

Oponentský posudek dizertační práce Ing. Václava Papeže

"Archetype-based approach for modelling of electroencephalographic/event-related potentials data and metadata"

Analýza biomedicínských signálů patří v současnosti mezi velmi důležité metody sloužící pro podporu rozhodovacích procesů jak v diagnostice, tak v terapii. V některých lékařských oborech se analýza signálů stává téměř nutností. Neurologie a neurofyzologie patří mezi takové obory již dlouhodobě. Přitom se ukazuje řada problémů spojená s neexistencí obecného datového standardu jak pro elektroencefalografické (EEG) záznamy, tak pro evokované potenciály (ERP). Předkládaná disertační práce se zabývá otázkou využití datových modelů pro popis EEG a ERP. Autor navrhuje nové archetypy a jejich propojení s existujícími daty a metadaty. Navržené metody a postupy byly implementovány a otestovány na reálných datech.

Práce je členěna do deseti kapitol, které jsou tematicky sloučeny do pěti bloků, a příloh; obsahuje seznam obrázků, tabulek, zkratk a symbolů a celkem 70 referencí.

Celkové hodnocení práce

Aktuálnost zvoleného tématu

Zvolené téma dizertační práce má interdisciplinární charakter a je aktuální nejen z vědecko-výzkumného, ale i praktického hlediska. Zároveň je to téma velmi náročné. Pro popis EEG a ERP dosud neexistují jednotné standardy ani pro datové formáty, ani pro datové modely. Dizertace je příspěvkem k modelování dat s využitím archetypů openEHR. Každý příspěvek k formalizaci popisů lze tak považovat za velký přínos.

Cíl dizertace

Cíle dizertace jsou explicitně, ale dosti stručně uvedeny v abstraktu. Celkem si autor stanovil 2 klíčové cíle: propojení existujících dat/metadat v EEGBase se strukturami archetypů a návrh nových archetypů popisujících EEG/ERP doménu. Hlavní cíle jsou rozpracovány na následující dílčí cíle: analýza běžných datových modelů za účelem vymezení společných prvků modelů a vystavění hierarchie datových modelů v závislosti na jejich vyjadřovací síle. Práce postupuje od popisu problémové oblasti a rešerše literatury, přes návrhy metod, jejich implementace až po testování navržených řešení a diskusi dosažených výsledků. Navržené metody byly experimentálně ověřeny na reálných datech. Autor své výsledky analyzoval velmi detailně. Všechny uvedené cíle autor splnil.

Zvolené metody zpracování

Práce je vedena metodicky správně. Po úvodu následují kapitoly 2 a 3, které stručně uvádějí do relevantních partií elektroencefalografie a neuroinformatiky, včetně problematiky neuroinformatických dat a metadat. Čtvrtá a pátá kapitola se věnují teoretickým otázkám datových modelů a mapování mezi modely s ohledem na vypovídací schopnost modelů. Následuje kapitola 6, která diskutuje elektronický zdravotní záznam z pohledu EEG/ERP záznamů a uvádí přehled standardů pro elektronické zdravotní záznamy. Kapitola 7 je klíčová a popisuje návrh archetypů pro specifikaci EEG/ERP a experimentů v rámci konceptu openEHR. Důležitým prvkem je integrace kontextové informace. Kapitola 8 je věnována popisu dosažených výsledků. Velmi důležité je vyhodnocení archetypů openEHR z hlediska požadavků na datový formát EEG. V kapitole 9 je diskutováno využití výsledků pro výkum a klinické použití. V závěrečné kapitole 10 jsou shrnuty výsledky ve vztahu k formulovaným cílům a jsou uvedena doporučení pro další rozvoj. V příloze jsou pak uvedeny definice vytvořených archetypů.

Výsledky dizertace a její hodnocení

Předložená dizertační práce je dobře strukturovaná a má správný metodický postup. Autor vede postup řešení jak ve vědecké, tak i implementační části jasně a správně. Hlavní přínos práce spatřuji v tom, že autor identifikoval základní problémy formalizace popisu EEG/ERP záznamů pomocí datových modelů a archetypů v rámci standardů používaných v oblasti elektronických zdravotních záznamů. Navrhl metodický postup tvorby modelů, definoval specifikace datových modelů a archetypů, ukázal výhody navržené standardizace. Navržené archetypy implementoval a otestoval na reálných datech. Bylo tak prokázáno, že nově navržená metodika a implementované archetypy jsou vhodné.

Práce je i po stránce formální na velmi dobré úrovni. Oceňuji, že je napsána v anglickém jazyce.

Význam pro praxi a další rozvoj vědy

Práce je bezpochyby přínosná nejen pro oblast návrhu a implementace podobných modelů a archetypů, ale zejména pro oblast managementu a sdílení dat v biomedicinském výzkumu a klinické praxi. Navržená metodika byla ověřena v praxi, takže též může být do budoucna využívána pro podobné úlohy. O přínosu práce svědčí i publikované práce doktoranda, zejména časopisecké články.

Připomínky

K práci mám následující připomínku:

- Postrádám samostatnou kapitolu definující cíle dizertace. Explicitně jsou vyjádřeny ve velmi stručné podobě v abstraktu. Doporučuji, aby uchazeč ve své prezentaci při obhajobě explicitně uvedl všechny cíle práce a aby na závěr ukázal, jak byly v práci splněny.

Dotazy do diskuse

- Jak snadno/obtížně bude možné integrovat kontejner pro data v jiném formátu než HDF5?
- Na obr. 3.3 je ilustrován popis analogového signálu v metadata modelu. Jak by byl reprezentován vícekanálový záznam?

Závěr

Přes výše uvedené připomínky mohu konstatovat, že předložená práce svědčí o cílevědomém a systematickém přístupu doktoranda ke zpracování dané problematiky jak z hlediska vědeckého přístupu, tak z hlediska praktického využití. Doktorand jednoznačně prokázal, že ovládá vědecké metody a disponuje potřebnou hloubkou teoretických vědomostí a praktických zkušeností.

Práce splňuje podmínky samostatné tvůrčí vědecké práce, obsahuje původní výsledky a odpovídá tak požadavkům k udělení akademického titulu PhD. Mohu tak konstatovat, že **práci doporučuji k obhajobě před příslušnou komisí.**



Doc. Ing. Lenka Lhotská, CSc.

V Praze dne 14.8.2017



Nathan C. Lea
UCL Institute of Health Informatics
222 Euston Road
London W1 2DA
UNITED KINGDOM

Doc. Rd. Ing. Vlasta Radová
Prof. Ing. Pavel Novák
Západočeská univerzita v Plzni
Děkanát FAV
Ing. Jaroslav Toninger
Univerzitní 8
306 14 Plzeň
Czech Republic

15th August 2017

Dear Doc. Dr. Ing. Radová and Prof. Ing. Novák,

Re.: Doctoral Thesis, ING. VÁCLAV PAPEŽ

Many thanks for your invitation to review the Doctoral Thesis of student Ing. Václav Papež on 12th July 2017, which I was delighted to accept on 21st July 2017. I am pleased to provide my recommendation below according to the guidelines that you have provided.

The thesis describes work to harness the use of the openEHR Archetype approach for representing clinical records within the experimental electroencephalographic /event related potential (EEG/ERP) domain to achieve a common data standard and reference implementation, which do not exist within the field.

The work applies a standardization approach that assesses the core components for data representation between persistence and communication, views each element's utility and justifies the need for the Archetype / two model approach for data representation.

It further assesses the field and considers the various standards available for providing a semantically viable clinical data representation, justifying the use of openEHR. It concludes by providing the developed screens and scripts and a treatise on their representation.

Overall I am satisfied with the quality of the work and I recommend the thesis for defence.

Significance of the Doctoral Thesis to the field:

The contribution made by this work is significant: it provides a detailed treatise on the complex task of developing a computable and shareable representation of a clinical intervention and the importance of understanding the context within which it would need to operate.

This serves as a fascinating and useful guide to the different technologies involved and emphasizes the need to understand and master each in order to faithfully and reliably represent the various clinical concepts digitally. This further provides an important critique on the nature of archetype and formal modelling of clinical concepts that would aid any clinical concept modeler and EHR system architect in understanding the layers involved in developing a semantically interoperable and standards compliant clinical information system.

Approach to solving the problem, methods used, and fulfilment of the given objective:

The student has taken a methodical approach to reviewing and explaining each layer of the design and software implementation paradigms and tools responsible for handling an EHR system. They have identified a set of independently developed desiderata against which the approach would need to be validated to define a successful outcome. Both the approach and the use of desiderata provided by research independent of the student's work provide a meaningful validation and justification for applying the openEHR standard to clarify and express the required clinical concepts.

The review has also established a basis upon which the student can explain the reasoning behind why he chose to model various concepts in the way that he did and how the openEHR model has helped the representation of those concepts.

The student has used established modeling approaches including Entity Relationship Modelling and Unified Modelling Language (UML) to explore the problem domain and provide a basis upon which the Archetypes themselves could be modelled more robustly. I note however one possible error: on page 41 Figure 4.10, I believe that the arrowhead pointing towards Laboratory from Institution should be pointing in the opposite direction; in its current form it suggests that an Institution is a kind of Laboratory, whereas I wonder whether it is vice versa.

By taking an approach that ensured he constructed the proposed Archetypes and deployed them for both server and client side technologies and related the results back to the desiderata, the student has been able to validate the work reasonably. I would have preferred more detail on how the Archetypes and screens he produced met the identified desiderata but I believe that this is something that could be probed during the defence.

Results and original contribution to the area:

The results are well presented and clearly articulated. It is noteworthy as an original contribution that the developed Archetypes are being used for a live system that is

currently under development. The original contribution is further supported by the critique that the student has applied against how the Archetypes are representable in the other semantic and persistence expressions, including Entity Relationships and odXML.

It is clear that the results provide a robust contribution to an area of neuroscience care and research that has no alternative expressions or implementations beyond an ad hoc handling of data items that lacks semantic robustness, expressiveness or any kind of meaningful interoperability.

It would however be useful to explore during the defence how the student believes that the screens and representations will work in actual use, particularly where there may be significant benefits or detriments.

Systematic approach, clarity, appropriateness of form and language:

It is clear that the student has provided a systematic review and critique of the expressions and persistence mechanisms available and how best to apply an openEHR approach to standardising the capture and representation of the data of interest both to a user in the context of neuroscience research and to the persistence tiers of a clinical system architecture.

I was impressed with the clarity and appropriateness of how the student expressed their arguments – the language was clear, concise and unambiguous, and the expression remained impartial and dispassionate, culminating in a well written piece of work. There was only one sentence that I had trouble following as it seems awkwardly expressed on page 67: the second sentence of the first paragraph needs to be rewritten, specifically, “...EEG/ERP can be deliberate in some way...” – deliberate as in non-routine?

The only points of concern I noted related to the failure to use the definite and indefinite article (e.g. page 7: **The** ERP method; Page 21: **The** CARMEN structure; page 68 first paragraph: **a** routine and throughout this page, **a** patient). This is fortunately an easily remedied problem and one that does not detract in any significant way from the clarity of the work, where its rectification would provide a polish that the work deserves.

Whilst there are very few typos that I can identify for which the student should be congratulated, but it may be advisable for the Student to do a check through for any outstanding issues prior to final publication (e.g. page 84 *codding* should I suspect be **coding** and page 90 *templates* should be **the templates**’.

The student’s publications

The student has cited in the thesis at least six of their own recent publications and their MSc. thesis. They have also provided a comprehensive list of their other journal publications and dissemination activities.

There appears to be a clear outline and progression of the student’s work during the thesis and a definite contribution to the overall thesis preparation. This helps to

provide clarity and rigor to the structure of the thesis and the themes, relating to semantic structure design, Archetype modelling and openEHR for meeting wider clinical domain concept and phenotyping requirements.

To that end, the publications independently show a peer-reviewed talent for assessing and critiquing both openEHR and other standards and technologies for semantically meaningful representation of a variety of clinical concepts. This provides further validation of the student's approach and techniques, and further assures me that I can recommend his thesis for defence.

Please do contact me if you need any further details or clarifications for this review.

Yours sincerely,

A handwritten signature in black ink, appearing to read 'NCL', written in a cursive style.

Nathan Christopher Lea B.A., MSc., Ph.D., MBCS
Senior Research Associate,
UCL Institute of Health Informatics