



Konceptualizace, komunikace a reprezentace znalostí

Lékařská informatika

Zimní semestr 2018/2019

Michal Huptych

Proč?



- Při technickém implementačním popisu se často ztrácí určitá část podstaty věcí.
- Snaha uchopit celou problematiku od základu, tak jak dává smysl.
- Vymezení/vyjasnění pojmů pro další diskuzi.
- Vše, co se v posledních 25 letech děje kolem sémantické interoperability v elektronizaci zdravotnictví vychází z následujícího.
- Tento přehled by tedy měl být základem pro další části.

Rozdělení pojmů



- Data
 - Signál měřený v čase a v milivoltech
- Informace
 - Měřený signál zaznamenává elektrickou činnost srdce
- Znalosti
 - Měřený signál nevykazuje žádnou abnormalitu, tj. po stránce elektrické aktivity se srdce jeví zdravé



- Základní reprezentace výsledku pozorování (měření)
- Data jsou nosičem informace
- Obecně (bez znalosti významu) popsitelná pomocí charakterizace nebo parametrizace
 - Statistický popis (četnost, rozložení, střední hodnota, rozptyl)
 - Frekvenční popis
 - Stanovení definičního oboru a oboru hodnot
 - Způsoby hledání informace bez znalosti významu dat
- Častokrát zatížena šumem
 - Komplikace pro následující zpracování



- Několik možných definicí
 - Významový obsah sdělení, zprávy
 - Z pohledu informatiky míra změny neurčitosti vlivem dat
 - dá se dát v protiklad k šumu – šum nepřináší informaci, resp. nezvyšuje míru informace
- Několik možných pohledů
 - Sémantická informace
 - Význam slov ve větě (slova = data, význam = informace)
 - Z pohledu informatiky
 - Míra entropie dat

$$H(S) = - \sum_{i=1}^N P(s_i) \log_2 P(s_i)$$

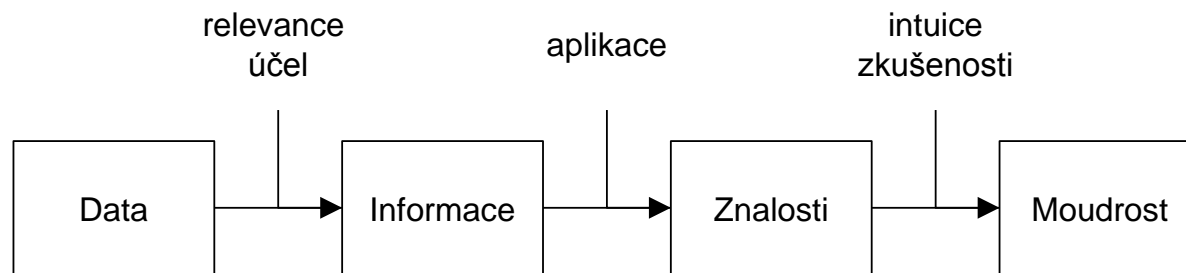


- Poznatky a zkušenosti v určité oblasti
- Dle možnosti zachycení
 - Explicitní - lze formalizovat a tedy jednoduše sdílet
 - Implicitní - není přímo viditelná, lze ji nalézt v datech a případně formalizovat, či aproximovat
 - Tacitní - skrytá znalost (expertní), častokrát v podstatě nevědomé
- Dle způsobu užití
 - Deklarativní - popis platného stavu
 - Procedurální - Popis, jak postupovat
- Dle způsob u získání
 - Znalosti získané od experta
 - Znalosti získané z dat (modelování znalosti)

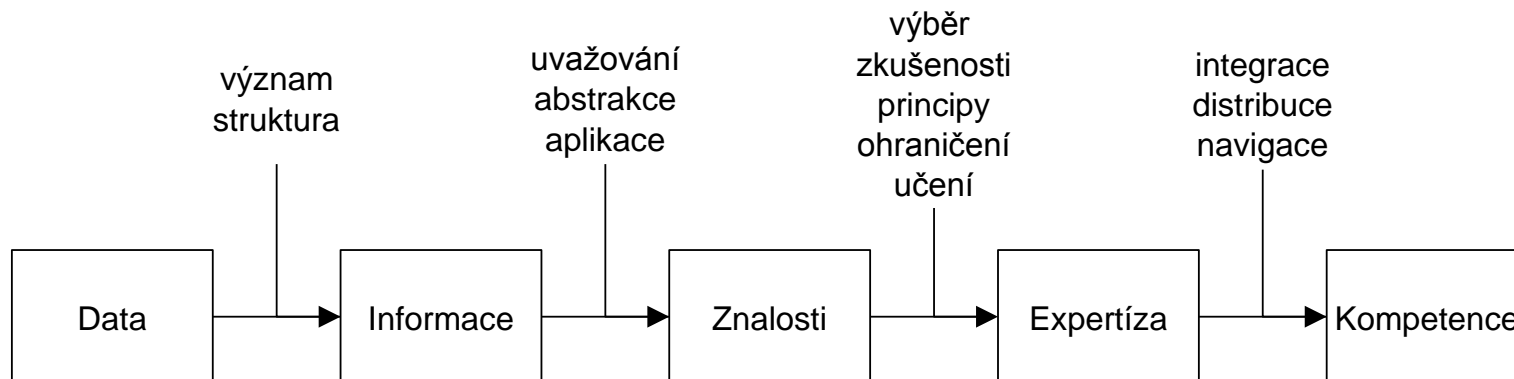
Data – Informace - Znalost



Tobin, 1996



Beckman, 1997



český překlad převzat z <https://sorry.vse.cz/~berka/docs/4iz430/P04-Reprezentace%20znalosti.pdf>

Začátek při řešení úlohy



- Pro každou úlohu máme v sobě nějaký způsob reprezentace světa v daném kontextu.
- Tvoříme kognitivní model, který nám pak umožňuje danou úlohu řešit.
 - v logice se hovoří o problému rámce (axiomů)
- Chceme-li data a informace z reálného světa ukládat a reprezentovat co možná nejvíce plnohodnotně
 - musíme být schopni rozeznat více podstatné a méně podstatné
 - kvůli kooperaci a předávání výsledků řešení dané úlohy musíme být schopni dorozumět se nad takto vytvořeným modelem
 - chceme-li zpracovávat data strojově, musíme být schopni kognitivní model v potřebném rozsahu definovat na formální úrovni

Konceptualizace



- Potřeba popsat reálný stav (problém)
 - Proces uvažování, kdy hledáme dostatečně přesný, ale zároveň v rámci úlohy řešitelný popis reality
 - Hlavní krok při řešení úloh
 - Většina úloh není dostatečně definována,
 - tj. chybí dostatečný popis počáteční nebo koncového stavu, kritérium pro splnění řešení, efektivní algoritmus řešení
- Konceptualizace určuje, co bude v rámci řešení úlohy bráno v potaz
 - které objekty reálného světa
 - jaké vlastnosti budou uvažovány a které budou zanedbány
 - hledání vhodného (formálního?) způsobu řešení

Konceptualizace



- Těžké na procesu konceptualizace je hledání míry abstrakce (jistá míra abstrakce je nutná)
- Recept na dort:
 - Cukrář 1: 00 extra mouka, 4 křepelčí vejce, třtinový cukr, bio mléko, Couverture
 - Cukrář 2: hladká mouka, 2 slepičí vejce, řepkový cukr, plnotučné mléko, hořká čokoláda
 - Nelze stanovit koncept „recept“ na dort bez abstrakce surovin na vyšší úroveň (supertřídu) surovin
- Ale neřešíme množství (kromě vajec)
 - Co když je znalost o množství surovin tacitní – núnost provést měření
- Ale je to dost na definici pojmu “dort” prostřednictvím jeho složení?

Komunikace/sdílení



- Schopnost předávat (sdílet) data, informace a znalosti
- Sdílení dat
 - Schopnost fyzicky (digitálně) předat výsledek pozorování
- Sdílení informací
 - Schopnost předat data v kontextu jejich využití
 - Sdílení dat i jejich význam
- Sdílení znalostí
 - Učení, závislá na druhu znalosti z hlediska možnosti zachycení
 - Otázka úplnosti, přesnosti a důvěryhodnosti znalosti

Interoperabilita



- Interoperabilita (funkční)
 - "interoperability" - "styková provozuschopnost" (Anglicko-český výkladový slovník výpočetní techniky, SNTL Praha, 1990)
 - schopnost vzájemně si rozumět, vzájemně spolupracovat, dosáhnout vzájemné součinnosti
- Sémantická interoperabilita
 - schopnost správně pracovat i s významem přenášené informace, rozumět jí na úrovni formálně definovaných doménových konceptů (ISO TR 20514)
 - Je nezbytná pro možnost automatického zpracování dat



- Ve smyslu terminologický slovník
 - tedy slovník pojmů z dané oblasti
- **Glosář**
 - seznam nových nebo nejednoznačných pojmů s jejich definicí většinou v přirozeném jazyce
 - nejčastěji součástí odborných a výukových materiálů
 - cílem je informovat čtenáře, aby neutratil kontext textu
- **Tezaurus**
 - řízený slovník (strukturovaný popis) pojmů z dané oblasti
 - použití např. k anotování a indexování dokumentů
 - pojem definován v přirozeném jazyce
 - definuje nad množinou pojmů některé relace
 - existují standardy pro tezaurus, např. ISO 2788-1986

Pojem a termín



- Pojem je zobecnění objektů našeho zájmu a to:
 - Konkrétních předmětů (pacient, orgánů, histolog. struktury)
 - Vlastností (cyanóza, nauzea),
 - Dějů (biochemické vyšetření, progrese vyšetření)
 - Název a definice
- Pojem je dán současně abstrakcí a generalizací, kde pak:
 - Klasifikace - řeší vztahy mezi příbuznými pojmy
 - Definice - přesná formulace významu pojmu
 - Název (termín) - pojem pojmenovává
- Terminologie-souhrn všech názvů pro danou oblast (lékařský jazyk, synonyma)
 - snaha používat jediný výraz: synonyma

Ontologie



- Ve filozofii nauka o „bytí“
- V informatice
 - stanovení terminologie a jejího uspořádání v určité oblasti
 - formální reprezentace
 - umožňuje sdílení znalostí - umožňuje i mapování znalostí ze dvou různých terminologií
 - hierarchie založena na subsumpci (vztah is-a) – taxonomie
 - Může obsahovat speciální vztahy (vlastnosti)
 - Popisují obecné znalosti
 - rozdělení do tříd a jejich atributů a axiomů
 - jedinečné znalosti – instance jednotlivých tříd s nabytím hodnot u jednotlivých atributů – jsou zahrnuty do **báze znalostí**
 - k jedné ontologii může existovat více bází znalostí

Formalizace problému



- Deklarativní popis
 - Otázkou je „Co?“
 - Matematika (vzorec) a statistika
 - Klasifikace
 - Logická notace
 - Grafy, množiny
- Procedurální popis
 - Otázkou se „Jak?“
 - Algoritmizace
- Znalostní pravidla
 - Terapeutická, „dávkování“, diagnostická
 - Nezbytný symptom, postačující symptom, přispívající symptom

Klasifikace



- Přirazení objektů k třídě
 - na základě jejich charakterizace a parametrizace
- Na základě stanovení hierarchie
 - Vztahy souřadnost, nad/podřazenost, zjemnění, ortogonalita
 - hierarchie generická (rod, druh)
 - hierarchie partitivní (celek, část)
- Na základě podobnosti měřených parametrů
 - klasická klasifikace v strojovém učení
 - rozdělení na základě podobnosti objektů
 - model klasifikace se nazývá klasifikátor
 - Vzniklá struktura nemusí být (a většinou není) hierarchií, ale složením pozorovaných parametrů v N rozměrném prostoru (kde N je počet parametrů)

Reprezentace znalostí



- Pravidla
 - Deklarativní pravidla
 - stav, předpoklad \Rightarrow závěr
 - klasifikační pravidla
 - Procedurální pravidla
 - událost, situace \Rightarrow akce
 - procedurální schémata (sada produkčních pravidel)
- Hierarchicky uspořádaná pravidla
 - Acyklický graf
 - Reprezentace znalosti o uspořádání pravidel (pořadí jejich aplikace)
 - Možnost výskytu jednoho parametru ve více úrovních
 - Ve strojovém učení reprezentování rozhodovacím stromem, resp. klasifikačně-regresním stromem

Reprezentace znalostí



- Predikátová logika 1. řádu
 - predikát je obdoba „funkce“, která může nabýt hodnot pravda a nepravda
 - predikát P říká, že objekt (proměnná) x je vzděláním biomedicínský inženýr
 - predikát S říká, že objekt (proměnná) x je technicky vzdělaný
 - Jestliže x přiřadíme Michal Huptych, pak $P(x)$ je pravda
 - $P(x) \rightarrow S(x)$
 - predikátová logika definuje proměnné, funkce a konstant
 - formule pred. logiky vznikají z atomických formulí pomocí logických spojek a kvantifikátorů
 - konjunkce (\wedge), disjunkce (\vee), implikace (\rightarrow) a negace (\neg)

Reprezentace znalostí

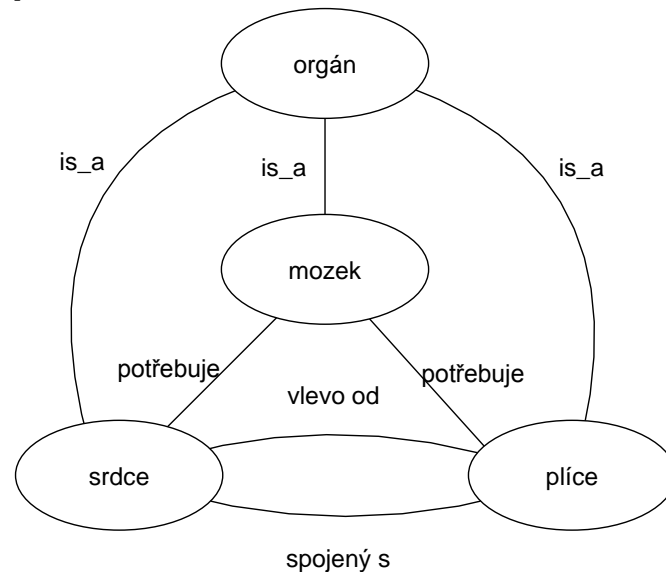


- Deskripční logika
 - Nejedná se o jeden formalismus, ale řadu příbuzných formalismů o různé schopnosti vyjádření (rozsahem výrazů)
 - Zachycení struktury třídy a relací mezi třídami podobně jako u ontologií, resp. rámců
 - Terminologická část (množina axiomů) (tzn. TBox)
 - definování nových konceptů a relací
 - obsahuje všechny inkluze ($C_1 \subseteq C_2$) a ekvivalence ($C_1 \equiv C_2$)
 - zobecněná terminologická znalost
 - Přiřazovací část (množina axiomů) (tzn. ABox)
 - přiřazování hodnot proměnným
 - obsahuje všechny axiomy $C(i)$ a $R(i_1, i_2)$
 - konkrétní relační struktura mezi objekty dané oblasti

Reprezentace znalostí



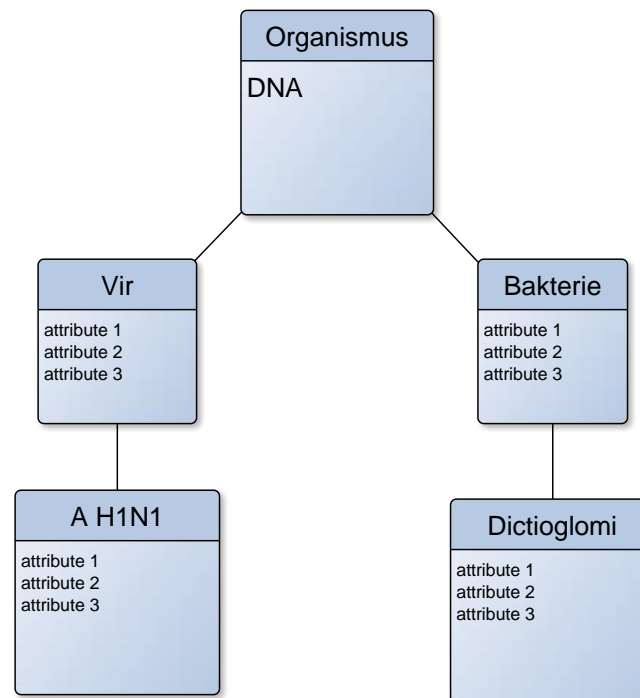
- Sémantické sítě
 - Navrženy v 60. letech
 - Reprezentace znalostí na základě grafu
 - Propojení uzlů (popisovaných objektů) hranami (vztahy mezi objekty)
 - Základní relace is-a (subsumpce), a_kind-of (hierarchie), part-of (kompozice)



Reprezentace znalostí



- Rámce
 - Navrženy Marvinem Minskym v 70. letech
 - Rámce lze propojovat podobně jako sémantické sítě (kind_of, part-of)
 - Základ pro objektový popis (reprezentaci v objektovém programování)



Shrnutí



- Je nutné uvědomit si, zda chceme pracovat s daty, informacemi nebo znalostmi.
- Každé řešení problému je zasazeno do určitého rámce/kontextu.
- Bez možnosti komunikace je sebelepší reprezentace neúčinná.
- Záleží jaká úroveň konceptualizace a formalizace je zvolena (nechodit s kanónem na komára).
- Reprezentace znalostí je otázka optimalizace vzhledem ke způsobu řešení.

Reference



V. Mařík, O. Štěpánková, J. Lažanský a kol.: *Umělá inteligence (1)*, Academia Praha, 1993 (použita kapitola 4, Z. Zdráhal: *Reprezentace znalostí*)

V. Mařík, O. Štěpánková, J. Lažanský a kol.: *Umělá inteligence (6)*, Academia Praha, 1993 (použita kapitola 4, Z. Zdráhal: *Reprezentace znalostí*)

P. Berka: podklady k přednáškám a cvičením z předmětu Principy inteligentních systémů, Dostupné na [https://sorry.vse.cz/~berka/4IZ430/] Kapitola Reprezentace znalostí na [https://sorry.vse.cz/~berka/docs/4iz430/P04-Reprezentace%20znalosti.pdf]