanovení konceptu Metathesauru. Existují také skupiny nehierarchických vztahů, které se seskupují do pěti hlavních kategorií:

1. „fyzicky související s“,
2. „prostorově souvisejí s“,
3. „časově související s“,
4. „funkčně související s“ a
5. „konceptuálně související s“.

Mezi informace spojené s každým sémantickým druhem patří specifický identifikátor, číslo stromu ukazující na pozici v „isa“ hierarchii, definice a přímý rodič a potomci. Mezi

informace spojené s každým vztahem patří specifický identifikátor, číslo stromu, definice a množina sémantických druhů, které mohou být přijatelným způsobem spojeny tímto vztahem.

Sémantická síť UMLS je poskytována ve dvou formátech: relační tabulka a záznamová jednotka. Je dostupná na <http://semanticnetwork.nlm.nih.gov/>. Její šíření podléhá smluvním podmínkám.

SPECIALIST lexicon (ODBORNÝ lexikon) UMLS

ODBORNÝ lexikon je jedním ze tří znalostních zdrojů UMLS, který byl vytvořen Národní lékařskou knihovnou jako část projektu UMLS.

ODBORNÝ lexikon byl vytvořen k poskytování lexikálních informací potřebných pro ODBORNÝ systém zpracovávající přirozený jazyk. Jeho cílem je vytvoření obecného anglického lexikonu, který by zahrnoval mnoho biomedicínských termínů. Pokrývá jak běžně se objevující anglická slova, tak i biomedicínskou slovní zásobu. Každý záznam ve slovníku zaznamenává syntaktické, morfologické a ortografické (pravopisné) informace, které jsou potřebné pro ODBORNÝ systém zpracovávající přirozený jazyk.

Lexikon se skládá z množiny lexikálních položek. Každá pravopisná varianta tvoří jednu položku. Lexikální položky mohou být víceslovné termíny vytvořené z jiných slov, jestliže je stanoveno, že víceslovný termín se chová v obecné angličtině nebo v medicínských slovnících nebo v medicínských tezaurech jako je např. MeSH jako jedna lexikální jednotka. Rozšíření obecně používaných akronymů a zkratk rovněž umožnilo brát je jako víceslovné termíny.

Slova pro lexikální kódování jsou vybírána z mnoha zdrojů. Jádro tvoří zhruba 20 000 slov. Ty jsou vzaty z Testovací sbírky UMLS MEDLINE abstraktů a to společně se slovy, která se objevují jak v Metathesauru UMLS, tak i v *Dorland's Illustrated Medical Dictionary*. Navíc byla snaha zahrnout i slova z obecné anglické slovní zásoby. Bylo

zakódováno 10 000 nejčastějších slov objevujících se v *The American Heritage World Frequency Book* a 2000 slov, která se používají v definicích v *Longman's Dictionary of Contemporary English*. Jelikož většina slov vybraných ke kódování jsou podstatná jména, byla snaha zahrnout také slovesa a přídavná jména a to ze současných citačních záznamů v MEDLINE, pomocí *Computer Usable Oxford Advanced Learner's Dictionary* a eventuálních přídavných jmen z *Dorland's Illustrated Medical Dictionary* pomocí heuristiky, kterou vyvinula v roce 1990 McCray a Srinivasan.

Při kódování lexikálních záznamů se využívá množství referenčních zdrojů. Kódování je založeno na současném používání v Testovací sbírce UMLS a MEDLINE, slovnících obecné angličtiny, zejména ve studentských slovnících, které zaznamenávají druh syntaktické informace potřebné pro zpracovávání přirozeného jazyka a z medicínských slovníků. Byly použity tyto slovníky:

- *Longman's Dictionary of Contemporary English*,
- *Dorland's Illustrated Medical Dictionary*,
- *Collins COBUILD Dictionary*,
- *The Oxford Advanced Learner's Dictionary* a
- *Webster's Medical Desk Dictionary*.

ODBORNÝ lexikon je poskytován ve dvou formátech: záznamová jednotka a relační tabulka. Informace spojené s každou lexikální položkou obsahují jedinečný identifikátor, základní tvar, kód syntaktické kategorie, informaci o dohodnuté shodě, komplementační informaci, jestliže je potřebná a další vlastnosti, které se vztahují k dané lexikální položce. Záznamová jednotka je rámcová struktura, která se skládá z mezer a výplní. Mezery jsou základní lexikální atributy a výplně vyjadřují možné hodnoty těchto atributů pro určitou lexikální položku. Data pro lexikální záznamy jsou také vyjadřována v sadě relačních tabulek. Relační formát lexikonu není plně normalizován. Záměrně je zde duplikace dat mezi různými

vztahy a uvnitř některých vztahů. Vývojáři se budou muset rozhodnout v jakém rozsahu tuto redundanci ve svých aplikacích zachovat, snížit nebo zvýšit. Mezi dalšími tabulkami jsou samostatné tabulky pro informace o shodě a skloňování, časování, komplementační vzory, pravopisné varianty, zkratky a akronymy a jejich plné formy.

ODBORNÝ lexikon je k dispozici jako open source zdroj jako část ODBORNÝCH nástrojů pro zpracování přirozeného jazyka (<http://SPECILIST.nlm.nih.gov>). Šíření podléhá smluvním podmínkám.

3. 2. 2. MetamorphoSys ULMS

MetamorphoSys je instalační průvodce a program, pomocí něhož se přizpůsobuje systém potřebám uživatele a který je součástí každého vydání UMLS. Instaluje jeden nebo více znalostních zdrojů UMLS. Jestliže si vyberete Metathesaurus, můžete si vytvořit přizpůsobené podmnožiny Metathesauru.

Uživatelé si přizpůsobují své podmnožiny Metathesauru ze dvou hlavních důvodů:

- aby vyloučili slovníky, které nepotřebují nebo které v lokálních aplikacích podléhají licenci;
- aby si přizpůsobili podmnožinu, která využívá množství datových výstupů a filtrů.

Metathesaurus se skládá z mnoha souborů, z nichž některé jsou neobyčejně velké. To, že se vyloučí některé zdroje, může značně snížit velikost výstupní podmnožiny. Vzhledem k počtu a škále slovníků, které se objevují v Metathesauru, je nepravděpodobné, že by se vyžadovaly všechny nebo většina z více než 100 slovníků. Navíc, některé zdroje vyžadují samostatné licenční smlouvy pro jejich specifické využití, které uživatel nebudete chtít získat. Jsou uvedeny v Příloze k licenční smlouvě. Dodatečné informace o některých zdrojových

slovníkách můžete nalézt na domovské stránce UMLS pod heslem Metathesaurus Source Vocabularies (Zdrojové slovníky Metathesauru).

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.

1. První část tvoří *administrativní údaje*, které jsou potřebné pro identifikaci pacienta.
2. Další částí je *rodinná anamnéza*, zahrnující informace o matce, otci a libovolném počtu sourozenců.
3. Dále následuje *sociální anamnéza a toxikománie*, která se zaměřuje na rodinný stav, fyzickou zátěž, psychickou zátěž, fyzické aktivity, míru kouření a míru požívání alkoholu.
4. Část MDMK je věnována *alergiím* pacienta, zejména alergiím na léky.
5. 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.

6. V části MDMK nazvané *Současné obtíže možného kardiálního původu* se lékaři zaměřují na dušnost, bolest na hrudi, palpitace, otoky, synkopu, kašel, hemoptýzu a klaudikaci.
7. Další část MDMK zjišťuje, jakou *léčbu* pacient podstupuje, jaký má předepsaný druh diety a jaké užívá léky.
8. V části *fyzikálních vyšetření* se zjišťuje pacientova hmotnost, výška, tělesná teplota, obvod boků, BMI, WHR, krevní tlak, tepová a dechová frekvence a patologické nálezy.
9. *Laboratorní vyšetření* se zaměřují na glykémii, kyselinu močovou, celkový cholesterol, HDL-cholesterol, LDL-cholesterol a triacylglyceroly.
10. 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 **D**atového **M**odelu EuroMISE 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.

Pacient: Rodné číslo: Datum narození:

Záznamy | Administrativa | RA | SA | Alergie | OA | Obtíže | Léčba | Fyz. vyš. | Lab. vyš. | EKG

Počet Záznamů:

Rodinný stav: Žije sám? Nejvyšší c
Celková psychická zátěž: Fyzická zátěž v zaměstnání: nekuřák
 příležitostný kuřák (méně než 1 cig. denně)
 kuřák Kouří od věku: Prům. počet cig. denně:
 exkuřák Kouřil od věku: Prům. počet cig. denně:

Pivo: Počet sklenic 0,5 l/týden
Vino: Počet sklenic 0,2/týden
Destiláty: Počet sklenic 0,05/týden

Obr. 9. Ukázka aplikace ADAMEK.

Pacient: Rodné číslo: Datum narození: Věk:

Záznamy Administrativa RA SA Alergie OA Obtíže Léčba Fyz. vyš. Lab. vyš. EKG

bratr 1

P Ř Í B U Z N Í Narození, úmrtí Vyjmenované choroby Jiná závažná onemocnění

matka
otec
bratr 1
sestra 1
bratr 2

Přidat Sestru
Přidat Bratra

Počet Záznamů:

Přítomnost následujících chorob:

Diabetes mellitus?	<input type="text" value="Ne"/>	
Hypertenze?	<input type="text" value="Ne"/>	
Hyperlipoproteinémie?	<input type="text" value="Nevím"/>	
I.CH. srdeční?	<input type="text" value="Ne"/>	
Infarkt myokardu?	<input type="text"/>	Rok zjištění: <input type="text"/>
Angina pectoris?	<input type="text"/>	Rok zjištění: <input type="text"/>
Cévní moz. příhoda?	<input type="text" value="Ne"/>	Rok zjištění: <input type="text"/>
I. CH. dolních končetin?	<input type="text" value="Ne"/>	Rok zjištění: <input type="text"/>

Přidat Záznam

Obr. 10. Ukázka aplikace ADAMEK.

Pacient: Jan Novák Rodné číslo: Datum narození: Věk: 58 let

znamy Administrativa RA SA Alergie OA Obtíže Léčba Fyz. vyš. Lab. vyš. EKG

I.Ch. Srdeční I. Ch. Periferních Tepen Cévní mozkové příhody Další rizikové faktory Jiné choroby Úmrtí

Přítomna? Ano

Počet Záznamů Formy ICHS Poznámky

2

24.5.2001
18.5.1998

Angina pectoris Némá ischemie InfarktMyokardu Srdeční selhání Arytmie

Přítomna? Ano

Jednotlivé případy

P Ř Í H O D Y Datum zjištění: 2001 Léčen? Ano

Příhoda 1
Příhoda 2

Poznámka:

Typ léčby: Konzervativní
 PTCA
 PTCA+Stent
 Chirurgická (bypass)
 Kardiopulmonální resuscitace
 Fibrinolýza

Přidat Příhodu

Přidat Záznam Zrušit

Uložit vše Zpět

Obr. 11. Ukázka aplikace ADAMEK.

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 takového 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.

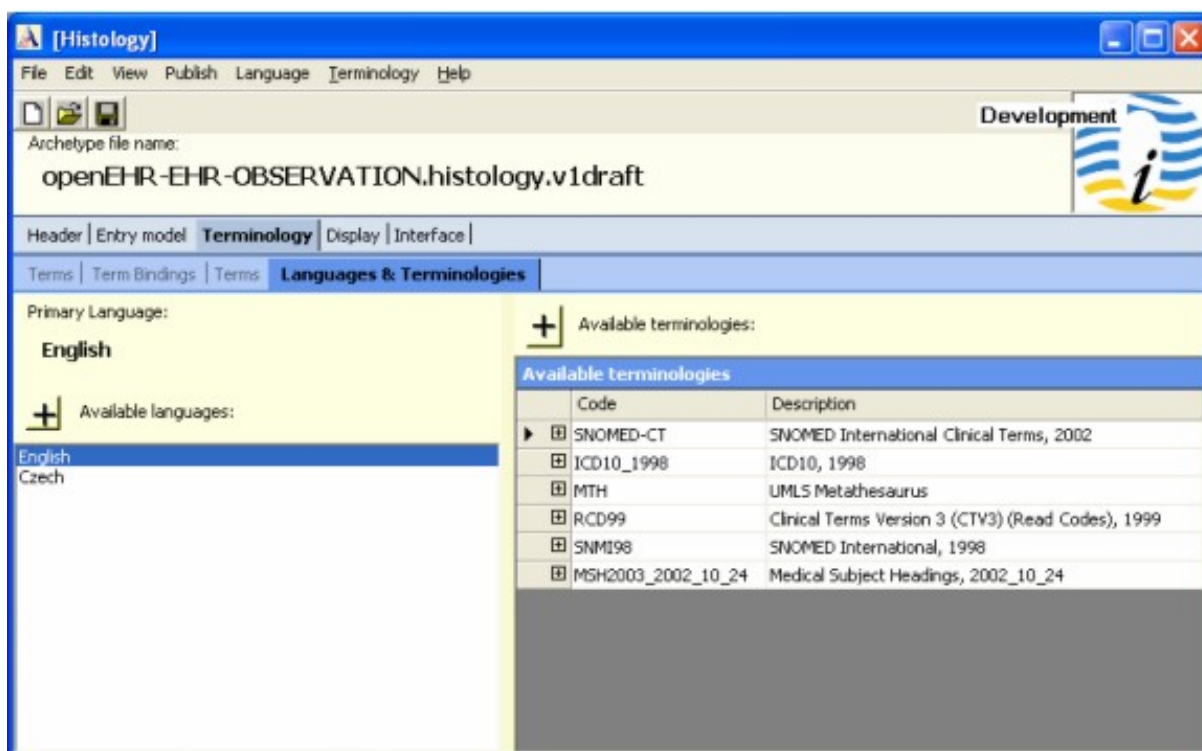
Referenční model Synapses Object Model (SynOM) vytvořený v rámci projektu Synapses, resp. SynEx (Synergy on the Extranet) [37] je velmi podobný modelu definovanému v CEN ENV 13606. Jako typy sbíraných hodnot jsou zde využity tzv. archetypy – definice strukturovaně sbíraných údajů v určité doméně obsahující 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.

4. 1. 1. Mapování terminologie a tvorba archetypů

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.

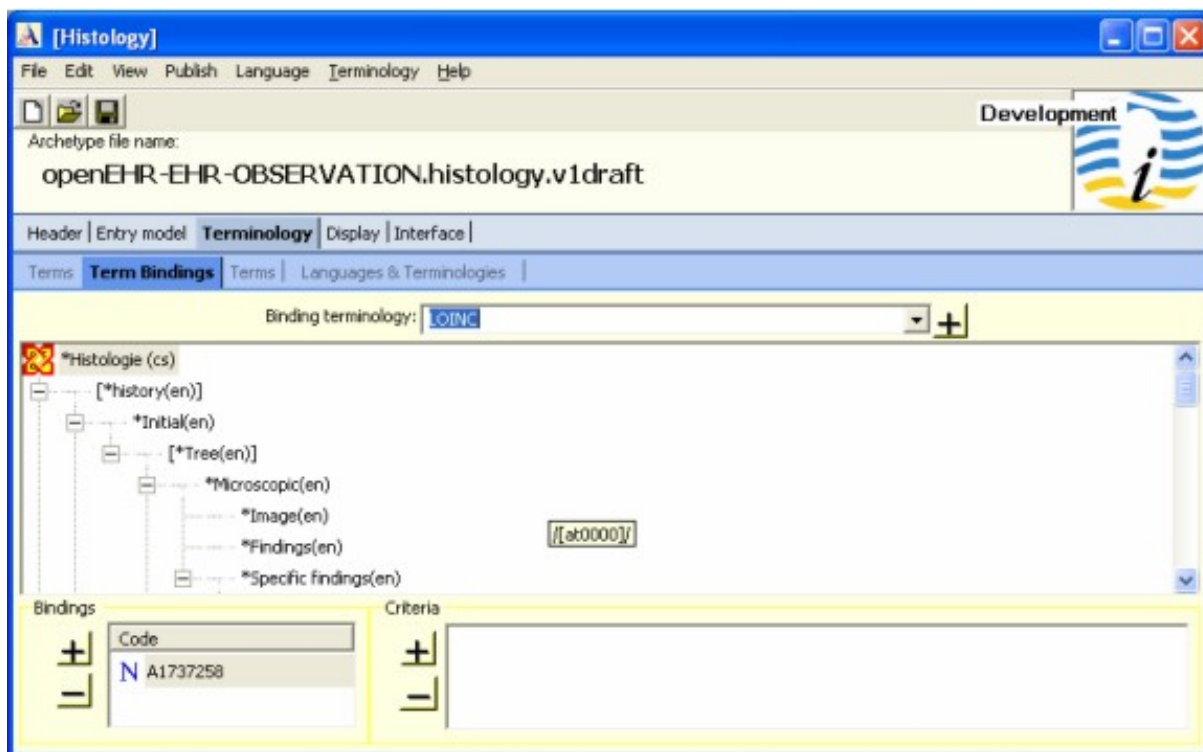
Jako příklad, jak správnou terminologii odkazovat již při vytváření archetypů, může fungovat editor od firmy *Ocean Informatics* [39] zobrazený na Obrázku 12.



Obr. 12. Práce s terminologií v editoru archetypů.

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 tak, jak je zobrazeno na Obrázku 13.



Obr. 13. Mapování použité terminologie na standardní kódovací systémy.

4. 1. 2. Standardizace klinického obsahu

Analýzu vhodnosti a využitelnosti jednotlivých terminologických slovníků jsem započala mapováním klinického obsahu tzv. Minimálního datového modelu pro kardiologii (MDMK) na různé terminologické klasifikační systémy. Jak již bylo zmíněno v kapitole 3. 3. MDMK je souborem přibližně 150 atributů, jejich vzájemných vztahů, integritních omezení a jednotek. Na těchto atributech se shodli přední odborníci v oblasti české kardiologie jako na základních údajích nutných při vyšetření kardiologického pacienta.

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.

Atributy z pohledu možnosti jejich mapování na standardní kódovací systémy lze klasifikovat následujícím způsobem:

- **Bezproblémové atributy** – tj. atributy, které lze mapovat přímým způsobem, tak, že u nich existuje právě jedna možnost mapování, případně existují pouze synonyma se zcela stejným významem a tedy i klasifikačním kódem (např.: křestní jméno pacienta, současný kuřák, hybnost, výška pacienta).
- **Částečně problematické atributy** – tj. atributy, které lze mapovat tak, že u nich existuje právě několik možností mapování na různá synonyma, která se ale lehce liší významem a tedy zpravidla i klasifikačním kódem (např.: ischemická cévní mozková příhoda, angína pectoris, hypertenze, městnavé srdeční selhání).
- **Atributy s příliš malou granularitou**, tj. atributy popisující určitou vlastnost na příliš obecné úrovni tak, že klasifikační systémy obsahují pouze termíny užšího významu (např. email v MDMK vs. email do zaměstnání / email domů / email lékaře atd. v klasifikačních systémech).
- **Atributy s příliš velkou granularitou**, tj. atributy popisující určitou vlastnost na úzké úrovni tak, že klasifikační systémy obsahují pouze termíny obecnějšího významu (např. šelest nad AO ústím, souměrný tep karotid, atd.).
- **Atributy, které se v klasifikačních systémech dohledat nedaří**, např. rodné číslo, dyslipidemie, atd.

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

4. 2. 1. Český jazyk

Český jazyk patří k západní skupině slovanských jazyků. Mezi slovanskými jazyky patří do východní, nebo-li satemové, skupiny indoevropských jazyků. Český jazyk se od ostatních slovanských jazyků odlišil několika změnami, které proběhly v rozmezí 10. až 16. století. Ke konci 15. století ztratila čeština prakticky duální číslo a dva ze slovanských minulých časů. Na druhou stranu ale významněji rostl mluvený aspekt a zvýšil se počet skloňování. Na začátku 15. století navrhnul církevní reformátor Jan Hus diakritický pravopis. Umístil diakritická znaménka nad některá písmena latinky a tím odlišil české palatální souhlásky (č, d', ň, ř, š, t', ž) a dlouhé samohlásky (á, é, í, ó, ú, ý). V 16. století bylo ještě přidáno písmeno „ů“, které označovalo dlouhé „u“. Jediné zdvojené písmeno v moderní češtině je „ch“.

Česky mluví zhruba 10 milionů lidí v České republice a asi 200 000 v dalších zemích. Jedná se většinou o emigranty a jejich děti, kteří opustili zemi ve značných migracích kolem 1. a 2. světové války a kolem roku 1948 a 1968. Mnoho česky mluvících obyvatel žije zejména v Rakousku (většinou ve Vídni), v Polsku, Německu, na Ukrajině, v Chorvatsku, v západním Rumunsku, v Austrálii a v Kanadě. Nicméně, největší skupina česky mluvících lidí mimo Českou republiku žije ve Spojených státech amerických, ve městech jako New York, Chicago a v několika venkovských komunitách v Texasu, Wisconsinu, Minnesotě a Nebrasce.

Čeština patří mezi jazyky s volným slovosledem [42].

4. 2. 2. Jazykový rozbor textových lékařských zpráv

Styl zapisování textových zpráv není nijak standardizován [43] a často 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. „Brieho 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. 2. 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 zaznamenán, je z důvodu jeho nepřítomnosti nebo zda se na něj lékař nezeptal.

V následujících tabulkách jsou ukázány různé způsoby, jak byly vybrané atributy Minimálního datového modelu zaznamenány různými lékaři ve zkoumaných 110 lékařských zprávách psaných volným textem.

Tab. 1. Způsoby zaznamenání atributu **Kouření** ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	kuřák	1
2.	kouří do x cig/den	1
3.	kouří do x cig.	1
4.	kouří cca x/den	1
5.	kouří x/den	1
6.	kouří x cigaret denně	2
7.	kouření x-y za den	1
8.	kouření x-y denně	1
9.	kouření x let	1
10.	kouření ano	2
11.	asi x cig./denne	2
12.	kouření x	1
negativní výskyt		
1.	nekouří	8
2.	kouření 0	15
3.	nekuřák/nekuřačka	9
4.	x roků nekouří	4
5.	kouřil do x let	2
6.	kouřil x/den	1
7.	nadále nekouří	1
8.	býv. kuřák	1
9.	x let kouřila	1
10.	x let nekuřák	1
11.	exkuřák	1
12.	kouření ne	3
13.	kouření negativní	1
14.	ex-kuřák	1
15.	nekouří od roku x	4
16.	nekuřák	1
17.	wxkurák	1
18.	t.č. nekouří	1

Tab. 2. Způsoby zaznamenání atributu *Alergie* ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	AA x,y	23
2.	alergie na	4
negativní výskyt		
1.	neudáva	4
2.	neudává	4
3.	neguje	19
4.	negat	6
5.	negat.	6
6.	negativní	5
7.	neg	5
8.	není znama	1
9.	AA 0	8
10.	AA: O	3
11.	neg.	2

Tab. 3. Způsoby zaznamenání atributu *Ischemická choroba srdeční* ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	susp. ICHS	3
2.	chron. ICHS	6
3.	chron ICHS	2
4.	chronická ICHS	1
5.	ICHS chron	1
6.	ICHS	41
7.	pacientka s ICHS	2
8.	pozitivní RA pro ICHS	2
9.	pozit. RA pro ICHS	4
10.	posit RA pro ICHS	1
11.	poz RA pro ICHS	1
12.	pozitivní rodinná anamnesa pro ICHS	1
negativní výskyt		
1.	ICHS neguje	1
2.	ICHS 0	1

Tab. 4. Způsoby zaznamenání atributu **Dušnost** ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	zadýchává se	7
2.	dušnost již při minimální námaze	1
3.	námahová dušnost	3
4.	námah dušnost	1
5.	námah. dušnost	1
6.	dušnost	5
7.	dušná	3
8.	horší dech	1
9.	hůž dýchá	1
10.	špatně se mu dýchá	1
11.	pacient s dušností	1
12.	udává dušnost	1
13.	občasné dechové potíže	1
14.	s klidovou dušností	1
negativní výskyt		
1.	není dušn	1
2.	není dušný/á	12
3.	dušnost nepozoruje	1
4.	dušnost neudává	3
5.	bez dušnosti	22
6.	dycha se dobře	1
7.	dýchá se jí dobře	3
8.	bez dušnos tIO	1
9.	dechové potíže nemá	1
10.	bez dušnosti	1
11.	dušnost 0	1
12.	dušnostO	1
13.	není dušnost	1
14.	dušná nebyla	1
15.	bez klidové dušnosti	1

Tab. 5. Způsoby zaznamenání atributu **Bolest na hrudi** ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	bolesti na hrudi	4
2.	bolesti na hrudní	1
3.	bolest na hrudi	2
4.	tlak za hrudní kostí	1
5.	bol. na hrudi	2
negativní výskyt		
1.	bez bolestí na hrudi	4
2.	bez bolesti na hrudi	6
3.	bolesti na hrudi neudává	2
4.	bolesti na hrudi 0	3
5.	bolesti na hrudi nemá	4
6.	bolesti na hrudi nejsou	1
7.	boelsti na hrudi nemá	1
8.	bolesti na hrudi neguje	2
9.	bol. na hrudi nemá	2
10.	bolesti na hriudi nemela	1
11.	není bol. na hrudi	1
12.	bez boelstí hrudníku	1

Tab. 6. Způsoby zaznamenání atributu **Palpitace** ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	palpitace	6
negativní výskyt		
1.	bez palpitací	3
2.	palpitace 0	3
3.	nepocit'uje palpiatce	1
4.	palpitace nemá	1
5.	palpitace neguje	1
6.	palpitace neg	1
7.	palpitace nemela	1

Tab. 7. Způsoby zaznamenání atributu **Otoky** ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	pozoroval otoky	1
2.	DK - mírné otoky	1
3.	DK - měkké otoky	1
4.	má otoky dolních končetin	1
5.	DK - otoky	1
6.	DK - oteká PDK	1
7.	DK s otokem	3
8.	otoky DK	1
9.	na DK otoky	1
10.	DK s otoky	1
11.	otok L lýtky	1
negativní výskyt		
1.	DK bez otoků	28
2.	DK bez otoku	8
3.	DK: bez otoku	5
4.	dk bez otoků	13
5.	DK - bez otoků	5
6.	DK: bez otoků	2
7.	DK ez otoků	1
8.	LDK bez otoku	1
9.	DKK – bez otoků	2
10.	DKk otoky O	1
11.	DKK otoky O	2
12.	otoky DK 0	5
13.	DK neotékají	4
14.	DK neotekají	1
15.	otoky DK nepozoruje	4

Tab. 8. Způsoby zaznamenání atributu **Hmotnost** ve 110 lékařských zprávách psaných volným textem.

počet možností	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	hmotnost	42
2.	váha	37
3.	vaha	3
4.	V	11
5.	phmotnost	1

Tab. 9. Způsoby zaznamenání atributu **Diabetes mellitus** ve 110 lékařských zprávách psaných volným textem.

	možnost zaznamenání	počet výskytů
pozitivní výskyt		
1.	DM	15
2.	DM II	1
3.	diabetes mellitus II	1
4.	DM 2.typu	8
5.	DM 2. typu	7
6.	DM II. typu	10
7.	DM II typu	1
8.	diabetes mellitus 2. typu	1
9.	diabetes mellitus II. typu	5
10.	diabetik/diabetička 2. typu	1
11.	diabetik	1
negativní výskyt		
1.	DM 0	9
2.	DM O	4
3.	DM neguje	3
4.	DM nepotvrzen	2

4. 2. 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.

V tabulce 10 je přehledně zobrazeno procentuální vyjádření o zaznamenaných hodnotách vybraných znaků MDMK v 1119 lékařských zprávách při užití softwarové aplikace ADAMEK a ve 110 textových lékařských zprávách.

Tab. 10. Procentuální vyjádření zaznamenaných hodnot vybraných znaků MDMK v lékařských zprávách při užití aplikace ADAMEK a v textových lékařských zpráv.

znak MDMK	aplikace ADAMEK n=1119	textové lékařské zprávy n=110
kouření	96,5 %	64,5 %
alergie	95,2 %	81,8 %
ischemická choroba srdeční	93,4 %	60,9 %
dušnost	93,6 %	71,8 %
bolest na hrudi	93,7 %	34,5 %
palpitace	94,2 %	15,5 %
otoky	93,8 %	86,4 %
výška	97,9 %	85,5 %
diabetes mellitus	95,9 %	62,7 %

Na rozdíl od textových lékařských zpráv je diverzita zaznamenávání atributů Minimálního datového modelu pro kardiologii ve strukturované softwarové aplikaci ADAMKE daleko nižší. V následujících tabulkách jsou ukázány různé způsoby, jakými byly vybrané atributy Minimálního datového modelu zaznamenány různými lékaři ve zkoumaných 1119 strukturovaných lékařských zprávách uložených v aplikaci ADAMEK.

*Tab. 11. Způsoby zaznamenání atributu **Kouření** v 1119 strukturovaných lékařských zprávách.*

počet možností	možnost zaznamenání	počet výskytů
1.	kuřák	203
2.	příležitostný kuřák	19
3.	exkuřák	128
4.	nekuřák	730

*Tab. 12. Způsoby zaznamenání atributu **Alergie** v 1119 strukturovaných lékařských zprávách.*

počet možností	možnost zaznamenání	počet výskytů
1.	ano	378
2.	ne	681
3.	nevím	6

*Tab. 13. Způsoby zaznamenání atributu **Ischemická choroba srdeční** v 1119 strukturovaných lékařských zprávách.*

počet možností	možnost zaznamenání	počet výskytů
1.	ano	44
2.	ne	992
3.	nevím	9

Tab. 14. Způsoby zaznamenání atributu **Dušnost** v 1119 strukturovaných lékařských zprávách.

počet možností	možnost zaznamenání	počet výskytů
1.	námahová	95
2.	klidová	18
3.	žádná	934

Tab. 15. Způsoby zaznamenání atributu **Bolest na hrudi** v 1119 strukturovaných lékařských zprávách.

počet možností	možnost zaznamenání	počet výskytů
1.	námahová	40
2.	klidová	47
3.	žádná	962

Tab. 16. Způsoby zaznamenání atributu **Palpitace** v 1119 strukturovaných lékařských zprávách.

počet možností	možnost zaznamenání	počet výskytů
1.	ano	150
2.	ne	904

Tab. 17. Způsoby zaznamenání atributu **Otoky** v 1119 strukturovaných lékařských zprávách.

počet možností	možnost zaznamenání	počet výskytů
1.	ano	56
2.	ne	994

Tab. 18. Způsoby zaznamenání atributu *Diabetes mellitus* v 1119 strukturovaných lékařských zprávách.

počet možností	možnost zaznamenání	počet výskytů
1.	ano	48
2.	ne	1025

4. 2. 5. Atributy Minimálního datového modelu pro kardiologii zakódované pomocí SNOMED CT

Tabulka 19 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. Celkově se povedlo namapovat více než 60 % atributů MDMK na SNOMED CT.

Tab. 19. Vybrané atributy MDMK zakódované pomocí SNOMED CT.

Atributy z MDMK		English equivalent	SNOMED CT (Concept ID)
rodinný stav			
	svobodný/á	Marital status: single, never married (finding)	125725006
	ženy/vdaná	Legally married (finding)	36629006

	vdovec/vdova	Widowed (finding)	33553000
	rozvedený/á	Divorced (finding)	20295000
	jiný	other	
žije sám		Lives alone (finding)	105529008
nejvyšší dosažené vzdělání			
	základní	Educated to secondary school level (finding)	224297003
		Continued education to sixth form (finding)	224298008
	středoškolské	Received higher education (finding)	224299000
		Received polytechnic education (finding)	224301007
		Received higher education college education (finding)	224302000
	vysokoškolské	Received university education (finding)	224300008
	alergie na léky		drug allergy (disorder)
Allergic reaction to drug (disorder)			416093006
hypertenze		essential hypertension (disorder)	59621000
		High blood pressure (& [essential hypertension])	194757006
		Essential hypertension NOS (disorder)	266228004
hyperlipoproteinémie		Hyperlipoproteinemia (disorder)	3744001
		Fredrickson type IV hyperlipoproteinemia (disorder)	238085009
		Fredrickson type I hyperlipoproteinemia (disorder)	238086005

	Familial type 5 hyperlipoproteinemia (disorder)	34349009
	Familial hyperlipoproteinemia (disorder)	238038003
	Familial type 3 hyperlipoproteinemia (disorder)	398796005
	Fredrickson type IIa hyperlipoproteinemia (disorder)	397915002
ischemická choroba srdeční	Ischemic heart disease (disorder)	414545008
dušnost	Asthma (disorder)	187687003
bolest na hrudi	Dull chest pain (finding)	3368006
palpitace	(Palpitations) or (awareness of heartbeat) or (fluttering of heart)	161965005
otoky	Swelling or edema (finding)	248477007
synkopa	Syncope (disorder)	271594007
klaudikace	Claudication (finding)	275520000
hmotnost	On examination - weight NOS (finding)	162770007
	Height and weight (observable entity)	162879003
výška	Body height measure (observable entity)	50373000
tělesná teplota	Body temperature finding	105723007
	Body temperature (observable entity)	276535009
obvod pasu	Abdominal girth measurement (procedure)	48094003
dechová frekvence	respiratory rate (observable entity)	86290005

Celá tabulka je ke zhlédnutí v Příloze.

4. 2. 6. Atributy Minimálního datového modelu pro kardiologii zakódované pomocí MKN-10

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, viz tabulka 20. Pro srovnání jsou uvedeny rovněž kódy atributů MDMK v systému SNOMED CT [49].

Tab. 20. Vybrané atributy MDMK zakódované pomocí MKN-10 a SNOMED CT.

1. část – Alergie

Atributy z MDMK	Termín v MKN 10	Kód MKN 10	English equivalent	SNOMED CT (Concept ID)
alergie přítomna	alergie	T78.4	allergy manifested	nenalezeno
alergie na léky	alergie na lék	T88.7	drug allergy (disorder)	416098002
			allergic reaction to drug (disorder)	416093006

Tab. 20. Vybrané atributy MDMK zakódované pomocí MKN-10 a SNOMED CT.

2. část – Osobní anamnéza

Atributy z MDMK	Termín v MKN 10	Kód MKN 10	English equivalent	SNOMED CT (Concept ID)
diabetes mellitus	diabetes typu I	E10.-	diabetes mellitus type 1 (disorder)	46635009
	inzulin dependentní	E10.-	insulin- treated non- insulin- dependent diabetes mellitus (disorder)	237599002
	těhotenský	O24.4	pregnancy and insulin- dependent diabetes mellitus (disorder)	237626009
hypertenze	Esenciální (primární) hypertenze	I10	essential hypertension (disorder)	59621000

hyperlipoproteinémie	hyperlipoproteinémie	E78.5	hyperlipoproteinemia (disorder)	3744001
	Fredericksonova typu IV	E78.1	Frederickson type IV hyperlipoproteinemia (disorder)	238085009
	Fredericksonova typu I	E78.3	Frederickson type I hyperlipoproteinemia (disorder)	238086005
	Fredericksonova typu IIa	E78.0	Frederickson type IIa hyperlipoproteinemia (disorder)	397915002

Tab. 20. Vybrané atributy MDMK zakódované pomocí MKN-10 a SNOMED CT.

3. část – Současné potíže možného kardiovaskulárního původu.

Atributy z MDMK	Termín v MKN 10	Kód MKN 10	English equivalent	SNOMED CT (Concept ID)
dušnost	dušnost	R06.8	asthma (disorder)	187687003
bolest na hrudi	bolest hrudníku	R07.4	dull chest pain (finding)	3368006
palpitace	palpitace (srdce)	R00.2	(palpitations) or (awareness of heartbeat) or (fluttering of heart)	161965005
synkopa	synkopa srdeční	R55	syncope (disorder)	271594007
kašel	kašel	R05	cough	158383001
hemoptýza	hemoptýza	R04.2	haemoptysis	158384007

Celá tabulka je ke zhlédnutí v Příloze.

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

4. 3. 1. Tradiční míry diverzity

Pro kompletní charakteristiku atributu je nezbytné stanovit si kategorie každého atributu. Pro daný atribut určíme kategorie A_1, \dots, 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 $\mathbf{p} = (p_1, \dots, p_k)$, $\sum_{i=1}^k p_i = 1$.

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

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

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 $\mathbf{p} = \mathbf{u} = (1/k, 1/k, \dots, 1/k)$ pro rovnoměrné rozdělení pravděpodobnosti. Původně byl tento index Ginim [50] navržen jako míra nerovnosti v příjmech a později diskutován Simpsonem [51] jako míra ekologické diverzity.

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

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

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 $\mathbf{p} = \mathbf{u} = (1/k, \dots, 1/k)$.

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 představuje 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].

4.3.2. 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 .

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

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 Shannonova entropie, které je dosaženo, jestliže $Y = X$. V případě, že atribut X má kategorie A_1, A_2, \dots, A_k , vyskytující se s pravděpodobnostmi p_1, p_2, \dots, p_k , potom je **Shannonova entropie** atributu X stejná jako Shannonův informační index

$$H_S(\mathbf{p}) = -\sum_{i=1}^k p_i \log p_i. \text{ Dále budeme nazývat tuto míru diverzity } \mathbf{Shannonovou diverzitou}.$$

Shannonova informace může být zobecněna na f-informaci

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

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, \dots, A_k a rozdělením pravděpodobnosti $\mathbf{p} = (p_1, p_2, \dots, p_k)$, potom je **f-entropie** atributu X

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

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 $H_f(\mathbf{p})$ mírou diverzity závislé na distribuci \mathbf{p} . $H_f(\mathbf{p})$ 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 \mathbf{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 součet 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 [58] 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 $\mathbf{p} = \mathbf{u}$. 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.

4. 3. 3. Míry vzácnosti, vlastní a marginální f-diverzita

V případě, že X je atribut s kategoriemi A_1, \dots, A_k a pravděpodobnostním rozdělením $\mathbf{p} = (p_1, \dots, p_k)$, 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

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

Tři široce používané indexy diverzity jsou:

Počet kategorií (Diverzita počtu kategorií)

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

Gini-Simpsonův index (Gini-Simpsonova diverzita)

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

a Shannonův index (Shannonova diverzita)

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

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

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

Vidíme, že pro $\beta = 0$ získáme Shannovou diverzitu, pro $\beta = 1$ Gini-Simpsonovu diverzitu a pro $\beta = -1$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

kde $\tilde{\mathbf{u}} = (1/2, 1/2)$.

4. 3. 4. 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 Tabulce 21 shrnujeme 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 následující počet kategorií a vypočítali jsme Diverzitu počtu kategorií (Tab. 21).

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.

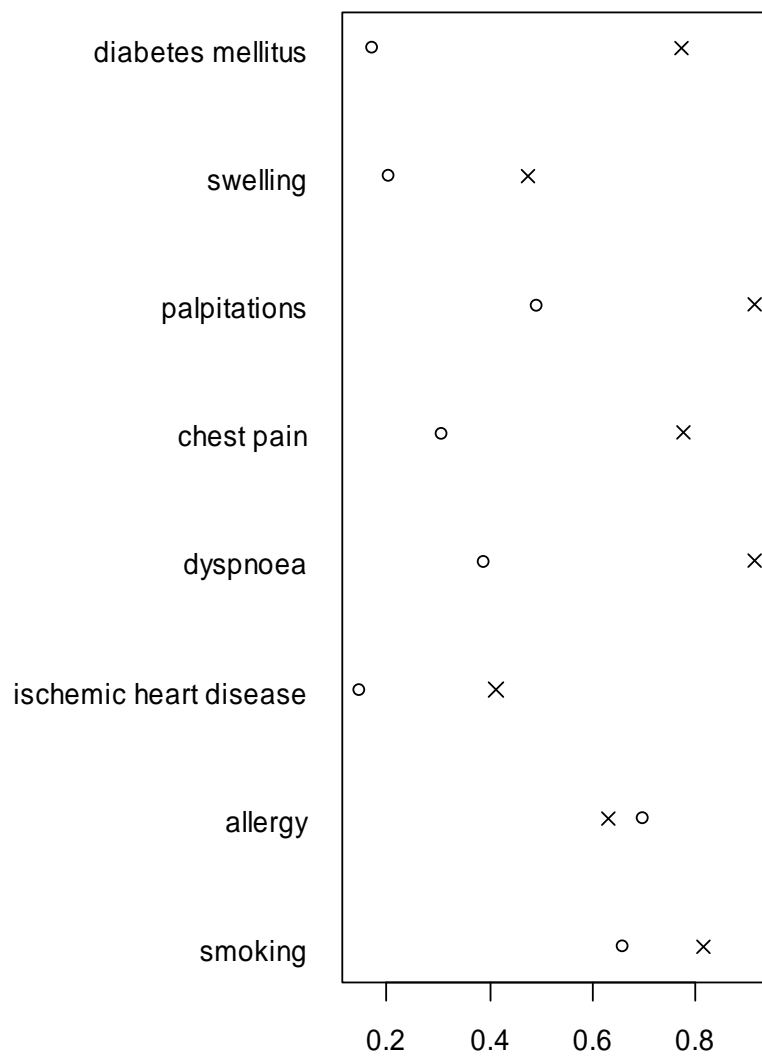
Tab. 21. Počet kategorií vybraných atributů MDMK nalezených ve 110 textových zprávách a počet kategorií v 1119 strukturovaných zprávách.

Attribut	Počet textových zpráv se zaznamenaným atributem	Počet textových zpráv s chybějícím atributem	Diverzita počtu kategorií (textové zprávy)	Počet strukturovaných zpráv se zaznamenaným atributem	Počet strukturovaných zpráv s chybějícím atributem	Diverzita a počtu kategorií (strukturované zprávy)
Kouření	71	39	29	1080	39	3
Alergie	90	20	12	1065	54	2
Ischemická choroba srdeční	67	43	12	1045	74	2
Dyušnost	79	31	28	934	72	2
Bolest na hrudi	38	72	16	1049	70	2
Palpitace	17	93	7	1054	65	1
Otoky	95	15	25	1050	69	1
Diabetes mellitus	69	41	14	1073	46	1

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 (Tab. 22) a Gini-Simpsonovy vlastní diverzity a relativní marginální diverzity (Tab. 23). 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 $\frac{1}{2}$. 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. Obrázek 14 znázorňuje Gini-Simpsonovy relativní diverzity vybraných atributů kategorizovaných podle MDMK v textových lékařských zprávách (X) a ve strukturovaných lékařských zprávách (O). Nicméně, předpokládáme, že odhady vypočítané z textových lékařských zpráv budou ovlivněny velkým množstvím chybějících pozorování.



Obr. 14. Gini-Simpsonovy relativní diverzity vybraných atributů v textových (X) a strukturovaných (O) lékařských zprávách.

Tab. 22. Gini-Simpsonovy relativní diverzity v analýze vybraných atributů ve 110 textových a 1119 strukturovaných lékařských zprávách.

Atribut	Počet kategorií	Počet textových zpráv	Gini-Simpsonova relativní diverzita	Počet strukturovaných zpráv	Gini-Simpsonova relativní diverzita	p hodnota
Kouření	4	71	0,8162	1080	0,6579	0,00073
Alergie	3	90	0,6300	1065	0,6977	0,30932
Ischemická choroba srdeční	3	67	0,4117	1045	0,1455	0,00359
Dušnost	2	79	0,9152	1047	0,3851	<0,00001
Bolest na hrudi	2	38	0,7756	1049	0,3042	0,00050
Palpitace	2	17	0,9135	1054	0,4882	0,00263
Otoky	2	95	0,4725	1050	0,2020	0,00914
Diabetes mellitus	2	69	0,7713	1073	0,1709	<0,00001

Tab. 23. Gini-Simpsonovy vlastní diverzity a relativní marginální diverzity ve 110 textových a 1119 strukturovaných lékařských zprávách pro vybrané atributy kategorizované podle MDMK.

Atribut	Kategorie	Počet textových zpráv	Gini-Simpsonova vlastní diverzita	Gini-Simpsonova relativní marginální diverzita	Počet strukturovaných zpráv	Gini-Simpsonova vlastní diverzita	Gini-Simpsonova relativní marginální diverzita
Kouření	kuřák	15	0,7887	0,6665	203	0,8120	0,6105
	příležitostný kuřák	0	1,0000	0,0000	19	0,9824	0,0691
	exkuřák	19	0,7324	0,7840	128	0,8815	0,4179
	nekuřák	37	0,4789	0,9982	730	0,3241	0,8762
Alergie	ano	27	0,7	0,84	378	0,6451	0,9158
	ne	63	0,3	0,84	681	0,3606	0,9222
	nevím	0	1,0	0,00	6	0,9944	0,0224
Ischemická choroba srdeční	ano	56	0,1642	0,5489	44	0,9579	0,1613
	ne	11	0,8358	0,5489	992	0,0507	0,1926
	nevím	0	1,0000	0,0000	9	0,9914	0,0342
Dušnost	ano	28	0,6456	0,9152	113	0,8921	0,3851
	ne	51	0,3544	0,9152	934	0,1079	0,3851
Bolest na hrudi	ano	10	0,7368	0,7756	87	0,9171	0,3042
	ne	28	0,2632	0,7756	962	0,0829	0,3042
Palpitace	ano	6	0,6471	0,9135	150	0,8577	0,4882

	ne	11	0,3529	0,9135	904	0,1423	0,4882
Otoky	ano	13	0,8632	0,4725	56	0,9467	0,202
	ne	82	0,1368	0,4725	994	0,0533	0,202
Diabetes mellitus	ano	51	0,2609	0,7713	48	0,9553	0,1709
	ne	18	0,7391	0,7713	1025	0,0447	0,1709

5. DISKUSE

Svoji prací jsem se snažila o ověření praktické použitelnosti mezinárodně používaných terminologických slovníků, tezurů, 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 neexistence odpovídajícího termínu v klasifikačních systémech. Velkým problémem při využití nomenklatur a metatezurů 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 (viz tabulka 20 nebo Příloha).

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 attributech 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. SNOMED CT ale může i překročit přímou péči o pacienty. 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 metatezurů 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 diverzity 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 měř 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 Telemedicine „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. Instruktažní 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 Organization: International Classification of Diseases (ICD). ©2011, domovská stránka 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., Jouvett 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 stránka 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 approach for solving terminologic problems, *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 patientských 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 patientských 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.

PŘÍLOHA

Porovnání atributů MDMK zakódovaných pomocí MNK 10 a SNOMED CT

Atributy z MDMK		Termín v MKN 10	Kód MKN 10	English equivalent	SNOMED CT (Concept ID)
Sociální anamnéza a toxikománie					
rodinný stav				marital status	
	svobodný/á	nenalezeno		Marital status: single, never married (finding)	125725006
	ženatý/vdaná	nenalezeno		Legally married (finding)	36629006
	vdovec/vdova	nenalezeno		Widowed (finding)	33553000
	rozvedený/á	nenalezeno		Divorced (finding)	20295000
	jiný	nenalezeno		other	
žije sám		nenalezeno		Lives alone (finding)	105529008
nejvyšší dosažené vzdělání				the highest educational attainment	
	základní	nenalezeno		Educated to secondary school level (finding)	224297003
				Continued education to sixth form (finding)	224298008
	středoškolské	nenalezeno		Received higher education (finding)	224299000
				Received polytechnic education (finding)	224301007
				Received higher education college education (finding)	224302000

	vysokoškolské	nenalezeno		Received university education (finding)	224300008
celková psychická zátěž		nenalezeno		Stress (qualifier value)	26218800
				Work stress (finding)	248265004
				Level of stress (observable entity)	405052004
	žádná	nenalezeno		none	nenalezeno
	mírná	nenalezeno		Moderate (severity modifier) (qualifier value)	6736007
				Increased stress (finding)	23085004
	střední	nenalezeno		middle	nenalezeno
	vysoká	nenalezeno		high	nenalezeno
fyzická zátěž v zaměstnání				physical load in job	
	žádná - sedí celou prac. dobu	Nedostatek tělesného cvičení (námahy)	Z72.3	any - sits all working time	nenalezeno
	mírná - 4 hodiny ve stoje nebo v pohybu	nenalezeno		moderate - standing or moving for 4 hours	nenalezeno
	střední - celou pracovní dobu v pohybu	nenalezeno		middle - moves all working time	nenalezeno
	vysoká - fyzicky náročné zaměstnání (kopáč, horník...)	nenalezeno		high - physically demanding job (digger, miner...)	nenalezeno
tělesná aktivita mimo zaměstnání		nenalezeno		physical activity out of job	nenalezeno
				Physical activity (observable entity)	68130003
				Leisure physical activity (observable entity)	4751000
				Physical activity finding (finding)	129861002
	žádná	nenalezeno		None	nenalezeno
				Physical activity target minimal exercise (finding)	408578001

	mírná	nenalezeno		Moderate	nenalezeno
		nenalezeno		Physical activity target light exercise (finding)	408580007
	střední	nenalezeno		middle	nenalezeno
	vysoká	nenalezeno		High	nenalezeno
Physical activity target strenuous exercise (finding)				408579009	
kouření		nenalezeno		Smoking	nenalezeno
				Tobacco smoking behavior (observable entity)	228484007
	nekuřák	nenalezeno		Non-smoker (finding)	8392000
				Never smoked tobacco (finding)	266919005
	příležitostný kuřák	nenalezeno		Occasional cigarette smoker (finding)	230059006
	kuřák	nenalezeno		Smoker	nenalezeno
				Moderate cigarette smoker (finding)	230062009
				Heavy cigarette smoker (finding)	230063004
				Heavy smoker (over 20 per day) (finding)	56771006
				Very heavy cigarette smoker (finding)	230064005
				Chain smoker (finding)	230065006
				Light cigarette smoker (1-9 cigs/day) (finding)	160603005
				Moderate cigarette smoker (10-19 cigs/day) (finding)	160604004
				Heavy cigarette smoker (20-39 cigs/day) (finding)	160605003
				Very heavy cigarette smoker (40+ cigs/day) (finding)	160606002

				Trivial cigarette smoker (less than one cigarette/day) (finding)	266920004
	od kdy kouří	nenalezeno		Age at starting smoking (observable entity)	228488005
	prům.množství cigaret denně	nenalezeno		Tobacco smoking consumption (observable entity)	266918002
				Cigarette consumption (observable entity)	230056004
exkuřák		nenalezeno		Cessation of smoking (life style)	360907006
				Ex-heavy cigarette smoker (20-39/day) (finding)	266924008
				Ex-very heavy cigarette smoker (40+/day) (finding)	266925009
				Ex-smoker (finding)	8517006
				Ex-heavy smoker (20-39/day) (finding)	160610004
				Ex-very heavy smoker (40+/day)	160611000
				Ex-moderate cigarette smoker (10-19/day)	137990008
				Ex-trivial smoker (<1/day) (finding)	160607006
				Ex-light smoker (1-9/day) (finding)	160608001
				Ex-moderate smoker (10-19/day) (finding)	160609009
				Ex-trivial cigarette smoker (<1/day) (finding)	266921000
				Ex-light cigarette smoker (1-9/day) (finding)	266922007
					266923002

Ex-moderate cigarette smoker

				(10-19/day) (finding)			
	od kdy kouřil	nenalezeno		Age at starting smoking (observable entity)	228488005		
	prům.množství cigaret denně	nenalezeno		Tobacco smoking consumption (observable entity)	266918002		
Ex-cigarette smoker amount unknown (finding)				266928006			
Ex-smoker - amount unknown				137996002			
	od kdy nekouří	nenalezeno		Time since stopped smoking (observable entity)	228486009		
alkohol		Alkohol, NS	T51.9	Alcohol	svobodný		
				Alcohol intake above recommended sensible limits (finding)	160592001		
				Social drinker	28127009		
				Pattern of alcohol consumption through week (observable entity)	228311005		
				Current non-drinker of alcohol (finding)	105542008		
				Požívání alkoholu	Z72.1	Alcohol intake (observable entity)	160573003
						Alcohol consumption unknown (finding)	160580001
						Drinks alcohol evenly through week (finding)	228312003
						Drinks alcohol unevenly through week (finding)	228313008
						Alcohol consumption NOS	160599005
				Alcohol units/week (qualifier value)	228958009		

	pivo	nenalezeno		Beer intake (observable entity)	230085005
				Beer drinker (finding)	160589000
				Drinks beer and spirits	137970004
				Beer drinker	137973002
				Drinks beer and hard liquor (finding)	160590009
	víno	nenalezeno		Wine intake (observable entity)	230086006
				Drinks wine (finding)	160591008
	destiláty	nenalezeno		Drinks beer and spirits	137970004
				Hard liquor intake (observable entity)	230088007
				Drinker of hard liquor (finding)	160588008
				Drinks beer and hard liquor (finding)	160590009
	Alergie				allergy
alergie přítomna	alergie, NS	T78.4	allergy manifested		
alergie na léky	alergie na lék	T88.7	drug allergy (disorder)	416098002	
			Allergic reaction to drug (disorder)	416093006	
Osobní anamnéza				personal history	
diabetes mellitus	diabetes typu I	E10.-	Diabetes mellitus (disorder)	73211009	
	inzulin dependentní	E10.-	Diabetes mellitus without complication (disorder)	111552007	
	Dříve existující diabetes mellitus, závislý na podávání insulínu	O24.0	Diabetes mellitus type 1 (disorder)	46635009	

	Dříve existující diabetes mellitus, nezávislý na podávání insulínu	O24.1	Insulin-treated non-insulin-dependent diabetes mellitus (disorder)	237599002
	Dříve existující malnutricí vyvolaný diabetes mellitus	O24.2	Malnutrition-related diabetes mellitus - fibrocalculous (disorder)	237600004
	Dříve existující diabetes mellitus, NS	O24.3	Secondary endocrine diabetes mellitus (disorder)	237601000
	Diabetes mellitus vzniklý v těhotenství	O24.4	Transitory neonatal diabetes mellitus (disorder)	237603002
	Diabetes mellitus v těhotenství, NS	O24.9	Diabetes mellitus autosomal dominant type II (disorder)	237604008
	Novorozenec ký diabetes mellitus	P70.2	At risk of diabetes mellitus (finding)	161641009
	těhotenský	O24.4	Unstable type I diabetes mellitus (disorder)	290002008
	Screeningové vyšetření specializované na diabetes mellitus	Z13.1	Type II diabetes mellitus with ketoacidosis (disorder)	408660003
	Diabetes mellitus v rodinné anamnéze	Z83.3	Pregnancy and insulin-dependent diabetes mellitus (disorder)	237626009
			Insulin dependent diabetes mellitus type IA (disorder)	23045005
			Brittle diabetes mellitus (disorder)	11530004

			Unstable diabetes mellitus (disorder)	275918005
			Protein-deficient diabetes mellitus (disorder)	57886004
			Drug-induced diabetes mellitus (disorder)	5368009
			Malnutrition related diabetes mellitus (disorder)	75524006
			Diabetes mellitus due to insulin receptor antibodies (disorder)	75682002
			Insulin dependent diabetes mellitus type IB (disorder)	28032008
			Diabetes mellitus associated with genetic syndrome (disorder)	5969009
			Diabetes mellitus type 2 in nonobese (disorder)	359642000
			Diabetes mellitus: [juvenile type, with no mention of complication] or [insulin dependent]	190322003
			Diabetes mellitus: [adult onset, with no mention of complication] or [maturity onset] or [non-insulin dependent]	190323008
			Type I diabetes mellitus without complication (disorder)	313435000

			Diabetes mellitus with no mention of complication (disorder)	190321005
			Type II diabetes mellitus with multiple complications (disorder)	190388001
			Type II diabetes mellitus without complication (disorder)	313436004
			Diabetes mellitus type 2 (disorder)	44054006
			Secondary diabetes mellitus (disorder)	8801005
			Type I diabetes mellitus with complication (disorder)	371054002
			Type II diabetes mellitus with complication (disorder)	371056000
			Diabetes mellitus with complication (disorder)	372069003
			Diabetes mellitus excluded	139838006
			Type II diabetes mellitus	190384004
			Type I diabetes mellitus	190362004
			Diabetes mellitus with no mention of complication	154674007
			Unstable type I diabetes mellitus	190367005
			Type I diabetes mellitus without complication	190373006
			Multiple complications of type II diabetes mellitus (disorder)	420414003

			Multiple complications of type I diabetes mellitus (disorder)	422228004
			Diabetes mellitus with other specified manifestation (disorder)	190417004
			Diabetes mellitus, juvenile type, with other specified manifestation (disorder)	190418009
			Diabetes mellitus, adult onset, with other specified manifestation (disorder)	190419001
			Other specified diabetes mellitus with other specified complications (disorder)	190420007
			Diabetes mellitus NOS with other specified manifestation (disorder)	190421006
			Diabetes mellitus with unspecified complication (disorder)	190422004
			Diabetes mellitus, juvenile type, with unspecified complication (disorder)	190423009
			Diabetes mellitus, adult onset, with unspecified complication (disorder)	190424003
			Other specified diabetes mellitus with unspecified complications (disorder)	190425002

				Diabetes mellitus NOS with unspecified complication (disorder)	190426001
	přítomný	nenalezeno		manifested	nenalezeno
	datum prvního zjištění DM	nenalezeno		date of first finding of DM	nenalezeno
	typ léčby	nenalezeno		type of treatment of DM	nenalezeno
	neléčen	nenalezeno		untreated	nenalezeno
	jen dietou	nenalezeno		Diabetic diet (finding)	160670007
	jen perorálními antidiabetiky (PAD)	nenalezeno		only with oral antidiabetic (OAD)	nenalezeno
				Oral hypoglycemic (product)	346597008
	jen inzulinem	nenalezeno		Insulin treated Type 2 diabetes mellitus	366909003
	PAD + inzulin	Insulin a perorální (hypoglykemizující) antidiabetická léčiva	T38.3	OAD + insulin	nenalezeno
hypertenze		Esenciální (primární) hypertenze	I10	essential hypertension (disorder)	59621000
		Hypertenzní nemoc srdce s (městnavým) srdečním selháním	I11.0	High blood pressure (& [essential hypertension])	194757006
		Hypertenzní nemoc srdce bez (městnavého) srdečního selhání	I11.9	Essential hypertension NOS (disorder)	266228004
	přítomná	nenalezeno		manifested	nenalezeno
	datum prvního zjištění EH	nenalezeno		date of first finding of EH	nenalezeno

	lčba hypertenze		nenalezeno		type of treatment of EH	nenalezeno
		nelčen	nenalezeno		untreated	nenalezeno
		režimovými opatřeními	nenalezeno		life style modification	nenalezeno
					Life style (life style)	60134006
		antihypertenzivy	nenalezeno		Antihypertensive therapy (procedure)	308116003
					Antihypertensive therapy	182823005
hyperlipoproteinémie			hyperliprotot einemie	E78.5	Hyperlipoproteine mia (disorder)	3744001
			Frederickson ova typu IV	E78.1	Fredrickson type IV hyperlipoproteine mia (disorder)	238085009
			Frederickson ova typu I	E78.3	Fredrickson type I hyperlipoproteine mia (disorder)	238086005
			Frederickson ova typu V	E78.3	Familial type 5 hyperlipoproteine mia (disorder)	34349009
			nenalezeno		Familial hyperlipoproteine mia (disorder)	238038003
			Frederickson ova typu III	E78.2	Familial type 3 hyperlipoproteine mia (disorder)	398796005
			Frederickson ova typu IIb	E78.2	nenalezeno	
			Frederickson ova typu IIa	E78.0	Fredrickson type IIa hyperlipoproteine mia (disorder)	397915002
	přítomná		nenalezeno		manifested	nenalezeno
	datum prvního zjištění HLP		nenalezeno		date of first finding of HLP	nenalezeno
	lčba	Treated with (attri-	nenalezeno		Hyperlipoproteine mia diet, type V (finding)	55765009

	bute): 289950 06				Hyperlipoproteine mia diet (finding)	56718006
					Hyperlipoproteine mia diet, type IIa (finding)	32671009
					Hyperlipoproteine mia diet, type IIb and type III (finding)	17902004
					Hyperlipoproteine mia diet, type I (finding)	9617008
					Hyperlipoproteine mia diet, type IV (finding)	77063006
	neléčen	nenalezeno			untreated	nenalezeno
	režimo vými opatřen ími	nenalezeno			life style modification	nenalezeno
	hypolip idemi- ky	nenalezeno			hypolipidaemic drugs	nenalezeno
ischemická choroba srdeční		ischemie koronární	I25.9		Ischemic heart disease (disorder)	414545008
	přítomná	nenalezeno	nenalezeno		manifested	nenalezeno
	datum zjištění	nenalezeno	nenalezeno		date of finding	nenalezeno
	forma	nenalezeno	nenalezeno		form	nenalezeno
	angina pectoris	Angina pectoris, NS	I20.9		angina pectoris (disorder)	367416001
	němá ischemie	ischemie němá (asymptonic ká)	I25.6		Silent myocardial ischemia (disorder)	233823002
	infarkt myokardu	infarkt myokardu, myokardiální (akutní nebo s dobou trvání 4 týdny nebo méně)	I21.9		myocardial infarction (disorder)	22298006
	srdeční selhání	selhání srdce hyperenzní	I11.0		cardiac failure	nenalezeno

		selhání srdce akutní (náhlé)	I50.9	Acute congestive heart failure (disorder)	10633002
		selhání srdce městnavé	I50.0	Congenital cardiac failure (disorder)	206586007
		selhání srdce revmatické (chronické)	I09.9	Acute left-sided congestive heart failure (disorder)	74960003
		Městnavé selhání srdce	I50.0	Acute heart failure (disorder)	56675007
		Selhání srdce, NS	I50.9	Chronic heart failure (disorder)	48447003
		Selhání levé komory	I50.1	Acute left-sided heart failure (disorder)	364006
	arytmie	arytmie (srdeční)	I49.9	irregularity of pulse	nenalezeno
		arytmie eytrasystolická	I49.4	Abnormal pulse rate (finding)	111972009
		arytmie komorová návratná	I47.0	Unequal pulse (finding)	248633009
		arytmie novorozence	P29.1	Slow rising pulse (finding)	248634003
		arytmie psychogenní	F45.3	Rapid upstroke pulse (finding)	248635002
		arytmie určená NJ	I49.8	Pulse rate altered (finding)	248630007
		Návratná - reciproční komorová arytmie	I47.0	Thready pulse	248643007
		Jiné určené srdeční arytmie	I49.8	Finding of alternating pulse volume (finding)	248644001
		Srdeční arytmie, NS	I49.9	Pulse regularly irregular (finding)	271638000
				Pulse irregularly irregular (finding)	271637005
				Abnormal pulse (finding)	67683007
	léčba			treatment	
		žádná	nenalezeno	none	nenalezeno

		konzervativní	nenalezeno		Conservative therapy (regime/therapy)	281131004
		PTCA	nenalezeno		PTCA	nenalezeno
		PTCA + stent	nenalezeno		PTCA + stent	nenalezeno
		chirurgická (bypass)	nenalezeno		surgical (bypass)	nenalezeno
					Single period of cardiopulmonary bypass (procedure)	233569005
					Multiple periods of cardiopulmonary bypass (procedure)	233570006
					Bypass graft with artificial transplant (procedure)	10855008
					Coronary artery bypass with autogenous graft, three grafts (procedure)	10326007
					Coronary bypass graft finding (finding)	251019006
					Coronary artery bypass grafting (procedure)	232717009
					Emergency coronary artery bypass graft (procedure)	414088005
cévní mozkové příhody		Cévní příhoda mozková (mrtvice) neurčená jako krvácení nebo infarkt		I64	Cerebrovascular accident (disorder)	230690007
		Cévní příhoda mozková v rodinné anamnéze		Z82.3	Cerebrovascular disease (disorder)	62914000

	Syndrom cévního postižení mozkového kmene (I60 - I67+)	G46.3	Left sided cerebral hemisphere cerebrovascular accident (disorder)	195216008
	Syndrom cévního postižení mozečku (I60 -I67+)	G46.4	Right sided cerebral hemisphere cerebrovascular accident (disorder)	195217004
	Jiná určená cévní onemocnění mozku	I67.8	Diffuse cerebrovascular disease (disorder)	302909007
	Cévní onemocnění mozku, NS	I67.9	Other cerebrovascular disease (disorder)	195219001
	Jiná cévní onemocnění mozku při nemocech zařazených jinde	I68.8	Other cerebrovascular disease OS (disorder)	195237000
			Other cerebrovascular disease NOS (disorder)	195238005
			Other specified cerebrovascular disease (disorder)	195248007
			Cerebrovascular disease NOS (disorder)	195249004
			CVA - cerebrovascular accident (& unspecified [& stroke])	195208004
			Cerebrovascular: [disease] or [accident]	266312006
			Stroke and cerebrovascular accident unspecified (disorder)	270883006
přítomné	nenalezeno		manifested	nenalezeno

	datum zjištění	nenalezeno		date of finding	nenalezeno
	hospitalizace pro CMP	nenalezeno		hospitalization for CVA	nenalezeno
	léčba pro CMP	nenalezeno		treatment of CVA	
	konzervativní	nenalezeno		Conservative therapy (regime/therapy)	281131004
	intervenční	nenalezeno		interventional	nenalezeno
ischemická choroba periferních tepen				ischemic peripheral artery disease	
	ICHDK	nenalezeno		ischemic lower extremities artery disease	nenalezeno
				Disorder of lower extremity (disorder)	118937003
				Atherosclerosis of arteries of the extremities (disorder)	51274000
				Vascular disorder of extremity (disorder)	363340006
				Vascular disorder of lower extremity (disorder)	373408007
	přítomná	nenalezeno		manifested	nenalezeno
	datum zjištění	nenalezeno		date of finding	nenalezeno
	léčba	nenalezeno		treatment	nenalezeno
	renální tepny			renal arteries	nenalezeno
	Vrozená stenóza renální arterie		Q27.1	Renal arterial hypertension (disorder)	39018007
	Jiné vrozené vady renální arterie		Q27.2	Multiple renal arteries (disorder)	46395002
				Atherosclerosis of renal artery (disorder)	45281005
	přítomná	nenalezeno		manifested	nenalezeno
	datum zjištění	nenalezeno		date of finding	nenalezeno

		léčba	nenalezeno		treatment	nenalezeno
	karotidy		Aneurysma krční tepny (karotidy)	I72.0	Carotids	nenalezeno
					Carotidynia (disorder)	230482003
		přítomná	nenalezeno		manifested	nenalezeno
		datum zjištění	nenalezeno		date of finding	nenalezeno
		léčba	nenalezeno		treatment	nenalezeno
aneurysmata aorty			aneuryzma aorty	I71.9	aneurysm of aorta	nenalezeno
			aneuryzma aorty hrudní (oblouku)	I71.2	Congenital aneurysm of ascending aorta (disorder)	253646008
			aneuryzma aorty syfilitické	A52.0	Syphilitic aneurysm of aorta (disorder)	12232008
			aneuryzma aorty kongenitální	Q25.4	Congenital aneurysm of aorta (disorder)	16972009
			aneuryzma aorty disekující (kterékoli části)	I71.0	Aneurysmectomy of aorta (procedure)	8102001
			aneuryzma aorty chlopně (srdeční)	I35.8	[X]Aneurysm of aorta in diseases classified elsewhere (disorder)	195622005
			aneuryzma roztržené aorty	I71.8	Aorta congenital: [aneurysm] or [dilatation]	204440006
			aneuryzma aorty břišní	I71.4	Repair of aneurysm of abdominal aorta (procedure)	405525004
			aneuryzma sestupné aorty	I71.9	Operation on aneurysm of aorta NEC (procedure)	175350008
			aneuryzma aorty sinu, vrozené	Q25.4	Aneurysm of ascending aorta (disorder)	425963007
			nenalezeno		Aneurysm of descending aorta (disorder)	426948001

				Aneurysm of supraceliac aorta (disorder)	427490006
				Aneurysm of suprarenal aorta (disorder)	426270006
				Chronic dissecting aneurysm of thoracic aorta (disorder)	428326005
	přítomno	nenalezeno		manifested	nenalezeno
	datum zjištění	nenalezeno		date of finding	nenalezeno
	chirurgická léčba	nenalezeno		surgical treatment	nenalezeno
ostatní relevantní choroby pacienta				other relevant diseases of patient	
	kód choroby			code of disease	
	datum prvního zjištění choroby			date of first finding of disease	
menopauza od		nenalezeno		menopause since	nenalezeno
	menopauza	N95.1		Menopause, function (observable entity)	161712005
	Nadměrné krvácení v období před menopausou	N92.4		Menopause present (finding)	289903006
	Postmenopauzální krvácení	N95.0		Normal menopause (finding)	237123000
	Menopauzální a klimakterické stavy u žen	N95.1		Post-hysterectomy menopause (disorder)	278063007
	Stavy spojené s arteficiální menopauzou	N95.3		Post-radiation menopause (disorder)	278064001
	Jiné určené menopauzální a perimenopauzální poruchy	N95.8		Menopause finding (finding)	276477006
	menopauza umělá	N95.3		Artificial menopause (qualifier value)	67886002

	Menopauzální a perimenopauzální porucha, NS	N95.9	Menopause symptoms present (finding)	170951000	
	menopauza předčasná	E28.3	Premature menopause NOS (qualifier value)	237789005	
	menopauza chirurgická	N95.3	Menopause	138977003	
			Postsurgical menopause (disorder)	371036001	
			Difficulty managing menopause (finding)	423478005	
			History of normal menopause (situation)	427956008	
datum úmrtí	nenalezeno		Date of death(observable entity)	408399004	
příčina smrti	nenalezeno		Cause of death (observable entity)	184305005	
			Death diagnosis (contextual qualifier) (qualifier value)	16100001	
	kardiovaskulární	nenalezeno	cardiovascular	nenalezeno	
	detailní příčina kardiovask.	nenalezeno	cardiovascular cause in detail	nenalezeno	
	náhlá smrt	Náhlá srdeční smrt	I46.1	Sudden cardiac death (disorder)	95281009
	ICHS	nenalezeno		ischemic heart disease	nenalezeno
	CMP	nenalezeno		cerebrovascular events	nenalezeno
	jiná	Jiné nepřesně určené a neurčené příčiny smrti	R99	other	nenalezeno
Současné obtíže možného kardiovaskulárního původu			current problems of eventual cardiovascular origin		

dušnost		Dušnost - dyspnoe	R06.0	Asthma (disorder)	187687003
	klidová			in rest	nenalezeno
	námahová			in effort	nenalezeno
bolest na hrudi		bolest hrudníku	R07.4	Dull chest pain (finding)	3368006
	klidová	nenalezeno		in rest	nenalezeno
	námahová	nenalezeno		in effort	nenalezeno
palpitace		palpitace (srdce)	R00.2	(Palpitations) or (awareness of heartbeat) or (fluttering of heart)	161965005
otoky		nenalezeno		Swelling or edema (finding)	248477007
synkopa		synkopa srdeční	R55	Syncope (disorder)	271594007
kašel		kašel	R05	[D]Cough	158383001
hemoptýza		hemoptýza	R04.2	[D] haemoptysis	158384007
klaudikace		nenalezeno		Claudication (finding)	275520000
Léčba				treatment	
dieta				diet	
	současná léčba dietou	dietní dohled a poradenství	Z71.3	current diet cure	nenalezeno
	speciální dieta	nenalezeno		special diet	nenalezeno
	omezení živočišných tuků	nenalezeno		restriction of animal fat	nenalezeno
	omezení uhlovodanů	nenalezeno		restriction of carbohydrate	nenalezeno
				High carbohydrate diet (finding)	226120004
				High complex carbohydrate diet (finding)	226121000
				High simple carbohydrate diet (finding)	226122007
				Carbohydrate-free diet (finding)	226123002

		omezení purinů	nenalezeno		restriction of purine	nenalezeno
		jiná	nenalezeno		other	nenalezeno
lék					drug	
	identifikace (firemní název)	viz Tabulka léků a chemikálií			identification (company term)	nenalezeno
	identifikace (chemický název)				identification (chemical term)	nenalezeno
	obsah účinné látky				active substance content	nenalezeno
	dáv kov ání				dosage	nenalezeno
	množství kdy				amount	nenalezeno
					when	nenalezeno
Fyzikální vyšetření					physical examination	nenalezeno
hmotnost			nenalezeno		On examination - weight NOS (finding)	162770007
					Height and weight (observable entity)	162879003
výška			nenalezeno		Height	nenalezeno
					Body height measure (observable entity)	50373000
tělesná teplota			nenalezeno		body temperature	nenalezeno
					Body temperature finding	105723007
					Body temperature (observable entity)	276535009
obvod pasu			nenalezeno		girth (waist size)	nenalezeno
					Abdominal girth measurement (procedure)	48094003
obvod boků			nenalezeno		hip size	nenalezeno
BMI			nenalezeno		BMI	162859006
WHR			nenalezeno		WHR (waist/hip ratio)	248367009

krevní tlak – pravá horní končetina	krevní tlak-pouze vyšetření	Z01.3	blood pressure - upper right extremity	nenalezeno
	systolický TK	nenalezeno	systolic pressure	nenalezeno
	diastolický TK	nenalezeno	diastolic pressure	nenalezeno
krevní tlak – levá horní končetina			blood pressure - upper left extremity	nenalezeno
	systolický TK	nenalezeno	systolic pressure	nenalezeno
	diastolický TK	nenalezeno	diastolic pressure	nenalezeno
tepová frekvence			pulse rate	nenalezeno
	nenalezeno		Pulse rate finding	301147003
			Pulse rate (observable entity)	78564009
dechová frekvence	nenalezeno		respiratory rate (observable entity)	86290005
somatický nález je fyziologický	nenalezeno		somatic finding is physiological	nenalezeno
patologický nález na karotidách	nenalezeno		pathological finding on carotids	nenalezeno
patologický nález na perif. tepnách	nenalezeno		pathological finding on peripheral arteries	nenalezeno
patologický nález na srdci při fyzikálním vyšetření	nenalezeno		pathological finding on heart when physically examined	nenalezeno
patologický nález na plicích při fyzik. vyšetření	nenalezeno		pathological finding on lungs when physically examined	nenalezeno
jiný patologický nález	nenalezeno		other pathological finding	nenalezeno
Laboratorní vyšetření	Laboratorní vyšetření	Z01.7	[V]Laboratory examination (situation)	316585001
glykémie	nenalezeno		glycaemia	nenalezeno
kyselina močová	kyselina močová v krvi (zvýšení)	E79.0	Uric acid (substance)	1710001
			Uric acid measurement (procedure)	86228006
celkový cholesterol	nenalezeno		Total cholesterol measurement (procedure)	121868005

			Total cholesterol (substance)	301860006
			Total cholesterol:HDL ratio measurement (procedure)	166842003
HDL-cholesterol	nenalezeno		High density lipoprotein (HDL) cholesterol (substance)	102737005
			High density lipoprotein cholesterol measurement (procedure)	28036006
LDL-cholesterol (výpočet)	nenalezeno		LDL-cholesterol (calculation)	nenalezeno
			Low density lipoprotein (LDL) cholesterol (substance)	102739008
			Low density lipoprotein cholesterol measurement (procedure)	113079009
			Calculated LDL cholesterol level	393855003
triacyglyceroly	nenalezeno		triacyglycerols	nenalezeno
EKG	nenalezeno		Standard ECG (procedure)	164847006
			ECG - electrocardiogram (& general)	271334005
			ECG finding (observable entity)	271921002
			ECG not done	142013001
parametry	nenalezeno		characteristics	nenalezeno
	rytmus	nenalezeno	rhythm	nenalezeno
	sinusový	nenalezeno	Sine wave (qualifier value)	262146004

	fibrilace síní	Fibrilace a flutter síní	I48	Atrial fibrillation (disorder)	49436004
				ECG: atrial fibrillation	142049000
	flutter síní	Fibrilace a flutter síní	I48	atrial flutter (disorder)	5370000
				ECG: atrial flutter	142050000
	stimulovaný	nenalezeno		stimulated	nenalezeno
	jiný rytmus	nenalezeno		other rhythm	nenalezeno
	frekvence	nenalezeno		Frequency	nenalezeno
				Unit of frequency (qualifier value)	259023001
	průměrné intervaly	nenalezeno		average periods	nenalezeno
	PQ	nenalezeno		PQ interval	nenalezeno
	QRS	nenalezeno		QRS complex feature (observable entity)	164949005
				ECG: QRS complex NOS	142126004
				ECG: QRS complex normal	142120005
				ECG: QRS complex abnormal	142121009
				Wide QRS complex (finding)	142122002
				Narrow QRS complex	142123007
				Finding of ECG QRS complex (finding)	365415001
hodnocení				ECG finding (observable entity)	271921002
	poruchy vedení	nenalezeno		Conduction (qualifier value)	278252000
				Cardiac conduction (observable entity)	251089002
				Cardiac conduction pathway (observable entity)	251099007

				Normal cardiac conduction (finding)	79525004
				Abnormal cardiac conduction (finding)	36358004
				(Conduction disorders (& [of heart])) or (heart block)	195038000
				Other conduction disorders NOS (disorder)	195065007
				Conduction disorders unspecified (disorder)	195066008
				Conduction disorders NOS (disorder)	195068009
				Conduction disorders (& [heart block])	266302004
				Conduction disorder of the heart (disorder)	44808001
				Conduction disorders (& [heart block])	155354000
				Conduction disorders NOS	155362008
				Cardiac conduction system feature (observable entity)	364069001
				Other conduction disorders (disorder)	195062005
				Finding of cardiac conduction (finding)	366171002
				typ poruchy vedení	nenalezeno
				type of conduction abnormality	nenalezeno
				1st degree block	nenalezeno
	blokáda I.stupně	Atrioventrikulární blokáda, prvního stupně	I44.0		

		atrio-ventrikulární blokáda (Wenckebach)	blokáda atrioventrikulární (neúplná) (částečná)	I44.3	ECG: Wenckebach phenomenon	142065004
		atrio-ventrikulární blokáda (Mobitz)	nenalezeno		Mobitz type I incomplete atrioventricular block (disorder)	54016002
	Mobitz type II atrioventricular block (disorder)				28189009	
	Electrocardiogram : Mobitz type I incomplete atrioventricular block (finding)				426913008	
	Electrocardiogram : Mobitz type II atrioventricular block (finding)				426183003	
		úplná a-v blokáda	úplná atrioventrikulární blokáda	I44.2	Complete atrioventricular block (disorder)	27885002
	Congenital complete atrioventricular block (disorder)				82580003	
	ECG: complete atrioventricular block				142066003	
	Complete atrioventricular block				195063000	
	Electrocardiogram : complete atrioventricular block (finding)				164906009	
		blokáda pravého raménka Tawarova	blokáda pravého raménka	I45.0	ECG: right bundle branch block	142067007

	blokáda levého raménka Tawarova	blokáda levého raménka	I44.7	ECG: left bundle branch block	142068002
	jiná	Jiné určené poruchy vedení	I45.8	other	nenalezeno
	extrasystoly	extrasystoly	I49.4	Extrasystoles (& [unspecified])	195092001
				ECG: extrasystole	142043004
	supraventrikulární	Supraventrikulární tachykardie	I47.1	Supraventricular arrhythmia (disorder)	72654001
				ECG: supraventric. arryth. NOS	142052008
				ECG: supraventricular arrhythmia	142047003
				ECG: no supraventric. arryth.	142048008
	komorové	Návratná - reciproční komorová arytmie	I47.0	ECG: ventricular arrhythmia	142053003
		Komorová tachykardie	I47.2	ECG: no ventricular arrhythmia	142054009
		Komorové kmitání (flutter) a mihání (fibrilace)	I49.0	ECG: ventricular tachycardia	142055005
				ECG: ventricular arrhythmia NOS	142057002
				ECG: ventricular fibrillation	142056006
	abnormální morfologie	nenalezeno		abnormal morphology	nenalezeno
	hypertrofie levé komory	nenalezeno		Left ventricular hypertrophy (disorder)	55827005

					ECG:left ventricle hypertrophy	142033002
					ECG: no LVH	142034008
					ECG: shows LVH	142035009
					ECG: LVH NOS	142036005
celkový popis			nenalezeno		overall ECG description	nenalezeno