



TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií ■

Editor konfiguračních souborů Flow123d

Diplomová práce

Studijní program: N2612 – Elektrotechnika a informatika

Studijní obor: 1802T007 – Informační technologie

Autor práce: **Bc. Tomáš Křížek**

Vedoucí práce: doc. Ing. Jiřina Královcová, Ph.D.





TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies ■

Editor for Flow123d configuration files

Master thesis

Study programme: N2612 – Electrotechnology and informatics

Study branch: 1802T007 – Information technology

Author: **Bc. Tomáš Křížek**

Supervisor: doc. Ing. Jiřina Královcová, Ph.D.



Tento list nahrad'te
originálem zadání.

Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Abstrakt

Český abstrakt

Abstract

English abstract

Poděkování

Obsah

Seznam zkratek	8
1 Rešerše	11
1.1 Software Flow123d	11
1.2 Jazyky pro popis dat	12
1.3 Formát konfiguračních souborů	13
2 Analýza	14
2.1 Konfigurační soubory	14
2.2 Input Structure Tree	14
2.3 Autokonverze	14
3 Návrh	15
4 Implementace	16

Seznam zkratk

EU Evropská unie

Úvod

Existuje celá řada softwarů, která pro zajištění požadované funkce potřebuje správné nastavení, podle kterého pak daný program přizpůsobí svoji činnost. Může se jednat o počáteční konfiguraci, jako je tomu například u serverových aplikací nebo e-mailových klientů. Tato konfigurace se zpravidla dále nemění, pokud nedojde k nějakým podstatným změnám.

Oproti tomu existují programy, od kterých se očekává, že budou spouštěny s širokou škálou různých nastavení. U těchto aplikací se typicky konfigurace předává při spuštění jako jeden ze vstupních parametrů. Činnost těchto programů se pak zásadně liší dle zvolené konfigurace.

Takovou aplikací je například simulátor Flow123d, který pracuje se zadanou výpočetní sítí, na které provádí výpočty dle definované úlohy. Úloha se definuje pomocí konfiguračního souboru, který pak spolu s výpočetní sítí a ostatními daty tvoří vstupní data pro aplikaci Flow123d.

Vzhledem k tomu, že Flow123d podporuje různé typy úloh a obsahuje rozsáhlé dílčí nastavení, vyvstává otázka toho, zda je zvolená konfigurace validní. Uživatel může omylem zvolit kombinace hodnot, které nejsou přípustné, ať už taková chyba vznikne logicky nebo například kopírováním. Dalším běžným případem chyby je překlep, kterého si uživatel nemusí všimnout.

Formát konfiguračních souborů pro Flow123d je poměrně rozsáhlý – samotná referenční příručka, která ho popisuje, obsahuje několik desítek stran. Čím je složitější formát pro konfiguraci, tím je pro člověka náročnější ověřit, zda byly zadány všechny povinné parametry, nebo zda při upravování souboru nedošlo k nějaké nepatrné změně, která do souboru zanesla chybu. Vzhledem ke složitosti formátu jsou na uživatele kladeny velké nároky – buď musí mít se softwarem rozsáhlé zkušenosti, nebo musí trávit čas prohledáváním dokumentace.

Celá situace je dále komplikována tím, že formát konfiguračních souborů se mění s tím, jak se vyvíjí funkce softwaru Flow123d. Některé změny ani nejsou zpětně kompatibilní. Tím pádem soubor, který byl validní pro starší verzi už nemusí být

validní pro nově vydanou verzi. Uživatele tedy opět čeká studium rozsáhlé referenční dokumentace, aby zjistil, jaké změny má provést.

Pokud se stane, že uživatel spustí Flow123d s nevalidní konfigurací, tak během inicializaci dojde k chybě, o které se uživatel dozví pomocí textového rozhraní, ve kterém se Flow123d spouští. Jelikož se může jednat o výpočetně náročné úlohy, které se často pouští na vzdáleném výpočetním clusteru, je tento proces poměrně zdlouhavý. Pokud se úloha spouští vzdáleně, musí dojít k navázání komunikace a přidělení zdrojů, než může vůbec dojít k inicializaci úlohy, při které pak může dojít k chybě. Odhalování takových chyb je pak časově náročné a uživatelsky nepříjemné.

Tyto důvody byly hlavní motivací ke vzniku speciálního editoru pro konfigurační soubory Flow123d, který práci s nimi značně zjednoduší a usnadní. Aplikace by měla zrychlit proces odstranění chyb v konfiguračních souborech tím, že je umožní odhalit už v průběhu jejich vytváření nebo upravování. To dává uživateli možnost chyby opravit ještě před tím, než předloží konfigurační soubor softwaru Flow123d. To vede ke značné časové úspoře obzvláště v případech, kdy se výpočetní úloha spouští vzdáleně.

Editor by měl uživateli poskytovat přívětivé uživatelské rozhraní, které mu zjednoduší přístup k dokumentaci. V rámci aplikace by měl mít uživatel k dispozici dokumentaci, která bezprostředně souvisí s právě upravovanou částí konfiguračního souboru. Tato forma nápovědy by měla uživateli poskytnout alternativu k manuálnímu prohledávání několika stránkové referenční dokumentace.

Dále by editor měl umožnit vizualizaci datové struktury, která tvoří konfigurační soubor. Kromě toho se od editoru očekávají základní funkce pro práci s textovými soubory, jako je podpora operací se schránkou, možnost vrátit či opakovat změny, vyhledávání či nahrazení textu a další. Editor by měl podporovat platformy Windows a Linux.

1 Rešerše

1.1 Software Flow123d

Flow123d je software, který slouží pro simulaci toků podzemních vod a transportu tepla. Jedná se o aplikaci, která je orientována na práci s daty. Vzhledem k tomu neobsahuje ani žádné grafické uživatelské rozhraní a uživatel s aplikací pracuje pouze pomocí předávání souborů a parametrů pomocí příkazové řádky.

Vzhledem k povaze úloh, které Flow123d řeší, je absence grafického uživatelského rozhraní pochopitelná. Simulace, které software provádí, jsou výpočetně náročné. Z pohledu uživatele je podstatný především výsledek této simulace, spíše než její průběh. Náročné úlohy jsou typicky spouštěny na vzdáleném výpočetním clusteru, kde by grafické rozhraní použití aplikace pouze komplikovalo.

Typický uživatel Flow123d si pomocí konfiguračních a dalších datových souborů vytvoří úlohu, kterou poté aplikaci předloží a ta provede simulaci nad zadanými daty. Po dokončení simulace má uživatel k dispozici výstupní datové soubory s výsledky simulace.

Uživatel ovšem potřebuje s datovými a konfiguračními soubory nějak pracovat. Pokud se nejedná o nějakou triviální úlohu, bylo by prakticky nemožné pracovat s výpočetní sítí bez pomoci nějakých dalších nástrojů, které jsou pro to určené. Obdobně jsou potřeba i nějaké specializované nástroje pro reprezentaci výstupních dat.

Pro práci s datovými soubory je k dispozici několik softwarů, které lze použít. Pro práci v s výpočetní sítí lze například použít software GMSH. Tvorbou aplikace pro vizualizaci výsledků jednoho typu úlohy jsem se zabýval ve své bakalářské práci. Vytvořená aplikace umožňovala generovat mapy koncentrací pro různé řezy výpočetní sítí a vytvářet grafy koncentrací pro jednotlivé elementy.

Doposud však neexistoval žádný nástroj, který by pracoval s konfiguračními soubory, pomocí kterých se úloha definuje. Definice této úlohy spočívá v inicializaci tříd ve Flow123d, které pak zajišťují samotnou simulaci. Úkolem konfiguračního souboru

je tedy doručit těmto třídám korektní data, z kterých se mohou inicializovat.

Tento proces je ovšem poměrně komplikovaný. Je potřeba si uvědomit, že vytváření úloh, tedy i tvorbu konfiguračních souborů, provádí lidé. Na druhé straně je pak samotný software Flow123d, pro který jsou podstatná pouze konkrétní data a hodnoty, ze kterých pak inicializuje potřebné třídy v C++. Je tedy potřeba, aby člověk komunikoval se softwarem pomocí nějakého rozhraní či specifikace.

Jelikož se pro předávání dat používají soubory, existují v principu dvě možnosti, jak předat tyto data. Formát souboru může být buď binární, nebo textový. Vzhledem k tomu, že soubory mají vytvářet lidé, tak by bylo krajně nepraktické, kdyby se použil binární formát souboru.

Textová reprezentace konfiguračních souborů s sebou kromě čitelnosti přináší i další výhody. Oproti binárnímu formátu není závislá na architektuře, jelikož všechna data jsou kódována ve formě textu. Navíc díky tomu, že textový soubor umožňuje kromě přenosu samotných dat i tyto data nějakým způsobem popsat, tak změny v interní struktuře Flow123d se nemusí nutně projevit ve formátu konfiguračních souborů.

1.2 Jazyky pro popis dat

Použití textového formátu konfiguračních souborů s sebou však přináší otázku, jakým způsobem tato data v textu reprezentovat. Je důležité, aby pomocí vybraného formátu bylo možné inicializovat libovolnou strukturu tříd v C++. Takové třídy obsahují atributy, které mají název (dále označován jako klíč), typ a hodnotu. Ve většině případů platí, že klíč jednoznačně implikuje typ. Potom je tedy dostačující ukládat dvojici klíč a hodnota.

Existují i situace, kdy z názvu atributu nelze jednoznačně určit jeho typ. To je způsobené použitím polymorfismu. Z klíče lze tedy odvodit pouze jakého typu musí být předek. Pokud má tento předek více potomků, pak je nutné vybraný typ explicitně uvést. Tyto situace prozatím zanedbám, jelikož se jim věnuje samostatná kapitola.

Je tedy potřeba ukládat dvojice název a hodnota. Hodnotou může být buď jednoduchého nebo složeného datového typu. Reprezentace jednoduchých datových typů je většinou triviální. Jediným problémem, či omezením, je přesnost reprezentace desetinných čísel.

Složeným datovým typem může být v jazyce C++ buď homogenní pole, nebo jiný objekt. Tím pádem vzniká hierarchická datová struktura, která může mít teore-

tický nekonečný počet vnořených úrovní. V praxi je samozřejmě počet úrovní vždy konečný, ale důležité je, aby použitý formát umožňoval reprezentovat libovolný počet vnoření.

Existuje celá řada možností. Pokud by konfigurační soubory byly triviální, dal by se použít například formát INI, který umožňuje jednoduchým způsobem zapisovat kombinace klíč a hodnota. Vzhledem k tomu, že konfigurační data pro Flow123d mohou tvořit složité hierarchické struktury, jsou však INI soubory nedostačující.

Hierarchické datové struktury lze vhodně popsat například pomocí formátu JSON, YAML nebo jazyka XML.

Flow123d v minulosti používal pro konfigurační soubory vlastní formát CON, který je podobný formátu JSON, ale mírně se od něj odlišuje. Díky tomu však nebyl příliš přenositelný, protože se pro jeho čtení nedaly přímo použít klasické knihovny, které umí číst formát JSON. To byl jeden z důvodů, proč došlo k rozhodnutí použít od verze Flow123d 2.0 jiný formát konfiguračních souborů.

1.3 Formát konfiguračních souborů

2 Analýza

2.1 Konfigurační soubory

Přiblížení konfiguračních souborů a jejich formátu.

Starý formát CON a přechod na nový formát YAML.

Diskuze XML/YAML(JSON).

2.2 Input Structure Tree

Popis formátu datových souborů.

Definice konkrétních datových typů.

Specifikovat možné atributy.

2.3 Autokonverze

Popsat možné konverze, uvést do souvislosti s XML transformacemi?

3 Návrh

Diagramy pro zpracovani YAML (inspirace SAX vs DOM).

4 Implementace

Závěr

V rámci této diplomové práce byl vytvořen editor konfiguračních souborů pro Flow123d. Jedná se o samostatně funkční aplikaci, která je ovšem navržena s ohledem na její použití jako součást softwarového balíku GeoMop, který obsahuje další nástroje, které usnadňují práci uživatelům Flow123d.

Editor uživatelům zjednodušuje vytváření a upravování konfiguračních souborů. Umožňuje ověřit správnost zadané konfigurace pro zvolenou verzi Flow123d a případně uživatele upozornit na detekované chyby. Tato funkce uživateli přináší časovou úsporu a uživatelsky příjemnější rozhraní při odhalování chyb.

Editor dále uživatelům poskytuje kontextovou dokumentaci a našeptávač. Obě tyto funkce přizpůsobují svůj obsah na základě pozice kurzoru v textu, tedy oblasti, kterou uživatel zrovna upravuje. Pro uživatele to představuje značné zjednodušení, jelikož může využít tyto funkce místo prohledávání rozsáhlé dokumentace.

V neposlední řadě editor obsahuje komponentu pro grafické znázornění datové struktury, která poskytuje alternativní pohled na zadaná data, a umožňuje rychlejší orientaci v rozsáhlých konfiguračních souborech. Kromě těchto stěžejních funkcí editor poskytuje i běžné nástroje pro manipulaci s textem, jako jsou například operace se schránkou, možnost vrácení provedených změn, vyhledávání a nahrazení nebo změna úrovně odsazení.

Aplikace je multiplatformní a podporuje systémy Windows (XP nebo novější) a Linux. S ohledem na požadavek multiplatformní aplikace byl pro vývoj použit jazyk Python 3 a grafická knihovna PyQt 5. K aplikaci byly vytvořeny instalační balíčky pro Windows a Debian.

V rámci budoucího vývoje jsou plánovány další dodatečné funkce. Jedná se např. o zlepšení zvýraznění syntaxe, které se by se mohlo přizpůsobit přímo formátu Flow123d. Další možné vylepšení spočívá v rozšíření funkcionality komponenty pro vizualizaci datové struktury. Ta by mohla v budoucnu podporovat kromě zobrazení i editaci dat nebo vylepšené zobrazení speciálních datových typů.