

ZAF – Validátor SIP dle Národního standardu pro elektronické systémy spisové služby



ARCHIV
UNIVERZITY KARLOVY
Univerzita Karlova

LightComp
Enhance Your Technology

Petr Cajthaml – Tomáš Dvořák – Petr Pytelka – Zdeněk Vašek



Validátor ZAF

<https://validatorzaf.github.io/zaf/>

- ▶ Aplikace pro kontrolu SIP balíčků z ESSS dle Nár. standardu, analytický nástroj, bez formální garance shody s validátorem NA
- ▶ Autory současné aplikace Stanislav Bečvář, Petr Pytelka a Karel Žáček
- ▶ Reakce na aktuální možnosti validátoru NA, potřeba využít validátor v provozu – hromadné testování (test, produkce)

ZAF – zpracování archivních formátů



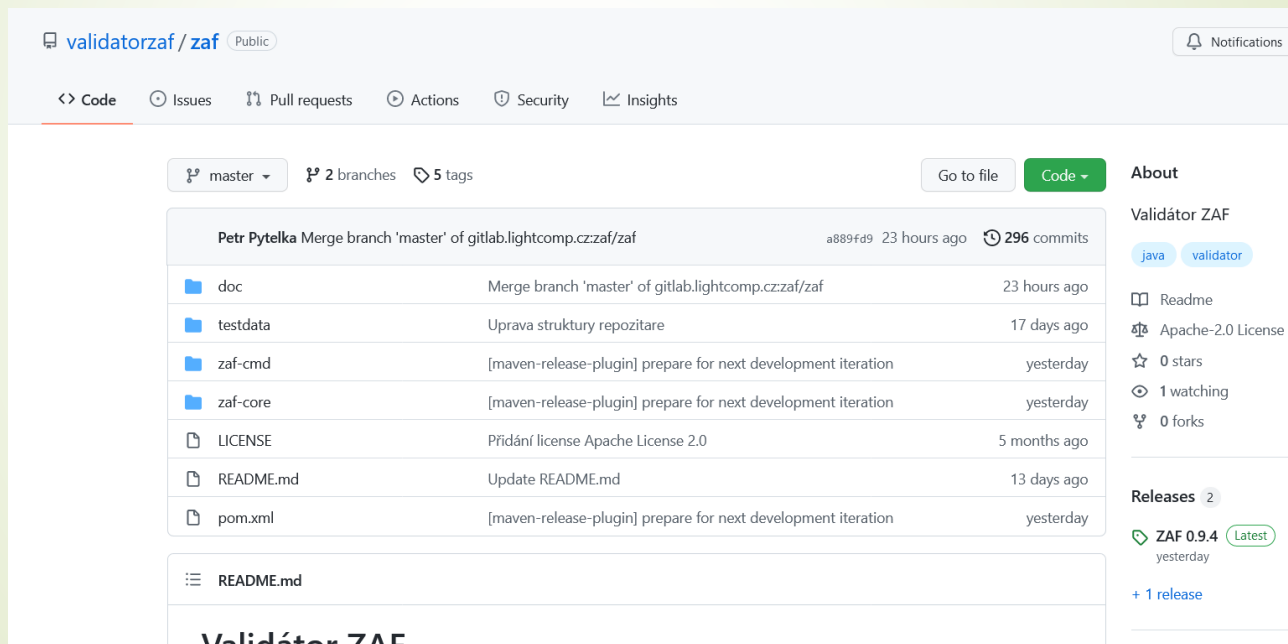
- Licence Apache 2.0
- Společný vývoj Archivu hl. m. Prahy, LightComp v.o.s. a Archivu UK, uzavřené Memorandum
- Od r. 2015 nasazen v Archivu hl. m. Prahy, od roku 2021 v Archivu UK a v ESSS UK



Odlišnosti oproti řešení NA

- Samostatně opensource dostupná aplikace včetně zdrojových kódů
- Lze zapojit jako knihovnu do vlastního systému (např. integrovat do ESSS, digitálního archivu)
- Rozhraní příkazového řádku
- Dávkové zpracování
- Strojově i lidsky čitelný výsledek validace, vysvětlení chyby a odkaz na předpis, lokalizace chyby v rámci SIP

Dostupný na <https://github.com/validatorzaf/zaf>



validatorzaf / zaf Public

<> Code Issues Pull requests Actions Security Insights

master 2 branches 5 tags Go to file Code

Commit	Message	Time
Petr Pytelka Merge branch 'master' of gitlab.lightcomp.cz:zaf/zaf		a889fd9 23 hours ago 296 commits
doc	Merge branch 'master' of gitlab.lightcomp.cz:zaf/zaf	23 hours ago
testdata	Uprava struktury repozitare	17 days ago
zaf-cmd	[maven-release-plugin] prepare for next development iteration	yesterday
zaf-core	[maven-release-plugin] prepare for next development iteration	yesterday
LICENSE	Přidání license Apache License 2.0	5 months ago
README.md	Update README.md	13 days ago
pom.xml	[maven-release-plugin] prepare for next development iteration	yesterday

README.md

Validátor ZAF

About

Validátor ZAF

java validator

Readme

Apache-2.0 License

0 stars

1 watching

0 forks

Releases 2

ZAF 0.9.4 Latest yesterday

+ 1 release

Zveřejněná dokumentace

<https://validatorzaf.github.io/zaf/doc/>

Obsah

Úvod

- Změny v pravidlech
- Licenční podmínky
- Autoři

Použití

- Samostatný CmdValidator
- Výsledek validace
- Java knihovna - přímá integrace
- Možnosti logování

Pravidla

- Kontrola škodlivého kódu
- Kontrola datové struktury
- Kontrola znakové sady
- Kontrola správnosti XML
- Kontrola jmenných prostorů
- Kontrola proti schématu
- Kontrola obsahu

ZAF 0.9.4: Dokumentace

Úvod

Validátor ZAF je knihovna pro kontrolu datových balíčků SIP podle [národního standardu pro elektronické systémy spisové služby z r. 2017](#) na základě sestavených [pravidel](#). Aplikace provádí načtení balíčků SIP (příp. rozbalení ze souborů ZIP), kontrolu a uložení [výsledků kontroly](#) do souboru XML vytvořeného podle schématu XSD. Kontrola škodlivého kódu probíhá externě a aplikace pouze načítá výsledky této kontroly podle použitých přepínačů. Pokud je v rámci jednoho z typů kontrol nalezena chyba, další typy kontrol se neprovádí.

Vývoj a další rozvoj Validátoru ZAF probíhá ve spolupráci Hlavního města Prahy, Magistrátu hlavního města Prahy, odboru Archiv hlavního města Prahy; [Univerzity Karlovy, Ústavu dějin a archivu Univerzity Karlovy](#) a Lightcomp, v.o.s.

Validátor ZAF je vyvíjen se snahou o zjištění shody a případných chyb validovaných balíčků SIP s NSESSS. Použití Validátoru ZAF neprokazuje splnění požadavků na balíčky SIP ze strany archivů rozhodujících ve skartačním řízení. Výsledky validace Validátorem ZAF nemusí být shodné s výsledky zjištěnými validátory digitálních archivů.

Změny v pravidlech

v1, 1.1.2022

- Přidána pravidla [obs100](#), [obs101](#), [obs102](#), [obs103](#), [obs104](#), [obs105](#), [obs106](#), [obs107](#)



Validační pravidla

- 115 validačních pravidel,
 - Obsah škodlivého kódu
 - Kontrola datové struktury
 - Kontrola znakové sady
 - Kontrola správnosti XML
 - Kontrola jmenných prostorů
 - Kontrola proti schématu
 - Logická kontrola obsahu (včetně formátů)
-
- Pravidla aplikována podle typu SIP

Dokumentovaná a zveřejněná validační pravidla

Pravidla

- Kontrola škodlivého kódu
- Kontrola datové struktury
- Kontrola znakové sady
- Kontrola správnosti XML
- Kontrola jmených prostorů
- Kontrola proti schématu
- Kontrola obsahu

obs101

Pravidlo	Pokud existuje jakýkoli element <nseess:Komponenty>, všechny dětské elementy <nseess:Komponenta> se stejnou hodnotou atributu poradi obsahují stejnou hodnotu atributu druh.
Kód	obs101
Zdroj	Informační list NA, č. 6/2020, č. 3/2020.
Popis chyby	Uveden je chybně druh komponent (počítačových souborů).
Aktivní	skartační řízení (jen metadata), skartační řízení (s komponentami), přejímka

443 ukázkových balíčků pro jednotlivá pravidla

🔍 master ▾ [zaf](#) / [testdata](#) / 06 KONTROLA OBSAHU /

Petr Pytelka Uprava struktury repozitare

..

📁 01-OK	Uprava struktury repozitare
📁 01-chyba	Uprava struktury repozitare
📁 02-OK1	Uprava struktury repozitare
📁 02-OK2	Uprava struktury repozitare
📁 02-chyba1	Uprava struktury repozitare
📁 02-chyba2	Uprava struktury repozitare
📁 02-chyba3	Uprava struktury repozitare
📁 03-OK	Uprava struktury repozitare
📁 03-chyba1	Uprava struktury repozitare
📁 03-chyba2	Uprava struktury repozitare
📁 03-chyba3	Uprava struktury repozitare
📁 04-OK	Uprava struktury repozitare

Výstupy validace (ukázka) – pravidlo 60, chyba

```
<kontrola nazev="souladu se schématem XSD" stav="ok"/>
<kontrola nazev="kontrola obsahu" stav="chyba">
  <pravidlo>
    <kod>obs39</kod>
    <zneni>Každý element &lt;mets:digiprovMD&gt; obsahuje v hierarchii dětských
    elementů &lt;mets:mdWrap&gt;, &lt;mets:xmlData&gt; právě jeden dětský element
    &lt;tp:TransakcniLogObjektu&gt;.</zneni>
    <zdroj>Bod 2.12. přílohy č. 3 NSESSS.</zdroj>
    <kodChyby>CHYBA</kodChyby>
    <popisChyby>Datový balíček SIP neobsahuje transakční protokol.</popisChyby>
    <vypisChyby>Element &lt;mets:xmlData> neobsahuje žádný dětský element
    &lt;tp:TransakcniLogObjektu&gt;.</vypisChyby>
    <mistoChyby>Řádek 234:23, element &lt;mets:xmlData&gt;.</mistoChyby>
  </pravidlo>
  <pravidlo>
    <kod>obs54</kod>
    <zneni>Pokud neexistuje žádný element &lt;nsesss:KrizovyOdkaz&gt; s atributem
    pevny s hodnotou ano, potom každý element &lt;mets:div&gt; obsahuje dětský element podle
    struktury entit/objektů (od spisového plánu po komponentu) v sekci dmdSec s atributem TYPE s
    hodnotou příslušné entity/objektu a s atributem DMDID s hodnotou příslušné entity/objektu v
    atributu ID a s atributem ADMID s hodnotou, která odpovídá hodnotě atributu ID příslušné
    entity/objektu v sekci amdSec (entita/objekt v hierarchii dětských elementů
    &lt;mets:digiprovMD&gt;, &lt;mets:mdWrap&gt;, &lt;mets:xmlData&gt;,
    &lt;tp:TransakcniLogObjektu&gt;, &lt;tp:TransLogInfo&gt;, &lt;tp:Objekt&gt;,
    &lt;tp:Identifikator&gt;, &lt;tns:HodnotaID&gt; a &lt;tns:ZdrojID&gt; odpovídá v hodnotách
```



Funkce validátoru a technické možnosti

- ▶ Nenáročný na provoz, možné integrovat do jiných aplikací
- ▶ Testování dávek stovek tisíc SIP
- ▶ Postupná validace podle jednotlivých sad pravidel
- ▶ OS Microsoft Windows, GNU/Linux, OS X nebo Oracle Solaris
- ▶ Java 1.8 a vyšší

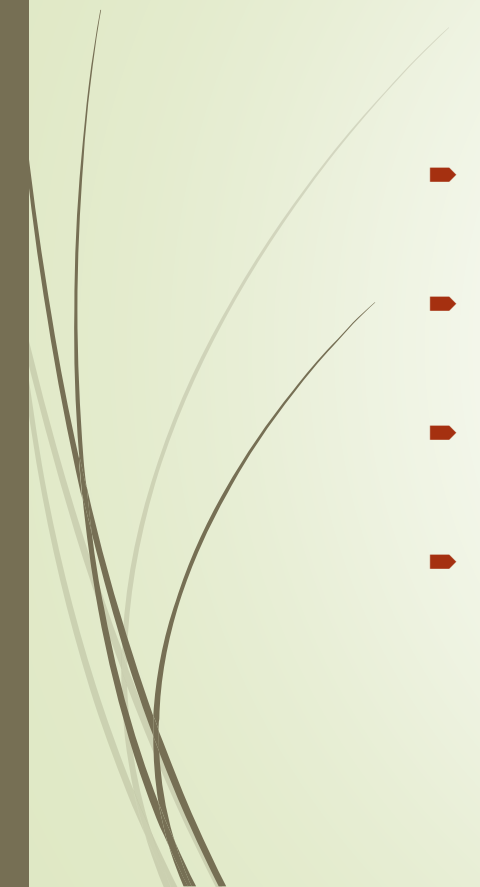


Další rozvoj

- ▶ Náhrada Apache Tika formátovým identifikačními nástroji Siegfried/Droid
- ▶ Doplnění API pro dávkové zpracování zapojenými nástroji
- ▶ Doplnění dalších pravidel
- ▶ Reakce na případné novelizace NSESSS



Možní uživatelé

- ▶ Zřizovatelé digitálních archivů
 - ▶ Archiváři testující skartace z ESSS
 - ▶ Původci provozující ESSS, digitální správní archivy a spisovny
 - ▶ Vývojáři a dodavatelé ESSS
- 



Možnosti spolupráce

- ▶ Uvítáme využití dalšími uživateli, zpětnou vazbu, hlášení chyb
- ▶ Testování i vývoj (optimalizace, refaktORIZACE)
- ▶ Metodická spolupráce (zpřesňování pravidel, tvorba nových)

DĚKUJEME ZA POZORNOST!

- ❖ petr.cajthaml@ruk.cuni.cz
- ❖ tomas.dvorak@praha.eu
- ❖ petr.pytelka@lightcomp.cz
- ❖ zdenek.vasek@ruk.cuni.cz



ARCHIV
UNIVERZITY KARLOVY
Univerzita Karlova

LightComp
Enhance Your Technology