

# Posudek oponenta diplomové práce

**Autor práce:** Lukáš Ščurko

**Název práce:** Standardizované datové formáty pro elektrofyziologická data

## Obsah práce

Autor v diplomové práci popisuje standardy pro ukládání elektrofyziologických dat. Hlavním cílem je najít vhodný standard pro převod dat z BrainVision formátu používaného v neuroinformatické laboratoři na Katedře informatiky a výpočetní techniky na Západočeské univerzitě v Plzni. Autor prozkoumal standardy z několika pohledů, podle kterých je nakonec vybrán výsledný standard Brain Imaging Data Structure (BIDS). Do tohoto standardu musí nakonec převést stávající datasety z neuroinformatické laboratoře. Při výběru standardu autor ověřuje, zda splňuje tzv. FAIR (Findable, Accessible, Interoperable, Reusable) principy definované mezinárodní organizací s názvem International Neuroinformatics Coordinating Facility (INCF). V práci je též řešena složitost převodu dat a podpůrných knihoven k prozkoumávaným standardům. Nakonec autor implementoval nástroj, který elektrofyziologická data z BrainVision formátu převede do vybraného BIDS standardu. Pro převedená data bylo zvoleno vhodné datové úložiště GIN německé neuroinformatické skupiny, do kterého se data uložila.

## Kvalita řešení a dosažených výsledků

Autor úspěšně naprogramoval aplikaci v jazyce Python pro převedení datasetů z neuroinformatické laboratoře do mezinárodního BIDS formátu. V práci je dobře popsána analýza FAIR principů pro sdílení a správu vědeckých dat. Úroveň porovnání dostupných datových úložišť, standardů i open source nástrojů pro export dat je velmi dobrá. Dále vybírá open-source nástroj MNE v jazyce Python pro zkoumání, vizualizaci a analýzu neurofyziologických dat. Autor vybírá GIN úložiště od německé neuroinformatické organizace jako vhodné datové úložiště pro elektrofyziologická data a metadata. Každý dataset má v rámci úložiště zdarma vygenerovaný jedinečný DOI identifikátor. Zdrojový kód je dobře okomentován a kvalita zdrojových kódů je dobrá. Autor implementoval jednotkové testy pro kontrolu jednotlivých metod aplikace. Aplikace mi byla předvedena autorem a žádný problém nenastal. Když jsem aplikaci testoval, tak jsem narazil na pár chyb. Při přidání datasetu jsem vybral složku PROJECT\_DAYS\_P3\_NUMBERS z elektronické přílohy. Textové pole obsahovalo jméno datasetu PROJECT\_DAYS\_P3\_NUMBERS. Rozhodl jsem se zdrojovou složku změnit a vybrat dataset řidiče (Driver's\_attention\_with\_auditory\_stimulation), který je řádově větší. Název datasetu se po změně automaticky nezměnil (při první zadání se změní) a při spuštění převodu se mi zobrazilo chybové okno „Error dataset name is not valid“. Dále jsem narazil na problémy s pamětí<sup>1</sup> při použití větších datasetů (použito z elektronické přílohy složka Driver's\_attention\_with\_auditory\_stimulation). U datasetů PROJECT\_DAYS\_P3\_NUMBERS aplikace fungovala bez problémů (např. 1 dataset má velikost přibližně 10 MB a načetl jsem všech 250 datasetů z elektronické přílohy, např. dataset řidiče „Driver's\_attention\_with\_auditory\_stimulation“ má velikost 459 MB).

---

<sup>1</sup>MemoryError: Unable to allocate 429. MiB for an array with shape (17, 3303880) and data type float64

## Formální úroveň a práce s literaturou

Po formální stránce je kvalita diplomové práce vyhovující. Práce je logicky strukturovaná. Text práce je v anglickém jazyce. Autor použil pro sazbu textu systém LaTeX. Počet překlepů je v toleranci vzhledem k rozsahu práce (55 stran práce, cca 39 normostran). Dále je v diplomové práci 46 zdrojů a 4 přílohy. Tištěná příloha obsahuje MME metody použité pro převod, uživatelský manuál pro převod dat, popis datové a metadatové struktury před a po převedení, workflow a datový model MME společně z NEO diagramem. Elektronická příloha obsahuje projekt aplikace se zdrojovými kódy, zdrojové datasety pro export (měření pozornosti řidiče s vizuální stimulací a zvukovým rušením, měření stimulace čísl z projektových dní na základních a středních školách), text diplomové práce v PDF + LaTeX zdrojové kódy, poster a výsledné datasety v BIDS formátu (celkem 251).

Autor se dostatečně seznámil s problematikou standardů pro ukládání elektrofyziologických dat. Seznam literatury obsahuje velké množství časopiseckých publikací a referencí na weby daných formátů (např. BIDS), open source nástrojů (např. MME), datových úložišť (např. GIN) zabývajících se problematikou standardizace. Uvedené zdroje jsou vhodně zvolené vzhledem k tématu práce.

## Splnění zadání

Všechny body zadání diplomové práce byly splněny.

Vzhledem k délce práce (39 normostran) a nalezeným nedostatkům (jsou uvedeny v sekci kvalita řešení a dosažených výsledků) navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

## Dotazy k práci

Jaké jsou možnosti vyhledávání metadat z našich BIDS datasetů v úložišti GIN?