

METODY FORMALIZACE ZNALOSTÍ

METHODS OF KNOWLEDGE FORMALIZATION

Arnošt Veselý

Anotace :

V příspěvku jsou diskutovány standardní metody formalizace znalostí. Dále je stručně popsána dnes často v lékařské informatice používaná formalizace pomocí GLIF modelů. Znalost formalizovaná prostředky GLIF modelu a zakódovaná v XML může být automaticky zpracována na počítači. Ve spojení s databází může být použita pro podporu rozhodování nebo pro analýzu shody s všeobecně používanými nebo doporučenými postupy a to nejen v medicíně, ale také v jiných oblastech, jako je například veterinární medicína, pěstitelství nebo řízení biotechnologických procesů.

Klíčová slova:

data, znalosti, ontologie, GLIF model, lékařská doporučení

Abstract:

In the paper the standard methods of formalization of knowledge are discussed. Then the GLIF formalization technique currently used in medical informatics is briefly described. Knowledge formalized by means of in XML coded GLIF model can be automatically processed on computer. In connection with database it may be used for decision support or for checking of compliance with commonly used or recommended standards not only in medical science, but also in other fields, as there are for example veterinary medicine, agriculture or biotechnological process control.

Keywords:

data, knowledge, ontology, GLIF model, medical guidelines

ÚVOD

K orientaci a správnému rozhodování v dnešním složitém světě je nezbytné mít k dispozici spolehlivé a relevantní informace. Informace lze zhruba rozdělit na data a znalosti. Data lze chápat jako ověřitelná elementární fakta. Znalosti jsou pravidla pro vyvozování nových faktů z elementárních faktů (tj. dat) nebo z faktů, které již byly z dat odvozeny. Znalosti tedy slouží k inferenci nutně pravdivých (logicky pravdivých) nových skutečností.

Data a znalosti se vztahují k určité oblasti popisu skutečnosti (doméně). Pojmy, které se v této doméně používají musí být jednoznačně definovány. Jejich definice ve stávajícím odborném jazyce není v mnoha případech postačující. Je tomu tak proto, že dosud nejsou k dispozici inteligentní systémy, které by rozuměly tvrzením, které jsou zapsány v přirozeném jazyce a znalosti formulované v přirozeném jazyce nelze proto na počítačích automaticky zpracovávat. Sémantiku domény je proto třeba popsat takovými formálními prostředky, které automatické zpracování dovolí neboli je třeba provést formalizaci domény. Znalosti vyjádřené formálními prostředky pak jsou formalizované znalosti. Jako obecným formalizačním prostředkem může sloužit predikátový počet. Pokud je formalizovaná doména založena šířeji, tj. obsahuje i tzv.

common sense pojmy, hovoříme o formalizované ontologii nebo jednoduše o ontologii. Vytvoření ontologií pro jednotlivé obory dovoluje přirozené spojování domén ve větší formalizované celky. Nejvíce známými ontologiemi jsou CYC a UMLS.

Budování CYC ontologie probíhá již řadu let. Prostředkem formalizace je predikátový počet. Vlastnosti definovaných pojmů jsou formulovány prostřednictvím axiomů. Ontologie CYC je velmi rozsáhlá. Dnes již obsahuje miliony axiomů (Guha, Lenat, 1990).

Ontologie UMLS byla vytvořena pro oblast medicíny. Definuje společný terminologický slovník, který podporuje použití alternativní terminologie. K definici pojmů je použita metodika sémantických sítí.

METODY

Základním prostředkem, kterým lze doménu formalizovat, je predikátový počet. Formalizace domény v predikátovém počtu je však velmi rozsáhlá a náročná úloha. Proto se pro formalizaci používají jednodušší formalizační prostředky. Použití těchto prostředků klade určitá omezení, ale dovoluje rychleji dospět k použitelným výsledkům. Protože predikátový počet je obecný formalizační prostředek, lze formalizace provedené s použitím jednodušších formalizačních prostředků do jazyka predikátového počtu přeložit. Překlad ale může být značně obtížný.

Z alternativních prostředků používaných pro formalizaci zde můžeme uvést sémantické sítě, objektové modely a konceptuální závislosti.

Sémantické sítě dovolují jednoduchým a efektivním způsobem zachytit hierarchii pojmů. Osvědčily se především při budování rozsáhlých databází. Objektové modely se nejlépe hodí pro popis a realizaci složitě strukturovaných systémů jako jsou například podnikové informační systémy. Pro popis objektových modelů se používá jazyk OML. Konceptuální závislosti dovolují formalizovat některé základní pojmy běžného jazyka a zjednodušeně zachycují skladbu vět běžného jazyka. Lze je proto s výhodou použít tehdy, pokud má mít formalizovaný zápis těsný vztah k původnímu zápisu v běžném jazyku.

Praxe ukazuje, že v některých případech je použití standardních formalizačních metod obtížné. Proto se vyvíjejí další specializované formalizační prostředky, vhodné pro určitou doménu. Účelem tohoto příspěvku je na jeden z nich upozornit. Jedná se prostředek GLIF (Guideline Interchange Format), vyvinutý pro formalizaci lékařských doporučení. Tento prostředek je výsledkem spolupráce amerických universit a sice Columbia University, Harvard University, McGill University a Stanford University (Ohno-Machado, 1998), (Peleg M., 2000).

Lékařská doporučení definují standardní vyšetření pacienta, standardní postupy při stanovení diagnózy a volby způsobu léčení. Jsou formulována lékařskými společnostmi v běžném odborném jazyce, který je lékařskou komunitou používán. Výsledkem formalizace lékařských doporučení v prostředí GLIF je GLIF model těchto doporučení. GLIF model je orientovaný graf. Jeho uzly odpovídají akci lékaře (provedení určitého vyšetření nebo nasazení určité terapie), stavu pacienta nebo modelují závazná či doporučená rozhodnutí lékaře v závislosti na předem provedených vyšetření.

Na obr.1. je jako příklad uveden malý úsek GLIF modelu lékařských doporučení pro Nestabilní anginu pectoris (Ascherman, 2001). Uzly 1,3 4 a 8 jsou tzv. akční kroky modelu a popisují akce lékaře. Krok 6 modeluje závazné rozhodnutí lékaře, které má provést na základě vyhodnocení logické formule α . Uzly 9 a 10 představují doporučené akce a uzly 2 a 3 zavádějí do modelu paralelismus. Na tomto místě nutno říci, že závazné doporučení znamená doporučení, kterým se lékař má řídit pokud se rozhodl podle lékařských doporučení postupovat. Je ale na lékaři samotném, aby rozhodl, zda bude v daném konkrétním případě podle lékařských doporučení postupovat nebo zda právě léčený pacient představuje výjimku,

na kterou nelze lékařská doporučení vztáhnout. Konečné rozhodnutí je tedy na lékaři, neboť odpovědnost za správnou léčbu pacienta nese vždy ošetřující lékař.

Formalizace lékařských expertních znalostí spočívá ve vypracování GLIF modelu. GLIF model je obvykle vypracován ve spolupráci lékařů specialistů a informatik. Jakmile je lékařská expertní znalost reprezentována GLIF modelem, může s ní dále pracovat informatik bez účasti lékaře experta. GLIF model tedy tvoří rozhraní mezi lékaři a informatikou. GLIF model může být dále zakódován a automaticky zpracováván počítačem. K zakódování lze například použít jazyk XML.

GLIF model lékařských doporučení lze použít k následujícím účelům:

- ověření jednoznačnosti navrhovaných lékařských doporučení
- automatickému porovnávání lékařských doporučení používaných v různých institucích nebo zemích
- podpoře rozhodování lékaře (zobrazení aktuální části GLIF modelu představuje podporu rozhodnutí lékaře) (Peleska, 2003)
- automatické kontrole postupu lékaře (systém elektronického záznamu pacienta s přístupem k GLIF modelu může automaticky porovnávat lékařem zvolené varianty postupu a v případě, že se lékař odchýlí od postupu, který je podle lékařských doporučení závazný, upozornit jej na tuto skutečnost) (Veselý, 2004)

ZÁVĚR

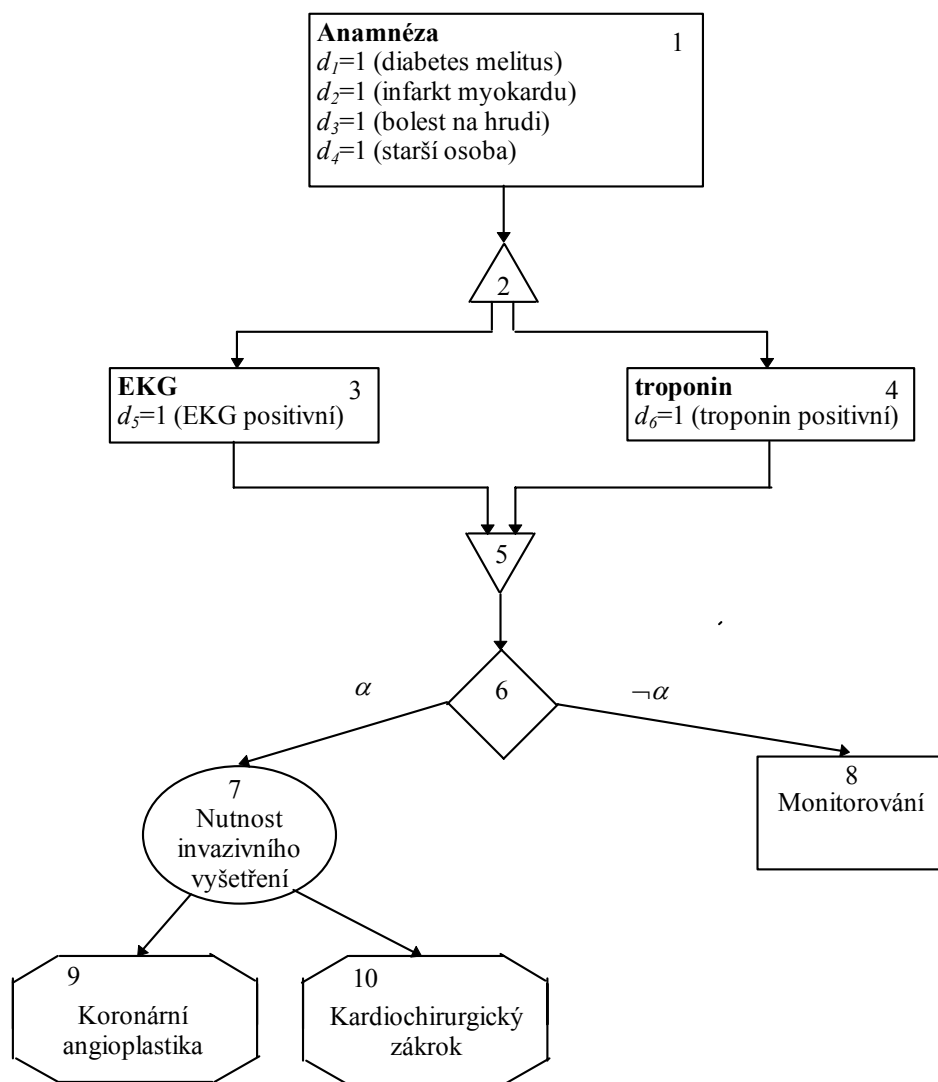
Ze stručného popisu GLIF modelu je zřejmé, že se jedná o prostředek formalizace, jehož použití není vázáno jen na oblast medicíny, ale může být použito pro formalizaci znalostí i v jiných znalostních doménách. Nabízí se například oblast veterinární medicíny, pěstivatelství nebo řízení biotechnologických procesů. Z oblasti medicínské informatiky je možné převzít metody, které již byly vyvinuty nebo na kterých se v současné době intenzivně pracuje. Jedná se především o metody, které těží z přímého navázání GLIF modelu na databázi. Byly navrženy a implementovány metody pro daty řízené zobrazování GLIF modelů, které mohou tvořit jádro systému pro podporu rozhodování. V současné době se také vyvíjejí metody pro automatické porovnávání aktuálního obsahu databáze s GLIF modelem, což umožní navrhovat automatické systémy pro sledování a kontrolu správnosti použitých postupů. V případě medicínské informatiky se bude jednat o sledování a kontrolu postupů lékaře. V jiných oblastech se může jednat o sledování a kontrolu postupů při řízení složitých systémů.

Literatura:

- Ascherman, M.: Doporučení k diagnostice a léčbě nestabilní anginy pectoris, 2001, <http://www.kardio-cz.cz>.
- Guha R.V., Lenat D.B.: Cyc: a midterm report, AI Magazine, 11 (3), 1990, pp. 32-59.
- Ohno-Machado L., Gennari, J. H., Murphy S.,N., Jain N.,L., Tu S., W., Oliver D., et al.: The GuideLine Interchange Format: A model for representing guidelines, Journal of the American Medical Informatics Association 1998, 5(4), pp. 357-372.
- Peleg M., Boxwala A.,A., et al.: GLIF3: The Evolution of Guideline Representation Format, In: <http://smi-web.stanford.edu/projects/intermed-web/guidelines>, 2000.
- Peleska J., Anger Z., Tomeckova M., Vesely A., Zvarova J.: Electronic form of the 1999 WHO/ISH Hypertension Guidelines, Thirteenth European Meeting on Hypertension , Milan, 2003.
- Vesely A., Zvarova J., Peleska J., Buchtela D., Anger Z.: Medical Guidelines Presentation and Comparing with Electronic Health Record, Proceedings of the International Joint Meeting, EuroMISE, Prague, April 2004.

Kontaktní adresa:

Doc. Ing. Arnošt Veselý, CSc., Katedra informačního inženýrství, PEF ČZU , Kamýcká 129, 165 21, Praha 6 – Suchbátka, tel. 224382356, e-mail: vesely@pef.czu.cz



Obr. 1: Malý a zjednodušený výsek GLIF modelu Lékařských doporučení pro Nestabilní anginu pectoris. Formule $\alpha = (d_1 \vee d_2 \vee d_3 \vee d_4) \wedge d_5 \wedge d_6$.