

Posudek školitele doktorské disertační práce Ing. Karla Zváry „Extrakce informací z lékařských textů“

Předkládaná disertační práce se věnuje vysoce aktuálnímu tématu: provázání lékařských zpráv, vznikajících v drtivě většině v podobě volného resp. jen mírně strukturovaného textu, a elektronického záznamu o nemocném, jehož efektivní zpracování předpokládá vysokou míru strukturovanosti a unifikovanosti. Tématu se ve světě věnuje velký počet týmů, přesto však dosud není uspokojivě vyřešen, a to ani pro anglický jazyk, pro který existuje velký počet lexikálních i odborných zdravotnických znalostních zdrojů. Disertace se nicméně zaměřila na texty zpráv v jazyce českém; pro tuto oblast je pokrytí existujícím výzkumem minimální, a jde tedy o práci do značné míry průkopnickou.

Práce je členěna do šesti kapitol, z toho klíčové jsou kapitoly 3, 4, a zejména 5, popisující vlastní výzkum autora, který byl také vesměs publikován v recenzovaných článcích (tyto kapitoly zaujmají cca 70% celého textu). Zbývající tři kapitoly představují zasazení této vlastní práce do kontextu předchozího výzkumu v tuzemsku i ve světě, do legislativního rámce v ČR, relevantních mezinárodních standardů, a dále pak závěrečné shrnutí.

V Kap. 3 autor na rozsáhlém vzorku lékařských zpráv analyzoval využitelnost vybrané množiny metod automatického zpracování přirozeného jazyka (zejména sofistikované tokenizace, lemmatizace a rozpoznávání slovních druhů) a mapování na zdravotnické nomenklatury. Výsledky ukázaly, že plně automatické zpracování zpráv naráží na zásadní limity, související zejména s nepřesným a zkratkovitým vyjadřováním v textech.

Kap. 4 je v jistém smyslu komplementární k předchozí: ukazuje, jak lze v případě přesně zadávaných vstupů (v kontrastu k ad hoc způsobu zapisování textových lékařských zpráv) dosáhnout vysoké kvality výsledné strukturované databáze. Širší projekt z oblasti orální medicíny, na kterém se doktorand podílel v oblasti tvorby a implementace datového modelu, se zaměřoval na mapování stavu chrupu pacienta prostřednictvím hlasové komponenty přímo rozpoznávající informace odpovídající typům z tohoto datového modelu.

Kap. 5 pak navazuje na obě předchozí (především však na Kap. 3): demonstruje přínos synergického zpracování textových lékařských zpráv, zahrnujícího jak aplikaci automatických procedur zpracování přirozeného jazyka, tak ruční úpravy prováděné experty – lékaři. Pro tento účel byla navržena tzv. třífázová metoda předzpracování zpráv (3PP), zahrnující fáze automatické tokenizace, interaktivní normalizace (opravu širokého spektra chyb a expanzi zkratk), a rovněž interaktivní sémantické anotace sadou zdravotnických nomenklatur. Metoda je podporována softwarovým nástrojem TOCESA, který doktorand rovněž vyvinul a otestoval. Jako modelový případ užití byl zvolen scénář automatického překladu textové zprávy z češtiny do angličtiny. Experimenty provedené za účasti dvou lékařských expertů na 49 zprávách z oblasti kardiologie jednak opět potvrzují, že normalizace zpráv je netriviální a vyžadující značnou expertizu, ale také, na druhé straně, že aplikace metody 3PP má na kvalitu zpráv zásadní vliv, odrážející se i na srozumitelnosti výsledného překladu.

Ing. Zvára své výsledky publikoval prostřednictvím významných mezinárodních publikačních platform. Zásadní význam má zejména článek v časopise *Methods of Information in Medicine* (IF=1,772), u kterého je hlavním autorem; článek obsahově odpovídá Kap. 5 disertační práce. Ze spoluautorských prací bych vyzvil článek v dalším prestižním časopise, *BMC Med. Inf. &*

Decision Making (IF 1,603). Za zmínku stojí i spoluautorství příspěvků na dvou ročnících MIE, nejvýznamnější evropské konference v oblasti zdravotnické informatiky; příspěvek z r. 2015, i článek v dalším IF časopisu *Biocybernetics and Biomedical Engineering*, se vztahují ke Kap. 4 disertace. Pokud jde o Kap. 3, odpovídající ranější fázi výzkumu, výsledky byly postupně publikovány ve dvou článcích v *Eur. J. Biomed. Inform.*

Nad rámec hodnocení disertační práce bych rád vyzdvihl velmi dobrou spolupráci s doktorandem samotným. Ing. Zvára ve svém výzkumu postupoval samostatně, aktivně si domlouval spolupráci s dalšími (lékařskými i nelékařskými) odborníky, a vyhledával si publikační cíle. Průběžně verze práce mi pravidelně postupoval k připomínkám, které vždy v krátké době reflektoval. Vedle jeho úsilí o dosažení originálních vědeckých výsledků si cením i jeho dlouhodobého úsilí přispět k řešení praktických problémů spojených se stavem zdravotnické dokumentace v ČR. Spolupráce s ním pro byla velmi obohacující jak pro mě, tak i pro mé další doktorandy a mladší kolegy, se kterými jsem měl příležitost ho zkontaktovat, a kterým přispěl užitečnými radami z oblasti své odbornosti.

Závěrem shrnuji, že předložená práce podle mého názoru splňuje požadavky kladené zákonem na disertační práci, a **doporučuji ji proto přijmout k obhajobě.**

V Praze dne 22.6.2017

doc. Ing. Vojtěch Svátek, Dr.