

BIG DATA VERSUS VZÁCNÉ PŘÍPADY

Lenka Lhotská, Miroslav Burša, Michal Huptych, Matěj Hrachovina

Anotace

Rozvoj technologií znamená generování, uchovávání, spravování a zpracování čím dál tím větších objemů dat, která jsou heterogenní, multimodální, strukturovaná i nestrukturovaná, někdy zašuměná, neúplná. Některé charakteristiky dat jsou společné pro data z různých problémových oblastí. Data v medicíně však vykazují určitá specifika, mezi něž patří i malé datové soubory popsané velkým počtem příznaků, vzácné případy vyjádřené odlehkými hodnotami příznaků, které nelze ignorovat, ale naopak zohlednit v dalších analýzách. V medicíně se také daleko častěji využívá a bude využívat interaktivních metod dobývání znalostí, kde expert je součástí procesu a jednotlivé kroky analýzy jsou řízeny jeho znalostmi. Na důležitosti nabývají metody z oblasti strukturního učení a grafové modely využívající pravděpodobnostní matematický aparát. Nedílnou součástí zpracování velkých dat je vhodná vizualizace dat, procesu zpracování a výsledků.

Klíčová slova

Big data, dobývání znalostí, vizualizace, multidimenzionální data, dolování dat

1. Úvod

Big data je pojem, který se v posledních letech objevuje čím dál častěji. S tím se také začaly tvořit různé definice. Ve většině najdeme následující formulaci: velká data jsou soubory dat, která kvůli jejich velikosti nelze uchovávat, spravovat a zpracovávat běžně používanými softwarovými prostředky v rozumném čase. Poznamenejme ještě, že tato data jsou velmi heterogenní, nestrukturovaná, někdy i částečně méně věrohodná. Velká data se v současnosti sbírají v mnoha oblastech lidské činnosti a každá z nich přináší jak problémy, tak požadavky spojené s charakterem oblasti a sbíraných dat. V našem příspěvku se zaměříme na problematiku velkých dat v medicíně, jejich specifik a vyvíjených metod pro zpracování takových dat. Na závěr uvádíme očekávané trendy v rozvoji metod pro zpracování a vizualizaci velkých dat v medicíně.

2. Big data

Základní charakteristiky velkých dat lze vyjádřit pomocí pojmu 3V, resp. 4V – z angličtiny volume, velocity, variety a veracity. **Volume** (objem) říká, že objem dat narůstá exponenciálně. **Velocity** (rychlost) reprezentuje požadavek řady úloh, které vyžadují okamžité zpracování velkého objemu průběžně vznikajících dat. **Variety** (různorodost) říká, že pracujeme s různými typy dat: strukturovanými, nestrukturovanými a multimediálními. Poslední vlastnost, **veracity** (věrohodnost), říká, že pořizovaná data nemusejí být vždy stoprocentně spolehlivá, úplná, atd. Tedy jejich věrohodnost může být nejistá právě v důsledku jejich nekonzistence, neúplnosti či nejasnosti. Příkladem mohou být údaje

čerpání z komunikace na sociálních sítích, ale také měřená data, která mohou být zašuměná či zatížena chybou měření.

Jedním, ale ne jediným, problémem objemných dat je způsob jejich ukládání tak, aby bylo možné s nimi následně efektivně pracovat. Dopředu se zpravidla dá velmi těžko odhadovat, jaká bude struktura všech sbíraných dat, jaké budou jejich vzájemné souvislosti, atd. Jednou z možností pro ukládání je použití **NoSQL databází**, které umožňují zpracovávat data bez jasné struktury, řešit závislosti mezi jednotlivými objekty a rychleji vyhledávat požadovaná data. Samozřejmě „klasické“ SQL databáze i NoSQL databáze mají svoje výhody i nevýhody. U existujících implementací se setkáváme s tím, že v NoSQL databázích lze realizovat SQL dotazy. NoSQL se často vykládá jako „Not only SQL“ = „nejen SQL“.

Existuje několik základních typů NoSQL, které se odlišují použitými datovými modely (v závorce uvádíme existující implementace):

- **Column** (Accumulo, Cassandra, Druid, HBase, Vertica);
- **Document** (Clusterpoint, Apache CouchDB, Couchbase, MarkLogic, MongoDB, OrientDB);
- **Key-value** (CouchDB, Dynamo, FoundationDB, MemcacheDB, Redis, Riak, FairCom c-treeACE, Aerospike, OrientDB, MUMPS);
- **Graph** (Allegro, Neo4J, InfiniteGraph, OrientDB, Virtuoso, Stardog);
- **Multi-model** (OrientDB, FoundationDB, ArangoDB, Alchemy Database, CortexDB).

Tyto modely lze porovnávat podle různých kritérií. Scofield [1] zvolil jako základní následující kritéria: výkon, škálovatelnost, flexibilita, složitost a funkcionalita. Volba bude vždy záležet na aplikační oblasti a predikovatelnosti budoucích možných dat, která budeme chtít do databáze přidávat.

Kromě ukládání velkých dat je důležitou otázkou a výzvou do budoucna vizualizace dat. Vizualizace může často napovědět, co se v datech skrývá. Problémem velkých objemů dat totiž je jejich obtížná zpracovatelnost ve smyslu podrobné analýzy tak, jak jsme na ni doposud byli zvyklí. U těchto dat často ani dopředu nevíme, na co se máme zaměřit a co v datech chceme najít. V tomto ohledu dochází k paradigmatickému posunu od klasické vědy, kdy nejprve máme otázku, resp. hypotézu, a následně sbíráme data, k vědám o datech, kdy nejprve máme data a následně klademe otázky. Velkou výzvou je klást relevantní otázky, abychom našli relevantní strukturální vzory a/nebo časové vzory („znalosti“) v takových datech, protože tyto vzory jsou často skryté.

3. Vlastnosti dat

Ještě než se budeme věnovat zpracování dat v medicíně, upřesněme si pro následující text některé pojmy, týkající se dat obecně a vedle toho i specifik medicínských dat. Zaměříme se na otázku struktury a standardizace dat. Data můžeme rozdělit z pohledu struktury do několika kategorií (z popisu pro jednoduchost záměrně vynecháváme matematický formalismus). **Slabě**

strukturovaná data jsou taková, kdy trajektorie mezi jednotlivými body připomíná náhodnou procházku. **Dobře strukturovaná data** jsou v reálném světě v minoritě a v ideálním případě má každý element dat přiřazenu definovanou strukturu, např. relační tabulky. Pojem **špatně strukturovaná data** se často používá jako protiklad dobře strukturovaných dat. Původně ale byl tento pojem použit v jiném kontextu [2]. **Částečně strukturovaná data** je taková forma strukturovaných dat, která není úplně v souladu se striktní formální strukturou tabulek a datových modelů relačních databází, ale obsahuje tagy a markery pro oddělení struktury a obsahu. Typickým příkladem je formalismus XML. **Nestruturovaná data** často nepřesně označují data v přirozeném jazyce. Nicméně i text má jistou strukturu: slova, věty, odstavce. Přesně řečeno, nestrukturovaná data by měla označovat čistě náhodná data, tedy šum. Duda, Hart a Stork [3] tak definují takovou vlastnost dat, která se objeví díky náhodnosti (v reálném světě, ze senzorů či měření). V informatice to mohou být nechtěná nerelevantní data bez významu a vztahu k ostatním datům.

Dále lze data rozdělit na standardizovaná (např. numerické položky v laboratorní zprávě) a nestandardizovaná (např. nestandardizovaný text v databázovém záznamu pacienta, který se často nevhodně označuje jako "volný text"). **Standardizovaná data** jsou základem přesné komunikace. V medicíně může se stejnými daty pracovat více lidí v různých časových okamžicích v různých místech. Datové standardy mají zajistit, že informace jsou prezentovány v takové formě, která podporuje interoperabilitu systému [4] a umožňuje koncovému uživateli porovnávat data při interpretaci. Standardy podporují znovupoužitelnost dat, zlepšují efektivitu zdravotnických služeb a zabraňují chybám, způsobeným např. duplikací vstupů. Standardizace dat se týká následujících oblastí:

- a) Obsah dat;
- b) Terminologie použité k reprezentaci dat;
- c) Způsob předávání dat;
- d) Způsob použití znalostí.

Zde se termín „znalosti“ vztahuje k lékařským doporučením, protokolům, rozhodovacím pravidlům, standardním operačním postupům, apod. Technické elementy pro sdílení dat požadují standardizaci identifikace, struktury záznamů, terminologie, posílání zpráv, ochrany dat, atd.

Nestandardizovaná data, kterých je většina, ohrožují kvalitu dat, výměnu dat a interoperabilitu. Samostatnou kategorií představují **nejistá data**, která jsou současně velkou výzvou v medicíně. Cílem je zjistit, které proměnné (představující prediktory) z milionů proměnných jsou spojené se specifickým výsledkem, jako je stav nemoci. Další komplikací je, že většina lékařských informací je neúplná, se širokou škálou typů a důležitosti chybějících informací. Analýza takových dat je podstatně náročnější než analýza dat s úplnými informacemi. Touto problematikou se metody dolování dat zabývají dlouhodobě. Výsledek je většinou velmi závislý na porozumění problémové oblasti, ze které data pocházejí.

4. Data v medicíně a jejich vlastnosti

Většina dat v medicíně je stále ukládaná izolovaně v různých úložištích a často ani v rámci jednoho státu nelze vytvořit jednoduše nad takovými daty analýzy.

Organizace ve zdravotnictví zpravidla v současnosti používají dvě množiny dat: retrospektivní data, základní informace o událostech sbírané z lékařských záznamů, a klinická data, sbíraná v reálném čase a prezentovaná v místě péče (obrazy, krevní tlak, saturace krve kyslíkem, tepová frekvence, apod.). Například když přijde diabetik do nemocnice a stěžuje si na pocit necitlivosti v prstech u nohou, může lékař místo okamžitého předpokladu, že příčinou je diabetes, vyšetřovat průtok krve a saturaci krve kyslíkem a potenciálně usuzovat, zda příčinou není něco více ohrožujícího, jako je aneurysma nebo mrtvice.

Metody dolování dat a datové analýzy umožnily úspěšně provázat oba typy dat tak, že lékaři mohou použít relevantní informace a použít je pro identifikaci trendů, které budou mít dopad na budoucí zdravotní péči – obor známý jako prediktivní analýza. Takže bude-li u většího množství diabetiků zjištěno znečitlivění prstů u nohou, lze propojením aktuálních klinických a retrospektivních dat pomoci lékařům analyzovat, jaká léčba bude na danou populaci zabírat. To může pomoci rozvíjet preventivní a dlouhodobou péči individualizovanou pro příslušné skupiny pacientů.

Když půjdeme dále a přidáme do této mozaiky data o genových sekvencích, můžeme odhalit souvislosti mezi geny a citlivostí na určité choroby. Zejména v případě některých infekčních chorob může rozhodovat rychlost nasazení léčby o jejím úspěchu. Kromě znalosti genů pacientů je také důležité mít k dispozici genom infekčního organismu a návazně informaci o jeho citlivosti na různá antibiotika. Tak lze potom účinnou léčbu nasadit podstatně rychleji.

Velká data, to není jen otázka velikosti / objemu. Metody dolování dat se snaží najít vzhled do komplexních, zašuměných, heterogenních, dlouhodobých a objemných dat. Snaží se najít odpovědi na otázky, které v minulosti zůstávaly nezodpovězené. Tato témata zahrnují získávání, ukládání, prohledávání, sdílení a analýzu dat.

Velká data na jedné straně, ale také málo dat a hodně příznaků – i to je v medicíně vcelku obvyklé. Vedle metod pro analýzu velkých dat se musejí rozvíjet metody pro analýzu a hledání souvislostí u vzácných případů [5]. V takových případech se okrajové hodnoty (outliers) nesmějí ignorovat, ale naopak zohlednit při dalších analýzách. Zde jsou na místě interaktivní metody dobývání znalostí, kde expert je součástí procesu a přináší své znalosti.

5. Nové trendy v medicíně

Jedním z největších cílů budoucí medicíny je modelování pacientů v celé jejich složitosti tak, aby bylo možné lékařské rozhodování, zdravotní praxi a léčbu přizpůsobit každému jednotlivému pacientovi. Tento trend k personalizované medicíně ale produkuje stále větší objemy dat. A i když experti zvládají na výbornou rozpoznávání v dimenzích ≤ 3 , tak ve vícedimenzionálních datech vznikají problémy. Většina biomedicínských dat však patří do kategorie multi-

dimenzionálních dat, což znesnadňuje manuální analýzu a často ji úplně znemožňuje. V každodenní lékařské praxi je téměř nemožné se věnovat takto složitým datům. Navíc lékaři potřebují znalosti a vhled do souvislosti v datech pro podporu svého rozhodování. Proto se začíná v počítačových vědách objevovat snaha vyvinout efektivní, použitelné a užitečné výpočetní metody, algoritmy a nástroje, které umožní dobývat znalosti a v interaktivním režimu získat vhled do multidimenzionálních dat. Součástí vývoje je také adaptace reprezentací dat. Velkým úkolem je mapování multidimenzionálních dat do méně dimenzí pro snazší interpretaci a vizualizaci. Dalším rysem většiny biomedicínských dat je jejich řídkost, šum a časová závislost.

Expertí ve vědách o živé přírodě se musejí vypořádat s velkými objemy komplexních, multidimenzionálních, heterogenních, zašuměných a málo strukturovaných dat a velkého množství nestrukturovaných informací.

V medicíně se objevuje trend směrem k tzv. **P4-medicíně** (v angličtině Personalized, Predictive, Preventive, Participatory). S tím se ale zvyšují nároky na zdravotnické systémy, protože dochází k nárůstu objemu a složitosti heterogenních, multidimenzionálních dat a také nestrukturovaných informací. Nejvíce narůstá objem dat v tzv. „-omics“ oblastech, např. genomika, proteomika, metabolomika, epigenetika. Trendem je také posun od reaktivní k proaktivní medicíně. P4-medicína je úzce svázána se systémovým přístupem k nemocem a využitím nástrojů pro analýzu obsahu. Známé výzvy týkající se takových dat zahrnují složitost dimenze příznaků (problémy škálování a mapování), heterogenity dat (problémy integrace dat, fúze dat), změny v čase a většinu problémů klasických lékařských dat: nejistota kvality dat, chybná a neúplná data a nebezpečí artefaktů. Často zmiňovaný problém velkých dat může být ale výhodou pro použití metod strojového učení: V biomedicíně používáme často jen několik stovek příkladů, což může znamenat nebezpečí náhodného hádání. Velká data znamenají, že je pro trénování a následné testování daleko více příkladů, a můžeme tak dosáhnout větší přesnosti.

Tudíž kognitivní složitost a mnohaúrovňová vizualizace představují výzvu pro správné porozumění informacím v klinickém kontextu. Za klíčové body považujeme uživatelsky přívětivý a srozumitelný návrh a návrh takových reprezentací informací, které budou „šité na míru“ specifčnosti lidského zpracování informací. To je důležité zejména proto, že roste diverzita koncových uživatelů v čím dál tím složitější oblasti biomedicíny. Cílem práce s komplexními informacemi v medicíně je podpora rozhodování uživatelů. Tudíž výstupy musejí být pro uživatele transparentní a interpretovatelné.

Pro práci s takovými typy dat je nutné hledat vhodné kombinace oblastí a nástrojů, které mohou nabídnout dobré podmínky pro zpracování velkých biomedicínských dat. Takovou synergickou kombinaci představují oblasti interakce člověk – počítač (HCI – Human-Computer Interaction) a dobývání znalostí / dolování dat (KDD – Knowledge Discovery from Data). Hlavním cílem tohoto synergického působení teorií, metod a přístupů HCI a KDD je provázání s koncepty personalizované medicíny. Takové přístupy potřebují samy o sobě

transdisciplinární metodologii. Například porozumění takovým složitým strukturám, jako jsou regulační sítě používané v P4-medicíně, je náročný cíl a nelze ho řešit v rámci jedné izolované disciplíny. Na druhé straně se rozvíjejí i metody zpracování dat. Díky tomu je možné zkoumat metody a míry, které pomohou analyzovat globální a lokální strukturní vlastnosti komplexních dat a jejich vzájemných vztahů.

Základní myšlenkou, která je v poslední době velmi intenzivně rozvíjena, je analýza velkých dat s využitím metod strojového učení a „doktora/experta-ve-zpětné-vazbě“. To znamená, že celý proces není úplně automatický, ale v určitých fázích vstupuje člověk se svými expertními znalostmi do procesu zpracování dat tak, aby zaměřil pozornost systému správným směrem v dalším kroku analýzy.

Kroky k dosažení výsledku v takovém procesu zahrnují:

1. Integraci dat (fúze, mapování, čištění dat, předzpracování)
2. Interaktivní strojové učení (a kognitivní výpočty, dobývání znalostí)
3. Vizualní analytiku (+ expert ve zpětné vazbě, podpora rozhodování)
4. Soukromí, ochrana dat, bezpečnost, etika, uživatelská a genderová studie, hodnocení.

6. Metody dolování dat v biomedicíně

Podívejme se nyní blíže na metody z oblasti dolování dat, které je možné použít na velká data v biomedicíně.

V několika minulých letech se výzkum v dolování dat zaměřil na vývoj efektivních metod strojového učení pro analýzu velkých datových souborů, které se objevují v reálných aplikacích a zejména v biomedicíně. Těmi nejčastěji rozvíjenými metodami jsou metody z oblasti strukturního učení a grafových modelů, jako pravděpodobnostní závislostní sítě (probabilistic dependency networks), pravděpodobnostní rozhodovací stromy, Bayesovské sítě, Markovova náhodná pole.

Řada úloh v této oblasti může být formulována jako pravděpodobnostní inferenční úlohy. Předpokládá se, že další vývoj bude zaměřen na přístupy vycházející z konceptů z oblasti teorie informace, topologického dolování dat, dolování dat využívajícího grafové struktury. Současné tyto metody musejí poskytnout data a výsledky v takových modelech, které budou transparentní a snadno vyjádřitelné pomocí vizualizací tak, aby byly přístupné a srozumitelné pro člověka. Mezi takové metody patří pravděpodobnostní grafové modely, které umožňují reprezentovat různé závislostní vztahy různých typů příznaků.

Zde je také nutné zdůraznit zřejmou skutečnost, že v biomedicíně nelze dolování dat zcela úplně zautomatizovat. Je vhodné a výhodné použít koncept „člověk/expert-ve-zpětné-vazbě“, kdy expert pomocí nástrojů interaktivního učení může jednak optimalizovat následné zpracování nebo odhalit neobvyklé vztahy v datech a tak zaměřit další kroky zpracování žádoucím směrem.

Ruku v ruce s výše zmíněnými postupy jde extrakce a selekce příznaků z dat. Pomocí příznaků lze pak nalézt podskupiny v integrovaných datech. Extrakce

a selekce příznaků je velmi podstatným krokem v analýze biomedicínských dat, ať už jde o analýzu genetických dat, měřených signálů či obrazových dat. Protože pracujeme s integrovanými zdroji heterogenních a multidimenzionálních dat, je nutné k selekci příznaků a hledání podskupin přistupovat trochu jinak – s využitím explorativních a distribuovaných postupů. Hlavním důvodem je rychle rostoucí výpočetní náročnost a tedy i náklady. Zejména pokud uvažujeme, že v budoucnu mohou přibýt nové zdroje dat. Základní strategií může být nalezení kandidátů na vhodné příznaky buď v každé, nebo některé skupině zdrojů dat v prvním kroku. Hledání bude využívat jednoduchá kritéria, jako např. minimální redundanci, nebo použití pravděpodobnostního grafového modelu, závisejícího na charakteristikách každého zdroje dat. Taková selekce může být realizována zcela nezávisle v distribuované podobě nebo s globální koordinací. Tito kandidáti na příznaky ze všech datových zdrojů budou pak použiti pro hledání podskupin a následně nalezení menších množin příznaků, relevantních pro jednotlivé podskupiny. Tyto úlohy vyžadují definování vztahů mezi zdroji dat, které mohou poskytnout experti z daných oblastí, případně je můžeme získat inferencí z dat, např. s využitím pravděpodobnostních grafových modelů. Tyto informace lze také využít pro koordinaci distribuované selekce příznaků, což může snížit redundanci kandidátů.

Ne vždy však máme dostatek dat, aby bylo možné použít pro analýzu např. osvědčené postupy matematické statistiky. Medicína je navíc velmi rozsáhlou oblastí, kde dosud často nejsou jednotlivé podoblasti dostatečně propojeny právě z hlediska souvislosti v měřených a analyzovaných datech. Navíc se neustále díky rozvoji technologií objevují nové způsoby získávání dat. Tudíž je nutné se také zaměřovat na hledání souvislostí mezi oblastmi, které jsou na základě dosavadních znalostí jen volně spojeny nebo nejsou dokonce dosud spojeny vůbec. Zde je důležité identifikovat významné vzory v datech napříč těmito oblastmi. I samotné vzory mohou být reprezentovány různými způsoby (např. koncepty, grafy, strukturní podobnost). Dosud nejsou propracovány vhodné metody a nástroje, které by umožňovaly formalizovat příznaky na vyšší abstraktní úrovni a následně popsat vzájemné vztahy mezi příznaky z různých oblastí. V procesu hledání souvislostí je v tomto případě role expertů nezastupitelná. Proto koncept „expert-ve-zpětné-vazbě“ hraje důležitou roli.

Další podstatnou vlastností medicíny je existence vzácných případů a chorob. Pokud bychom data zpracovali slepě statisticky, tak vzácné případy budou nejspíše označeny jako šum nebo chybná data. Proto je nutné rozvoj metod také zaměřit na detekci odlehlých dat (outliers) a dobývání znalostí z takových dat. V této oblasti je úloha experta ještě podstatnější, protože kromě hledání v datech je součástí zkoumání těchto vztahů prohledávání textů a identifikace implicitních vazeb. Přitom je nutné mít na paměti i zohlednění časově závislých jevů, které je nutné reprezentovat v dlouhodobých sledováních chronických pacientů. Ve všech případech je důležité nezapomenout na zaznamenání kontextových informací, které mohou často objasnit i nečekané souvislosti. Příkladem může být zaznamenání vnějších podmínek, za kterých byla konkrétní data sbírána.

7. Vizualizace informací

V předchozích částech jsme se zmínili o tom, že kromě nárůstu objemu dat v medicíně je velmi podstatným rysem jejich multidimenzionalita a také nárůst nestrukturovaných informací. Přitom lidské kognitivní schopnosti zpracovat taková data a převést je do podoby použitelných znalostí jsou omezené. Proto velkou výzvu představuje nejen vývoj vizualizačních nástrojů pro vyjádření sémantických aspektů informací [6], ale také pochopení kognitivních a komunikačních procesů vnímání informací. Vzhledem k tomu, že takto vyjádřené informace se budou využívat i přímo v reálném čase při poskytování lékařské péče (často s požadavkem na časově kritická rozhodnutí), musí být vizualizace navržena tak, aby odpovídala lidským schopnostem rychle pochopit znázorněné informace. Je nutné mít na paměti variabilitu uživatelů. Každý je schopen vyhodnotit vizualizované informace s jinou rychlostí s ohledem na zkušenost, vnímání vizuální informace, věk, případně pohlaví.

Zpracování multimediálního obsahu, jako je kombinace textu, zvukové informace, obrazů, videa a 3D modelů, klade velké nároky na efektivitu, spolehlivost a flexibilitu použitého hardwaru a softwaru. Jak jsme již zmínili výše, významnou roli při návrhu metod bude čím dál tím více hrát porozumění lidskému vnímání a porozumění obsahu v souvislosti s kognitivní složitostí.

8. Závěr

Dosažení pokroku v dobývání znalostí v komplexních multidimenzionálních datech vyžaduje spolupráci v různých tématech, od předzpracování dat, fúze dat, integrace a mapování dat až po interaktivní vizualizaci v nízkodimenzionálním prostoru. Jako slibné metody se jeví grafové a topologické metody, dále různé optimalizační metody pro výběr popisných příznaků. Je zajímavé, že jak v kognitivních vědách, tak ve strojovém učení se velká část prací soustřeďovala na učení s učitelem a učení bez učitele. Avšak v reálném světě může být výhodnější tzv. semi-supervised learning, tedy kombinace učení s učitelem a bez učitele, kdy pouze malá část dat je klasifikována a daleko větší část klasifikována není.

Další významnou výzvou ve zpracování velkých dat je nalezení vhodných měř „vzdálenosti“, např. pro shlukování, detekci odlehklých případů, měř podobnosti, apod., protože zajímavé vzory v datech se mohou objevit v různých podprostorech celého souboru.

Poděkování

Práce byla podporována projektem č. SGS13/203/OHK3/3T/13 Českého vysokého učení technického v Praze.

Literatura

- [1.] Scofield, B. (2010-01-14). „NoSQL - Death to Relational Databases(?)“. Retrieved 2015-02-26. <http://www.slideshare.net/b scofield/nosql-codemash-2010>

- [2.] Simon, H.A. (1973). *The Structure of Ill Structured Problems*. *Artificial Intelligence*, 4 (3-4), 181-201
- [3.] Duda, R.O., Hart, P.E., Stork, D.G. (2000). *Pattern Classification*, 2nd edition, Wiley, New York.
- [4.] Huptych, M., Lhotská, L. (2014). *Multi-layer Data Model: A Contribution to Semantic Interoperability*. In *IFMBE Proceedings - 6th European Conference of the International Federation for Medical and Biological Engineering*. Heidelberg: Springer, vol. 45, 749-752.
- [5.] Hrachovina, M., Huptych, M., Lhotská, L. (2013). *Data Acquisition and Storage System in a Cardiac Electrophysiology Laboratory*. In *Information Technology in Bio- and Medical Informatics, ITBAM 2013*. Heidelberg: Springer, 77-87
- [6.] Burša, M., Lhotská, L., Chudáček, V., Spilka, J., Janku, P., et al. (2013). *Visualization in Information Retrieval from Hospital Information System*. In *Soft Computing Models in Industrial and Environmental Applications*. Heidelberg: Springer, 459-467

Kontakt:**Lenka Lhotská**

tel.: 224353933

e-mail: lhotska@fel.cvut.cz**Miroslav Burša**e-mail: bursam@fel.cvut.cz**Michal Huptych**e-mail: michalhuptych@seznam.cz**Matěj Hrachovina**e-mail: matej.hrachovina@gmail.com**ČVUT FEL Praha****katedra kybernetiky**

Technická 2

166 27 Praha 6

tel.: 224357325

fax: 224311081

<http://cyber.felk.cvut.cz>