

Víceková formalizace obsahu lékařských doporučených postupů

Svátek Vojtěch, Růžička Marek, EuroMISE - Kardio, VŠE Praha

Pracovní text, 25.9.2002. Zkrácená verze publikována v časopisu Lékař a technika 4/2003.

Abstrakt

Převod textu lékařských doporučených postupů do formální podoby využitelné pro strojové usuzování je velmi náročný. K jeho usnadnění může sloužit rozklad celého procesu do několika explicitních kroků. Vyvinutá metodika a interaktivní softwarový nástroj, využívající značkovací jazyk XML a soustavy transformačních pravidel, umožňují formalizovat dokumenty přehledným způsobem, a odstraňují významnou část rutinní práce. Zúčastněný lékařský expert se pak může soustředit na odbornou stránku formalizace. Podrobné anotování textu je přínosné zejména v případech, kdy je nutno zachovat konsensuální charakter znalostí a vyvarovat se vnesení subjektivních interpretací; taková situace nastává např. pokud je cílem zpětná analýza shody léčby s obsahem (formalizovaných) doporučených postupů. Víceková metodika postupuje od značkování povrchové struktury dokumentu přes specifikaci sémantiky rozsáhlejších bloků textu, doplňování podrobné struktury obsažených znalostí (včetně implicitních), a jejich restrukturaci, do podoby znalostní báze, až k prototypové implementaci ve spustitelném kódu, který může být vygenerován do značné míry automaticky. Přístup je v současnosti ověřován na doporučených postupech z oblasti kardiologie.

Úvod

Využívání lékařských doporučených postupů (LDP) jako prostředku pro komunikaci lékařských znalostí a pro prosazování standardu zdravotní péče je v posledních letech velmi diskutovaným tématem v celosvětovém i tuzemském měřítku. Vedle distribuce LDP v textové (ať už tištěné, nebo elektronické) podobě se stále větší význam přikládá možnosti zpracovávat vhodně formalizované LDP pomocí počítače. Článek se věnuje problematice převodu textového obsahu LDP do podoby formálního jazyka zpracovatelného specifickou počítačovou aplikací (odvozovacím programem).

Role konceptuálního modelu a textu v procesu formalizace

Formalizace znalostí obsažených v textu je vedle získávání znalostí přímo od experta na danou problematiku (tzv. elicítace) jednou ze základních úloh znalostního inženýrství. Několik desetiletí vývoje této disciplíny ovšem ukázalo, že přechod od znalostí formulovaných přirozeným jazykem (ať už v písemné nebo ústní podobě) k zápisu ve formálním jazyce umožňujícím strojové odvozování nemůže dost dobře být přímočarý, s ohledem na syntaktickou i sémantickou odlišnost obou typů jazyků. Roli zprostředkovatele standardně hraje *abstraktní model*, který je na obou konkrétních způsobech vyjádření znalostí relativně nezávislý; v posledních letech se pro takový typ modelu ujal (z filosofie převzatý) termín *ontologie*. V případě LDP může ontologie zahrnovat pojmy jako je „procedura“, „rozhodování“, „klinický parametr“, „stav pacienta“ apod. Jednotlivé dílčí znalosti je pak nutno:

- identifikovat v textu (tj. odlišit od částí textu, které pro daný účel významnou informaci nenesou)
- přiřadit abstraktním prvkům ontologie (jako jejich tzv. instance)
- převést do formální podoby s využitím předpřipravené struktury prvku definované v ontologii.

V posledních letech se výzkumné úsilí v oblasti formalizace LDP rozdělilo na dva proudy. První z nich bere jako východisko formalizace právě *model*: hlavní činnost lékařského experta ve spolupráci se znalostním inženýrem spočívá v naplňování předdefinované struktury modelu formalizovanými znalostmi abstrahovanými z příslušných pasáží textu. Druhý naopak přiznává ústřední roli samotnému *textu*: jednotlivé pasáže jsou (rovněž ve spolupráci lékařského experta a znalostního inženýra) systematicky převáděny do formální podoby. Jako softwarová podpora se v prvním případě uplatňují zejména *graficky orientované editory* (s důrazem na vývojové diagramy aktivit, rozhodovací stromy apod.), ve druhém případě spíše *značkovací editory* (na bázi jazyka XML) sloužící pro sémantické značkování fragmentů textu.

Oba přístupy mají své výhody a nevýhody, které vystupují do popředí v souvislosti s konkrétními úlohami. Formalizace *řízená modelem* umožňuje poměrně rychle postoupit od neformální analýzy textu k tvorbě formální báze znalostí, zahrnující často jen určitý výsek problematiky řešené v LDP. Nese však s sebou riziko, že formální podoba LDP sémanticky neodpovídá podobě textové. Aby totiž bylo s pomocí formálního modelu možné úspěšně odvozovat, je nutné do něj dodat dva typy informací, které v původním textu obsaženy nejsou. První z nich jsou *implicitní předpoklady*, zpravidla odpovídající „učebnicovým“ znalostem (např. „o esenciální hypertenzi se jedná, pokud u pacienta není diagnostikována choroba vyvolávající nárůst tlaku sekundárně“). Druhý typ chybějící informace je pro laika lépe patrný, avšak jeho doplnění o to obtížnější: jde o *vágní formulace*, jako je např. „léčba má dobrou odezvu“ či „měření je nutno opakovat několikrát v krátkých intervalech“. Vágnost bývá zapříčiněna snahou ponechat lékaři jistou volnost v rozhodování, někdy jde ovšem přímo o kontroverzní problematiku, o které nebyla při tvorbě LDP dosažena shoda. Lékař pak při tvorbě formální reprezentace často použije takovou interpretaci, na jakou je z praxe zvyklý, aniž si třeba i uvědomí, že tím obsah LDP *subjektivizuje*.

Formalizace *řízená textem* tento problém alespoň částečně eliminuje tím, že každé doplnění informace je možné v označeném textu výslovně zdokumentovat. Na druhé straně je časově i intelektuálně náročnější – jednak proto, že víceméně předpokládá zpracování celého textu „slovo za slovem“, jednak v důsledku velkého *syntaktického i sémantického odstupu* mezi strukturou přirozeného jazyka a strukturou počítačem zpracovatelné reprezentace. Tento odstup je obvykle překonáván postupně, v několika iteracích, i tak však textově orientovaná formalizace inklinuje k tvorbě modelů spíše „pasivních“, tj. jen částečně využitelných pro odvozování. Důraz je kladen na zachycení nejen základních diagnostických a terapeutických „algoritmů“, ale i celkového klinického kontextu, včetně „meta- znalostí“ o aplikovatelnosti LDP samotných.

Přístupy orientované na model ve světě počtem převládají; mezi nejvýznamnější patří např. projekty EON [9], Asgaard [17], GLIF [10], Guide [13] či Prodigy [5]. Hlavním představitelem textově orientovaného směru je naopak GEM [20]; můžeme předeslat, že do téže skupiny patří i přístup popisovaný v tomto článku.

Pro jistotu ještě upřesněme, že oba směry se shodují na nutnosti existence konceptuálního modelu, ze kterého vychází v prvním případě grafické znázornění v interaktivním editoru, ve druhém případě struktura elementů značkovacího jazyka¹; k problematice volby konceptuálního modelu pro naše účely se vrátíme později. Naopak některé modelově orientované přístupy (EON, Asgaard) v poslední době usilují o zachování vazby prvků formalizované znalostní báze na odpovídající místa původního textu, viz např. [6], [19], [19].

¹ U textově orientovaných přístupů se při formalizaci nepracuje s modelem samotným (jako se svého druhu „logickou teorií“), nýbrž jen s jeho podstatně jednodušší stromovou „serializací“.

Účel formalizace – znalostní úlohy nad LDP

Důvodů, proč LDP formalizovat, může být celá řada. Uvedme si ty nejčastěji zmiňované:

- Při formalizaci se odhalí řada dílčích nedostatků, které mohou při neformálním zkoumání textu uniknout. Nejčastěji jde o nekonzistenci, neúplnost či nadbytečnost definic pojmů či doporučených postupů, viz [7]. Formalizace může být tedy užitečná i tehdy, když nám nejde o automatické zpracování LDP, nýbrž jen o *zpřesnění jejich textové podoby*. Cílovou skupinou zde budou tvůrci textových LDP, tedy zejména národní a mezinárodní lékařské společnosti.
- Klasickou oblastí aplikace již formalizovaných LDP je *podpora rozhodování* lékaře, a to jednak při stanovení, zda jsou daná LDP vůbec způsobilá k použití v konkrétním případě (tzv. „eligibility“), jednak při vlastním diagnostickém a/nebo terapeutickém postupu, který je předmětem LDP.
- LDP též sehrávají významnou roli při *edukaci* (zvl. začínajících) lékařů, tím, že jim umožňují otestovat si vlastní představu o postupu (na hypotetických příkladech), a upozorní je na zásadní chyby dříve, než by se projevíly v klinické praxi. Příklad formalizovaných LDP (pro problematiku hypertenze) s převážně edukativním zaměřením je popsán v [12].
- Poslední charakteristickou úlohou, kterou zde uvedeme, je *zpětná analýza shody* léčby (zachycené prostřednictvím elektronického záznamu o nemocném - EZN) s obsahem LDP. V anglické literatuře se zpravidla označuje jako „compliance analysis“; někdy se používá i termín „adherence“² [3].

V tomto článku budeme věnovat specifickou pozornost právě *zpětné analýze shody* - tato úloha má totiž vlastnosti, které podle našeho názoru potlačují nevýhody a naopak vyzdvihují výhody *textově orientované* formalizace LDP oproti modelově orientované:

- Ve srovnání s jinými úlohami je přidání *subjektivních znalostí* (které má textově orientovaný přístup lepší šanci eliminovat) obzvláště nežádoucí, protože nám primárně jde o porovnání EZN s původními (a nikoliv „vylepšenými“) LD.
- Náročnost systematické analýzy textu může být kompenzována *hromadnou aplikací* výsledné formální verze LDP na velký počet EZN dlouhodobě uchovávaných v databázích³. To platí zejména pro diagnózy s vysokou mírou výskytu v populaci. Cílem analýzy pak není vyhodnotit kvalitu léčby jednotlivého pacienta, nýbrž posoudit celkovou míru dodržování LDP (resp. jejich konkrétních částí) lékaři na daném pracovišti či v daném regionu.
- Nalezené *vzory neshod* (s vysokou četností na daném vzorku dat) je vhodné porovnat s původním textem, což textově orientovaná formalizace dobře podporuje. Teprve na základě prozkoumání přesného znění textu je možné posoudit, zda jde o systematickou chybu léčby, vědomou odchylku vynucenou lokálními specifiky, chybu v původním textu LDP, nebo kupříkladu nepřesnost vzniklou při samotné formalizaci.

Metodika víceetapové formalizace

Pro naše účely (i s ohledem na významnou roli zpětné analýzy shody) jsme se rozhodli pro textově orientovaný přístup k formalizaci LDP. Od začátku bylo zřejmé, že vyvíjená metoda musí zohlednit celou řadu aspektů

² Ten by ovšem, např. podle [16], měl být spíše rezervován pro *teoretickou* shodu lékaře s obsahem LDP. Ta se, jak ukazuje citovaný článek, nemusí vždy „proměnit ve skutek“, tj. odrážet se v *empirických* datech o léčbě.

³ Uplatní se zde techniky známé z oblasti *dolování dat*, viz [15].

přechodu od textového dokumentu k plně formální reprezentaci, představujících jádro výše zmíněného syntaktického a sémantického odstupu. Zejména pak:

- *Obecné jazykové výrazy* vyjadřující např. strukturu definic, rozhodovacích procesů či kauzálních vztahů musí být nahrazeny *standardizovanými formálními strukturami*.
- *Specifické jazykové výrazy* odpovídající odborným lékařským pojmům musí být převedeny na jednoznačné *identifikace* těchto pojmů, včetně vyřešení problému synonymie.
- Jak již bylo řečeno výše, chybějící *implicitní znalosti* musí být doplněny a *vágní formulace* zpřesněny.
- Prvky znalostí musí být *modularizovány*, tj. učiněny nezávislými na okolním kontextu dokumentu.

Je zřejmé, že textově orientovaná formalizace má (jak uvádí i [20]) vzhledem ke své náročnosti nutně víceřadový, iterativní charakter. Dosud známé přístupy ponechávaly členění těchto kroků na *ad hoc* rozhodnutí znalostního inženýra a lékařského experta. Pro urychlení jejich práce a usnadnění její softwarové podpory je ovšem výhodnější obvyklé členění kroků *explicitně předdefinovat*. To by mělo přinést zejména následující výhody:

- Celý proces se díky rozložení do jednodušších transformačních kroků stane *lépe kontrolovatelnějším*. Případné doplňování *subjektivních znalostí* expertem může být snadněji zdokumentováno, a riziko *ztráty informace* se naopak sníží. Navíc v případě zjištění *chybné formalizace* stačí vrátit se zpět jen k předchozí úrovni a ne až k původnímu (nestrukturovanému) dokumentu. To vše má zásadní význam i pro koncepci softwarové podpory, o které bude řeč dále.
- Některé kroky nebudou vyžadovat *účast experta*, nebo jen v malém rozsahu. Rozložení požadavků na odbornost do různých fází procesu se týká nejen lékařského experta, ale také specialisty na cílový jazyk *implementované* znalostní báze (který nemusí být totožný se znalostním inženýrem zaměřeným na konceptuální modelování LDP), ev. tvůrcem čistě *formální struktury* vstupního elektronického dokumentu.

Víceřadový proces formalizace vychází z původního textu, a každá následující úroveň zpracování rekurentně slouží jako základ pro další zpracování. Přestože je věcná podstata fází formalizace v principu nezávislá na používaném formálním jazyce, přijmeme již zde předpoklad, že každé úrovni odpovídá *specifický značkovací jazyk* založený na XML (*eXtensible Mark-up Language*). Před začátkem samotného zpracování textu je proto potřeba mít pro každou úroveň (s výjimkou první a poslední) vytvořen příslušný soubor DTD (*Definice Typu Dokumentu*). Daná posloupnost souborů DTD pak může být samozřejmě opakovaně používána pro více dokumentů téhož typu (např. různé LDP).

Prozatím jsme pro formalizaci LDP zavedli následující úrovně:

- **Vstupní textový formát.** Výchozím bodem formalizace je pro nás dokument s definovanou *syntaktickou strukturou*. Výhodné je použití standardu XHTML [2], který je zobrazitelný běžnými webovými prohlížeči podobně jako HTML, a současně zpracovatelný nástroji určenými pro XML. Velký počet existujících LDP je již dostupný prostřednictvím WWW právě ve formátu HTML, ze kterého je lze do XHTML automaticky převést. Pro převedení prostého textu do XHTML zase postačí kvalifikace návrháře webových stránek.
- **Základní sémantická struktura.** V první fázi formalizace se odstraní syntaktické formátování, a naopak se označí, někdy relativně rozsáhlé, bloky textu (ev. tabulky npod.), které obsahují určitý typ znalosti. Tato úroveň může být zpravidla bez větších problémů vyhotovena znalostním inženýrem (obeznámeným s podstatou hlavních doménových pojmů), který pouze v případech závažných nejasností kontaktuje lékařského experta.

- **Podrobná sémantická struktura.** V této fázi se základní bloky rozčlení do detailní struktury subelementů. Původní tok textu se tím přeruší, jeho jednotlivé části však zůstanou v původním pořadí. Na této fázi formalizace se již významným způsobem podílí lékařský expert, který doplňuje chybějící informace. Typ a míra jeho zásahů je zaznamenávána přímo v dokumentu, přesněji řečeno, ve specifických attributech elementů XML, kterých se zásah týkal.
- **Univerzální znalostní báze.** Dalším krokem je systematické přeuspořádání elementů, které se stávají nezávislými na kontextu uvedení v dokumentu. To vyžaduje další doplnění informací. V této fázi se rovněž odstraní případné duplicity informací, a související elementy se propojí křížovými odkazy. Pojmy odpovídající strukturovaným datům (diagnózy, testy apod.) se soustředí do *datového slovníku*. Vedle toho mohou vzniknout další struktury nad rámec původních znalostních bloků, vyjadřující např. hierarchie nebo sekvence. Řada dílčích činností má technologický charakter a může být vykonávána znalostním inženýrem (softwarová podpora je zde ovšem prakticky nezbytná!), v některých případech je však nutná účast lékařského experta.
- **Znalostní báze specifická pro export.** Poslední úroveň značkování XML již zohledňuje skutečnost, že vlastní odvozování nad znalostmi z LDN bude probíhat v jiném prostředí. Struktura je proto upravena tak, aby následný export do cílového jazyka byl co nejsnazší. Lze předpokládat, že tato univerzální znalostní báze může být používána pro různé cílové jazyky právě díky různým variantám specifické „meziúrovně“. Přitom stále ještě zůstáváme na syntaktické platformě XML.
- **Cílová výpočetní reprezentace.** Výsledný formát může být buď kódem některého jazyka způsobilého pro logické odvozování (v našem projektu uvažujeme OCML [8] a Prolog), nebo i běžného programovacího jazyka (v našem případě struktura tříd v jazyce Java). Export může být s výhodou proveden prostřednictvím stylového jazyka XSLT [4], běžně používaného pro transformace dokumentů XML do různých formátů.

Softwarová podpora vícekrokové formalizace

Jak bylo zmíněno výše, vícekroková formalizace není dost dobře myslitelná bez podpory interaktivním softwarovým nástrojem. Proto byl experimentálně vyvinut program pojmenovaný *Stepper*.

Práce s programem *Stepper* je v zásadě rozdělena do tří částí. První část, představující přípravu před samotnou formalizací, zahrnuje vyhotovení sady souborů DTD jednotlivých úrovní pro určitý typ dokumentu (např. LDP). Ke každé úrovni je potřeba dále blíže specifikovat vnitřní *závislosti* mezi elementy a atributy. Obdobně je nutné definovat pravidla pro přípustné *transformace elementů* z jedné úrovně do druhé. Pro *pravidla závislosti* i pro *transformační pravidla* nabízí *Stepper* vlastní editory. Přípravu pravidel vykonává erudovaný znalostní inženýr, a to pouze jednorázově. Pravidla jsou nezávislá na vstupních LDP.

Poté již může s aplikací pracovat i uživatel, který není s konceptuálním modelem detailně obeznámen, a chce formalizovat jednotlivé dokumenty odpovídajícího typu. V prvním kroku *vyznačí* základní bloky textu, přiřadí jim některý ze znalostních elementů (obr. 1) na základě aplikovatelných transformačních pravidel (ty jsou reprezentovány automaticky generovanými tlačítky v pravé horní části), a případně vyplní hodnoty atributů těchto elementů. V následujících krocích formalizace pak pracuje se stromovou reprezentací znalostních elementů, kde horní část obrazovky vždy odpovídajícím zdrojové a dolní část cílové úrovni daného kroku (obr. 2). Opět v závislosti na transformačních pravidlech *vybere* nástupníka (nástupníky) znalostního elementu a částečně *doplní* jeho vnitřní strukturu a hodnoty atributů. Rutinní část jeho práce je ovšem odstraněna díky

pravidlům, která se při vytváření nových elementů aktivují a některé hodnoty buď přímo vyplní, nebo pro ně nabídnou množinu alternativ konzistentní se zbytkem dokumentu.

Po dokončení formalizace najde uplatnění poslední skupina funkcí věnovaná nahlížení „napříč“ formalizačním procesem a případným dodatečným úpravám. Hlavní předností těchto funkcí je názorná grafická reprezentace vývoje a vzniku jednotlivých znalostních elementů. Tímto způsobem lze přehledně hledat vazby mezi dokumentem a formálním modelem, které jsou ze samotného zápisu ve formátu XML jen těžko patrné.

Navíc *Stepper* využívá některé standardní nástroje spojené s XML, které ovšem uživateli předkládá pouze skrze grafické rozhraní, a jejich skutečná, mnohdy nesrozumitelná podoba se tak stává neviditelnou. Funkce *Stepperu* lze tedy shrnout do následujících bodů:

- Podpora *vyznačování* znalostí ve zdrojovém textu, a *doplňování hodnot* jejich atributů.
- Plně automatická tvorba a aktualizace *odkazů* mezi úseky textu a znalostními elementy, napříč úrovněmi formalizace.
- Export znalostí *do spustitelné aplikace* pomocí XSLT.
- Uživatelsky přívětivý editor *transformačních pravidel* i *pravidel závislosti*
- *Vyhledávání* znalostí vzniklých ze zadaného úseku textu, nebo naopak textu, ze kterého vznikly zadané znalostní elementy.
- Zakrytí syntaxe XML i ostatních nástrojů spojených s XML (XLink, XPointer, XSLT).

Fungování programu i sémantika transformačních pravidel jsou podrobně (včetně příkladů) popsány v [14].

Volba konceptuálního modelu

Konceptuální model představuje, na nejvyšší úrovni abstrakce, jakýsi „zorný úhel“, přes který se díváme na problematiku popisovanou v daném LDP. Nedávno zveřejněná studie [11] zobecňuje různorodé spektrum používaných modelů LDP na čtyři hlavní typy, jejichž jádrem jsou postupně: modely *řešení úloh* (EON [9]), *skeletální plány* (Asgaard [17]), modely *workflow* (Guide [13]), a *stavy pacienta* s odpovídajícími *scénáři* (Prodigy [5]). Poslední z nich se od ostatních odlišuje větší mírou modularity: jednotlivé scénáře představují do jisté míry nezávislé jednotky, které se s ostatními propojují teprve druhotně. To na jedné straně diskvalifikuje použití některých vysoce sofistikovaných obecných metod (např. temporálního) odvozování, na druhé straně však vede ke snazšímu a přehlednějšímu vyhodnocování typických situací vzniklých při léčbě, což zřejmě souvisí s určením nástrojů Prodigy pro hromadné využívání praktickými lékaři při léčbě častých, spíše nefatálních chorob.

Model, ke kterému jsme nezávisle dospěli (viz [22]), vykazuje řadu podobností právě s Prodigy, a to i včetně používané terminologie. Klíčovým prvkem jsou *scénáře*, vyjadřující vztah mezi *stavy* pacienta (včetně událostí z historie léčby) a *rozhodnutími* či *akcemi*, které by měl lékař provést. Klinické parametry i další potřebné pojmy jsou popsány pomocí *definic pojmů*. Doplnkový charakter z hlediska možností odvozování mají *cíle* léčby (pro celé LDP nebo jednotlivé činnosti) a *kauzální vztahy*. Poslední částí modelu je sekundární *grafová struktura* nad scénáři, která by měla do jisté míry suplovat provázání akcí a rozhodnutí hrající podstatnou roli ve většině ostatních modelů.

Námi vyvinutý model je oproti většině modelů používaných ve světě konceptuálně jednodušší, a pro některé účely by mohl být nedostačující (neobsahuje např. dekompozici LDP na dílčí moduly). Chápeme ho proto spíše jako dočasný prostředek pro ověřování funkčnosti nástrojů pro formalizaci i pro zpětnou analýzu shody. Pro

budoucí formalizaci LDP směřující k rozsáhlejšímu klinickému využívání počítáme spíše s adaptací existujícího, bohatšího modelu Prodigy (nebo jemu podobného).

Aplikace

Metodiku víceřadkové formalizace jsme ověřovali na LDP pro léčbu hypertenze [1] vydaných organizací WHO, v rámci mezinárodního projektu „Medical Guideline Technology“, v letech 2000-2001. Detailní vyznačení znalostí v dokumentu se stalo základem aplikace pro zpětnou analýzu shody [15], vyvinuté v jazyce OCML. Vzhledem k tomu, že v té době ještě neexistovala softwarová podpora (první verze programu Stepper byla zprovozněna až koncem roku 2001), musela se ovšem formalizace uskutečňovat ručně, a nebyla plně dotažena do pokročilých úrovní. Aplikace proto nebyla generována automaticky a její vazba na sémanticky označovaný text je pouze nepřímá.

V rámci spolupráce partnerských organizací EuroMISE Centra – Kardio (Ústavu informatiky AV ČR a Vysoké školy ekonomické v Praze) probíhá v současnosti nová fáze formalizačních experimentů, které se zaměřují jednak opět na problematiku hypertenze, jednak na LDP pro nestabilní anginu pectoris. Možnost dotažení víceřadkového procesu formalizace (s využitím programu Stepper) až do poslední fáze – exportu do spustitelné aplikace – byla ověřena na části textu hypertenzních LDP [1] týkající se splnění cíle léčby ve vazbě na některé rizikové faktory. Ukázka rozhraní demonstrační „aplikace“ automaticky vygenerované v jazyce Java je na obr. 3. Vedle toho je v přípravě export znalostí o léčbě hypertenze do prototypového programu pro zpětnou analýzu shody, vyvíjeného v jazyce Prolog.

Závěr

Popsaná metodika a softwarový nástroj pro víceřadkovou formalizaci představuje možnost, jak dospět od textové podoby LDP k jejich formální, počítačem zpracovatelné podobě, s minimálním rizikem nechtěné ztráty či subjektivizace obsažených znalostí. Samotná textově orientovaná formalizace tím sice nepřestává vyžadovat značné úsilí ze strany lékařů a informatiků, díky jasné koncepci a softwarové podpoře se však stává podstatně lépe zvládnutelnou. Vložené úsilí se, podle našich předpokladů, vyplatí zejména v případě aplikací zaměřených na hromadnou, zpětnou analýzu shody LDP s reálným postupem léčby zachyceným v EZN.

Metodika i nástroj mají prozatím daleko do dokonalosti. Nejvýznamnější nedostatek ve srovnání s obdobnými přístupy, kterého jsme si vědomi, je absence propojení s *terminologickými zdroji* typu ICD-10, SNOMED apod., které by umožnily systematicky převádět *ad hoc* lékařské termíny použité v textu na termíny standardizované. Již stávající DTD však s takovým propojením do jisté míry počítají, a jeho realizace je převážně programátorskou záležitostí.

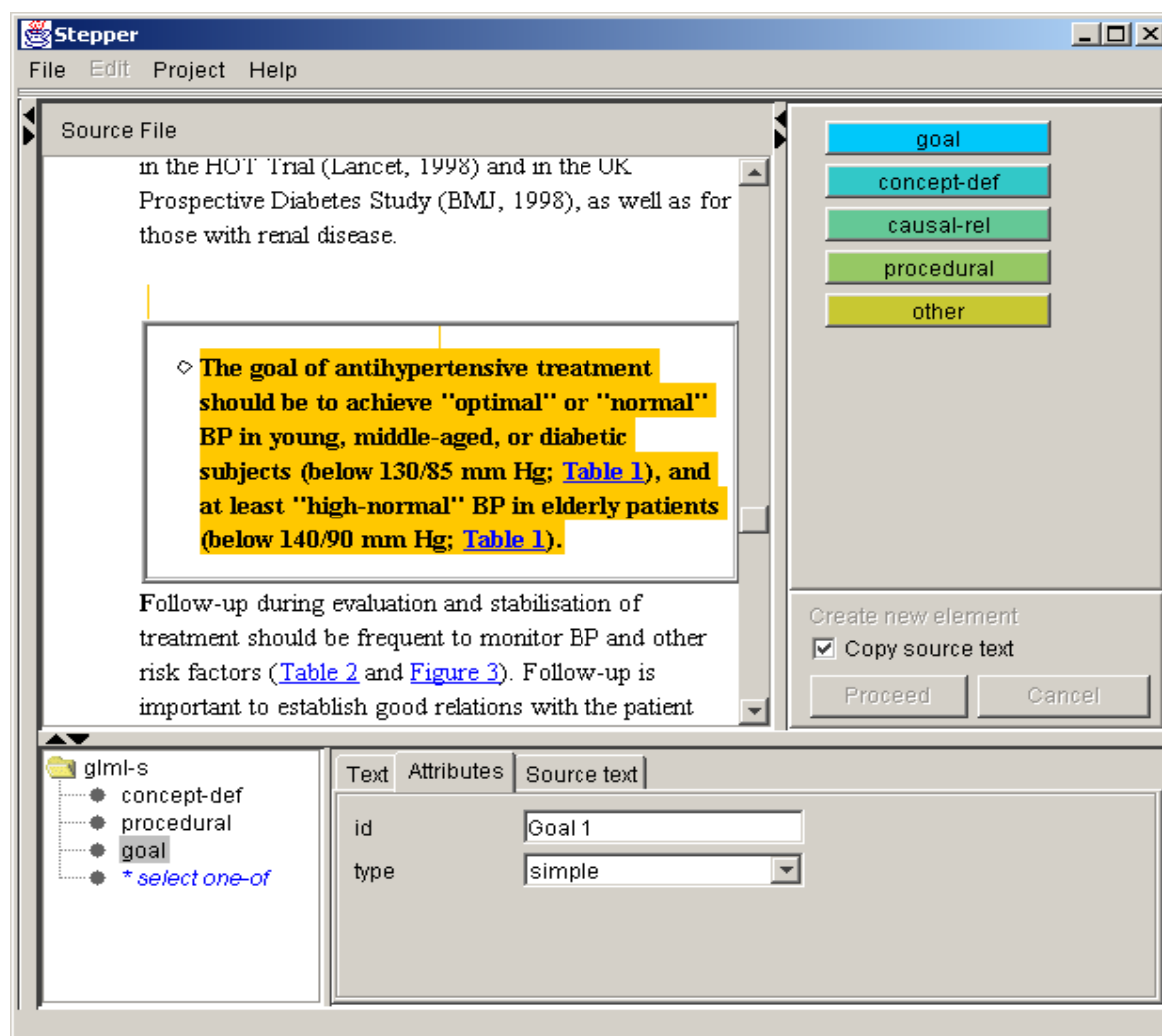
Další okruh problémů k řešení souvisí s *implicitními a vágními znalostmi*. V současnosti je jejich přidávání resp. zpřesňování pouze dokumentováno; v budoucnu bychom chtěli explicitně předcházet subjektivizaci, např. tím, že by problematické části textu interpretovalo více expertů, a výsledný znalostní element by zahrnoval „konsenzuální agregaci“ jejich zkušeností.

Naposledy se zmíníme o nutnosti ověřit celý přístup na *větším počtu dokumentů LDP*, a to zejména z českého prostředí. Výhodou zde může být snazší kontakt na samotné autory LDP při řešení nejasností spojených s výkladem textu.

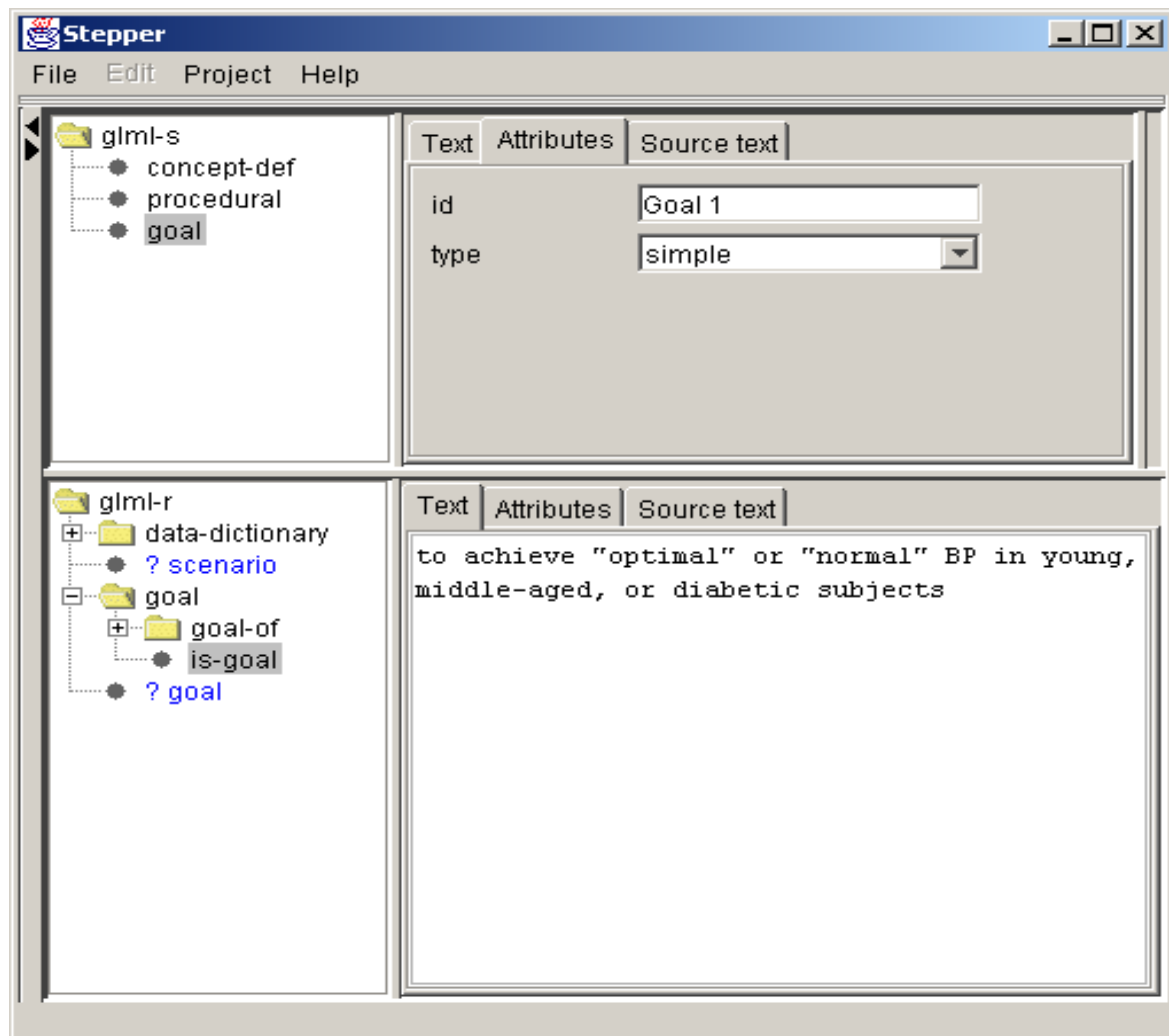
Literatura

- [1] WHO/ISH Guidelines for the Management of Hypertension. *Journal of Hypertension*, 17, 1999, 151–183.
- [2] XHTML 1.0: The Extensible HyperText Markup Language, <http://www.w3.org/MarkUp/>.
- [3] A. Advani, Y. Shahar, M. A. Musen. Medical Quality Assessment by Scoring Adherence to Guideline Intentions. 2001 Annual AMIA Conference, Washington, DC, Hanley and Belfus 2001.
- [4] Clark J.: XSL Transformations (XSLT) Version 1.0. W3C, 1999. <http://www.w3.org/TR/xslt>.
- [5] P. D. Johnson, S. Tu, N. Booth, B. Sugden and I. N. Purves: Using Scenarios in Chronic Disease Management Guidelines for Primary Care. AMIA Annual Symp., Los Angeles, CA, 389–393. 2000.
- [6] Kosara, R.; Miksch, S.; Seyfang, A. and Votruba P.: Tools for Acquiring Clinical Guidelines in Asbru, to appear in Proceedings of the Sixth World Conference on Integrate Design and Process Technology (IDPT'02).
- [7] Mar Marcos, Hugo Roomans, Annette ten Teije and Frank van Harmelen: Improving medical protocols through formalisation: a case study. To appear in Proceedings of the Sixth World Conference on Integrate Design and Process Technology (IDPT'02).
- [8] E. Motta: Reusable Components for Knowledge Modelling: Principles and Case Studies in Parametric Design. IOS Press, Amsterdam, 1999.
- [9] Musen M.A., Tu S.W., Das A.K., Shahar Y.: EON: A Component-Based Approach to Automation of Protocol-Directed Therapy. *JAMIA* 3:367–388, 1996.
- [10] M. Peleg, A. A. Boxwala, O. Ogunyemi, Q. Zeng, S. W. Tu, E. Bernstam, L. Ohno-Machado, E. H. Shortliffe and R. A. Greenes: GLIF3: The Evolution of a Guideline Representation Format. AMIA Annual Symposium, Los Angeles, CA, (20 Suppl):645–649. 2000.
- [11] Peleg, M.; Tu, S.; Bury, J.; Ciccarese, P.; Fox, J.; Greenes, R.; Hall, R.; Johnson, P; Jones, N.; Kumar, A.; Miksch, S.; Quaglini, S.; Seyfang, A.; Shortliffe, E.; Stefanelli, M.: Comparing Computer-Interpretable Guideline Models: A Case-Study Approach, to appear in *The Journal of the American Medical Informatics Association (JAMIA)*, 2002.
- [12] Peleška, J., Švejda, D., Zvárová, J.: Počítačová podpora rozhodování při léčbě arteriální hypertenze. *Lékař a technika* 28, 1-2, 1997.
- [13] S. Quaglini, M. Stefanelli, G. Lanzola, V. Caporusso, S. Panzarasa: Flexible guideline-based patient careflow systems, *Artificial Intelligence in Medicine*, Volume 22, Issue 1, April 2001.
- [14] Růžička, M.: Víceúrovňová formalizace obsahu textových dokumentů. Diplomová práce, VŠE Praha, 2002.
- [15] Říha A., Svátek V., Němec P., Zvárová J.: Medical guideline as prior knowledge in electronic healthcare record mining. In: 3rd International Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields, 25-27 September 2002, Bologna, Italy. WIT Press, to appear.
- [16] Seroussi B., Bouaud J., Antoine E. - C., Zelek L., Spielmann M.: An experiment in sharing and reusing OncoDoc's breast cancer guideline knowledge. In: (B. Heller , M. Löffler , M. Musen and M. Stefanelli, eds.) *Computer-Based Support for Clinical Guidelines and Protocols*, IOS Press, Vol. 83, Studies in Health Technology and Informatics 2001.

- [17] Shahar Y., Miksch S., Johnson P.: The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines. *Artificial Intelligence in Medicine* 14:29-51, 1998.
- [18] Shahar, Y.: A Hybrid Framework for Representation and Use of Clinical Guidelines. In: *Proc. AMIA 2002*, San Antonio, Texas 2002.
- [19] R. D. Shankar, S. W. Tu, S. B. Martins, L. M. Fagan, M. K. Goldstein, and M. A. Musen. Integration of Textual Guideline Documents with Formal Guideline Knowledge Bases. In: *AMIA 2001*.
- [20] R. N. Shiffman, B. T. Karras, A. Agrawal, R. Chen, L. Marengo and S. Nath: GEM: A proposal for a more comprehensive guideline document model using XML. *JAMIA* 2000; 7(5):488–498.
- [21] V. Svátek, T. Kroupa and M. Růžička: Guide-X – a Step-by-step, Markup-Based Approach to Guideline Formalisation. In: (B. Heller , M. Loeffler, M. Musen and M. Stefanelli, eds.) *Computer-Based Support for Clinical Guidelines and Protocols*, IOS Press, Amsterdam, 2001, 97–114.
- [22] Svátek V., Zvárová J., Jiroušek R.: A Two-Tiered Model of Medical Guideline. In: *MIE2000/GMDS2000, Medical Infobahn for Europe*. (CD-ROM). Berlin : Quintessenz Verlag, 2000.



Obr. 1: Vyznačování znalostí v textu



Obr. 2.: Transformace znalostí mezi úrovněmi zpracování

Nr.	BP	Age	Diabetic	Treatment	RESULT
1	Normal	middle-aged	no	monotherapy	GOAL ACHIVED
2	Grade 1	middle-aged	no	none	!!! BEGIN TREATMENT
3	Grade 3	elder	yes	monotherapy	!!! GOAL WASNT ACHIVED
4	Optimal	young	no	none	NOT APPLICATED

Obr. 3: Automaticky vygenerovaná demonstrační aplikace