

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Jan Janda
Název práce CSV file validator according to the CSV on the Web W3C recommendations
Rok odevzdání 2024
Studijní program Informatika **Studijní obor** Softwarové a datové inženýrství

Autor posudku Martin Svoboda
Pracoviště KSI

Role Oponent

Text posudku:

Datový formát CSV se řadí mezi tradiční způsoby ukládání dat relační povahy v podobě textových souborů. Bohužel není dostatečně standardizovaný, sám o sobě navíc ani neumožňuje dostatečný popis vnitřní struktury dat aneb očekávaného schématu. Cílem hodnocené diplomové práce bylo seznámení se s existujícími řešeními umožňujícími validaci takových CSV souborů vůči schématům reprezentovaných pomocí W3C konceptu *CSV on the Web* a následný návrh, implementace a evaluace autorova vlastního rozšiřitelného nástroje řešícího slabé stránky těch existujících. Popsané zadání práce vnímám jako splněné.

Jádrem navrženého řešení je knihovna napsaná v jazyce Java, která umožňuje provádět validace na základě nejrůznějších dílčích kontrol konkrétních aspektů doplněných o automaticky řešené vzájemné závislosti mezi nimi. Přístup k funkcím knihovny je pak možný přes webové rozhraní a lokálně přes příkazovou řádku. Díky rozsáhlosti příslušné specifikace W3C byly v souladu s původním záměrem implementovány jen vybrané kontroly.

Text práce je napsán relativně dobrou angličtinou a jen s malým množstvím chyb. Styl vyjadřování však často není úplně plynulý a uhlazený, některé formulace nebo části nejsou dostatečně přesné nebo odborné. Přestože v obecné rovině jsou obvyklé konvence psaní odborných textů dodržovány, některé části nejsou úplně konzistentní, např. pokud jde o psaní velkých písmen nebo chybějící reference v úvodu.

Celkový rozsah textové části je adekvátní, obsahuje všechny očekávané součásti a většina textu obsahuje autorovy původní myšlenky. Rozložení některých informací napříč prací však možná chtělo promyslet lépe. Např. analýzu existujících řešení by bylo vhodnější uvést před soupisem požadavků a dalších částí specifikace navrhovaného řešení, pokud na ni logicky tyto části mají navazovat. Název projektu Otava je uveden až téměř v půlce práce, přestože už je používán před svým vysvětlením. Jiné části jako náčrty grafického uživatelského rozhraní působí jako nedodělané.

Pokud jde o reference jako takové, celkově je citováno opravdu velké množství informačních zdrojů. Naprosto všechny ale tvoří jen nejrůznější dokumentace, specifikace nebo další zdroje na webu, odborné publikace chybí zcela. Zdrojový kód působí dobře dekomponovaný a stejně tak i napsaný dobrým a konzistentním stylem. Chybí ale technická JavaDoc dokumentace přímo v kódu.

Z hlediska zvolené metody řešení mám určité pochybnosti ohledně rozdělení procesu validace do relativně velkého množství relativně malých dílčích kontrol. Pokud by totiž tyto validace měly být univerzální, uživatelé by jednoduše očekávali, že se zkontroluje vše, a neměli by důvod cokoli nastavovat nebo ovlivňovat. Aneb šlo by jen o interní implementační detail validátoru a přínosy takto jemné granularity by byly diskutabilní.

Pokud by naopak záměrem bylo umožnit implementovat specifické kontroly pro konkrétní situace, data, domény nebo projekty, uživatelé (myšleno programátoři) by patrně

spíše volili komplexnější kontroly ad hoc řešící ideálně vše na jednom místě s cílem minimalizovat programátorské úsilí. Vše pak vynikne zejména u velkých souborů, které by se nepodařilo načíst celé do systémové paměti. V takovém případě by díky izolovanosti kontrol mohlo být nutné provádět sekvenční čtení vstupu každou kontrolou, což by značně snížilo efektivitu a použitelnost celého řešení.

Práce má čistě softwarový charakter, avšak i z implementačního pohledu mi chybí nějaké hlubší výzvy. Jinými slovy i přes téměř až marketingově znějící nekonkrétní vychválení přínosů práce samotným autorem v závěru je pro mě obtížné identifikovat zajímavější a netriviálnější části z hlediska složitosti. I přes uvedené výhrady práci k obhajobě doporučuji aneb jako celek dle mého názoru splňuje požadavky na diplomové práce kladené.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Datum 2. 2. 2024

Podpis