

Procedurální znalosti expertů a model GLIF

Ondřej Říha

Katedra informačního inženýrství, PEF ČZU v Praze
Kamýcká 129, 165 21 Praha 6 - Suchdol

ondrej.riha.praha@gmail.com

Abstrakt. Text příspěvku má za cíl seznámit čtenáře s problematikou expertních procedurálních znalostí. Tyto znalosti mohou být zaznamenány pomocí modelu GLIF, kde při práci s reálnými záznamy vznikají specifické problémy. Řešení problémů jako posouzení času v oborových doporučeních a porovnávání reálných záznamů s doporučeními v podobě nedeterministického modelu GLIF může znamenat zlepšení v procesech využívajících expertní znalosti. Komplikovanost porovnávání s modelem GLIF nevychází pouze z jeho nedeterministické formy. V praxi často nejsou dostupné kompletní záznamy provedených postupů, případně mohou být některé informace ztraceny při převodu na strukturovanou formu záznamu. Text představuje některé z možností řešení uvedených problémů jako je práce s časem v modelu GLIF a porovnávání nekompletních záznamů pomocí nahrazení chybějících hodnot případně kontrolu všech relevantních možností dle dostupných dat.

Klíčová slova: GLIF, oborová doporučení, EHR, nekompletní záznam

1 Úvod

Díky moderním komunikačním prostředkům jsou dnes informace v obrovské míře dostupné široké veřejnosti. Bez ohledu na kvalitu dostupných informací s nimi umíme efektivně pracovat. Mnohem složitější je to v případě záznamu znalosti. Experti při řízení složitých systémů používají tzv. procedurální znalosti, tedy znalosti často velmi komplexní, které reagují na průběžné změny a chování systému. Pro takové znalosti je obtížné definovat matematický model chování. Procedurální znalosti se používají například v humánní či veterinární medicíně, zemědělství, biotechnologii a dalších zejména pro diagnostiku a řešení určitých situací.

2 Metody

K formalizaci znalostí je dostupných mnoho prostředků od predikátového počtu, přes sémantické sítě, produkční systémy až k rámcům a mnoha dalším způsobům. Všechny tyto způsoby mohou být složitě aplikovatelné a nesnadné k porozumění expertům v jiných oblastech než jsou počítačové vědy. Procedurální znalosti lze formalizovat

*J. Steinberger, M. Zíma, D. Fiala, M. Dostal, M. Nykl (eds.)
Data a znalosti 2017, Plzeň, 5. - 6. října 2017, pp. 172-175.*

prostředkem, který je zápisem, respektive zobrazením bližší široké veřejnosti. Takovým prostředkem může být model GLIF (GuideLine Interchange Format). Tento model je silným formalizačním prostředkem pro jeho přehlednost a snadnou pochopitelnost [1].

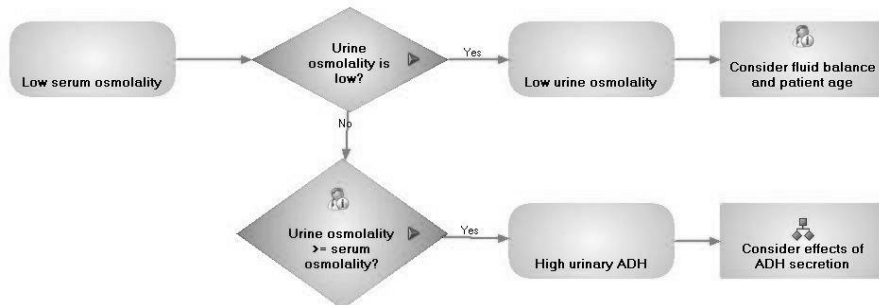
Cílem práce je nalézt vhodnou metodu zavedení časového faktoru do modelu GLIF a navrhnout postup srovnávání nekompletních záznamů o postupu expertů s nedeterministickým modelem GLIF.

2.1 Model GLIF

Model GLIF byl vyvinut ve spolupráci InterMed (Portland), Kolumbijské univerzity, Harvardské univerzity (Brigham and Women's Hospital a Massachusetts General Hospital) a Stanfordské univerzity [2]. GLIF model je procesně a objektově zaměřený a je reprezentován, jako orientovaný graf chronologicky udávající posloupnost úkonů. Samotné doporučení lze vytvářet ve specializovaných editorech nebo psát rovnou ve formě XML souboru (Extensible Markup Language), který je univerzální formát pro přenos informací. Současná verze modelu GLIF 3.5 používá OCL (Object Constraint Language) ve formě jazyka GELLO (“guideline expression language, object-oriented”). Model GLIF podporuje UMLS (Unified Medical Language System) [4].

GLIF model mapuje informace do tří vrstev abstrakce [3]

- jádro modelu GLIF
- RIM (Reference Information Model)
- lékařské znalosti



Obr. 1. Sublevel postupu léčby hyponatrémie [5]

Tento graf je skládán z pěti různých prvků, některé je možné vidět na obrázku.

Prvek ROZHODNUTÍ umožňuje několik subvariant [4].

- Utility Choice
- Weighted Choice
- Rule In Choice

Přes zdánlivou podobnost modelu s vývojovým diagramem je zde několik podstatných rozdílů a nejvýznamnější je ten, že model GLIF nemusí být deterministický.

3 Výsledky

Autor se ve své práci zabývá posouzením použití faktoru času v modelu, který je při záznamu některých procedurálních znalostí kritický a model GLIF nebyl přímo vybaven pro posouzení času v modelu jiným prostředkem, než prvkem *Rozhodnutí*, který může být v záznamech komplexních postupů značnou nevýhodou pro svoji schopnost zhoršit přehlednost modelu. Řešením může být rozšíření specifikace modelu GLIF o časový parametr či pod jednotnou ontologií definovat časové platnosti jednotlivých stavů systému a ty následně opakovaně používat bez nutnosti je v postupu pokaždé znovu určovat.

Dále je řešena otázka srovnávání existujících záznamů postupů expertů s modelem, který představuje doporučení v konkrétní oblasti užití. Zde je největší výzvou porovnávání záznamů s nedeterministickým modelem GLIF a porovnávání nekompletních záznamů.

Pro srovnání oborových doporučení s nekompletním záznamem autor zvážil užití tří metod.

- Empty metoda
- Imputace hodnot
- Komplexní varianta

Empty varianta zahrnuje triviální doplnění chybějící hodnoty prázdným znakem.

Pro imputaci hodnot bylo zváženo několik metod a zdrojů dat pro odhad chybějící hodnoty. Stejně tak bylo posuzováno, jaké hodnoty je možné doplnit a u jakých je metoda imputace nepřijatelná. Výsledkem aplikace metody imputace bylo také zpřesnění dělení prvků v modelu GLIF.

Komplexní metoda představuje úplné prohledání všech dostupných variant průchodu modelem vzhledem k dostupnosti záznamu a posouzení zda mezi několika finálními stavy je alespoň jeden, který odpovídá doporučenému postupu.

4 Diskuze a závěr

Po zvážení několika možností zavedení času do modelu GLIF, bylo na dostupných oborových doporučeních zjištěno, že vhodnost rozšíření specifikace modelu je vhodná až od určité složitosti modelu.

Pro porovnávání záznamů a doporučení je nutná unifikace strukturálních záznamů a oborových doporučení pod jednotnou ontologií. Právě použití jednotné ontologie umožňuje řešit problematiku času v modelu definicí časových platností jednotlivých parametrů nutných pro rozhodovací kroky modelu.

Empty metoda a imputace chybějících hodnot se prokázaly jako metody s velice omezenou možností použití a s výsledkem mající pouze pravděpodobnostní charakter.

Komplexní metoda je schopna zjistit s naprostou jistotou některé chyby v postupu expertů i v případě nekompletních záznamů. Při nekompletním záznamu ovšem nelze určit, zda postup byl zcela správný, ale pouze nalézt záznam, který neodpovídá doporučenému postupu.

Tato metoda může být po dalším zpřesnění využita pro hledání chyb například v postupu léčby pacientů, kde je vhodnost nalezení závažných pochybení obzvláště vhodná.

Literatura

Reference na články v časopisech:

1. Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, et al. Comparing computer-interpretable guideline models: a case-study approach. *Journal of the American Medical Informatics Association*, roč. 10, č. 1, s. 52-68, led. 2003.
2. L. Ohno-Machado et al., „The GuideLine Interchange Format: A Model for Representing Guidelines“, *J. Am. Med. Inform. Assoc. JAMIA*, roč. 5, č. 4, s. 357, srp. 1998.
3. A. A. Boxwala et al., „GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines“, *J. Biomed. Inform.*, roč. 37, č. 3, s. 147–161, čer. 2004.

Reference na technickou specifikaci:

4. M. Peleg, A. Boxwala, S. Tu, D. Wang, O. Ogunyemi, a Q. Zeng, „Guideline Interchange Format 3.5 Technical Specification“. *InterMed Collaboratory*, 04-kvě-2004.
5. Ch. Appleton, *QML Pathology document Investigation of Hyponatraemia*, Medical-Objects GLIF Editor, pros. 2016

Annotation:

Experts procedural knowledge and model GLIF

The text aims to acquaint the reader with the issue of expert procedural knowledge. This knowledge can be recorded by using of the model GLIF, where when working with real records arise specific problems. Address issues such as the assessment of time in the industry recommendations, and comparison of the real records with the recommendations in the form of non-deterministic of the model GLIF may be an improvement in the processes using expert knowledge. In practice, often not available complete records of the executed procedures, where appropriate, may be some information lost when converting to the structured form of the record. The text presents some of the options to address these problems such as work with time in the model GLIF and comparing the incomplete records using the replace missing values or control of all relevant possibilities according to the available data.