



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost



MINISTERSTVO ZDRAVOTNICTVÍ  
ČESKÉ REPUBLIKY

Projekt „Strategické řízení rozvoje elektronického zdravotnictví v resortu MZ“,  
registrační číslo CZ.03.4.74/0.0/0.0/15\_025/0006212,  
je spolufinancován Evropskou unií.

## ZPRACOVÁNÍ METODIK TVORBY NÁSTROJŮ PRO IMPLEMENTACI NÁRODNÍ STRATEGIE ELEKTRONICKÉHO ZDRAVOTNICTVÍ

**WP.E - Výběr a doporučení vhodného SW  
nástroje pro správu základní terminologie a  
datového slovníku elektronického zdravotnictví  
P8.1 – Doporučení SW nástrojů pro správu a  
publikaci terminologického slovníku NCeZ**



Projekt:	Strategické řízení rozvoje elektronického zdravotnictví v resortu MZ, registrační číslo CZ.03.4.74/0.0/0.0/15_025/0006212 je spolufinancován Evropskou unií		
Klíčová aktivita:	Zpracování metodik tvorby nástrojů pro implementaci Národní strategie elektronického zdravotnictví		
Datum:	12. 4. 2019	Stav:	final
Balík práce:	WP.E - Výběr a doporučení vhodného SW nástroje pro správu základní terminologie a datového slovníku elektronického zdravotnictví		
Název produktu:	P8.1 – Doporučení SW nástrojů pro správu a publikaci terminologického slovníku NCeZ		
Autor:			
Dodavatel:	Asseco Central Europe, a.s.		
Zákazník:	Ministerstvo zdravotnictví ČR		
Číslo dokumentu:		Verze:	2.0

## Schválení

Jméno	Podpis	Pozice	Datum
Ing. Martin Zeman		Sponzor projektu	12. 4. 2019
MUDr. Miroslav Zvolský		Hlavní uživatel	12. 4. 2019
Ing. Hynek Kružík		Hlavní dodavatel	12. 4. 2019
Ing. Martina Hábová, Ph.D.		Manažer projektu	12. 4. 2019

## Distribuční seznam

Jméno	Subjekt / organizační jednotka	Datum	Verze
Ing. Martin Zeman	Ministerstvo zdravotnictví ČR	12. 4. 2019	2.0
Ing. Jiří Borej	Ministerstvo zdravotnictví ČR	12. 4. 2019	2.0
Ing. Eliška Urbancová	Ministerstvo zdravotnictví ČR	12. 4. 2019	2.0



## Obsah

1	Obsah balíku práce WPE.....	4
2	P8.1 – Doporučení SW nástrojů pro správu a publikaci terminologického slovníku NCeZ .....	4
2.1	Vazba na další produkty balíku práce .....	4
2.2	Zadání pro produkt P8.1 .....	4
2.3	Úvod do problematiky .....	5
2.4	Způsob hledání relevantních nástrojů .....	5
2.5	Klasifikace nástrojů .....	5
2.6	Požadavky na výběr nástrojů .....	6
2.7	Požadavky vztažných systému elektronického zdravotnictví .....	6
2.8	Analýza nástrojů .....	6
2.8.1	Nástroje pro tvorbu glosáře .....	7
2.8.2	Nástroje pro tvorbu konceptuálního modelu .....	8
2.8.3	Nástroje pro tvorbu formálních ontologií .....	9
2.9	Sumarizace analýzy .....	9
2.9.1	Nástroje pro tvorbu glosáře .....	9
2.9.2	Nástroje pro tvorbu ontologického model .....	10
2.9.3	Nástroje pro tvorbu formálních ontologií .....	10
2.10	Doporučení SW nástrojů pro implementaci v ČR .....	10
2.10.1	Nevýhody řešení a doporučení k další implementace .....	11
2.10.2	Návrh organizačního zajištění .....	11
2.10.3	Instalace doporučené sady nástrojů .....	11
2.11	Reference .....	12



## 1 Obsah balíku práce WPE

Součástí balíku práce WP.E - Výběr a doporučení vhodného SW nástroje pro správu základní terminologie a datového slovníku elektronického zdravotnictví jsou celkem dva produkty:

- **P8.1 - Doporučení SW nástrojů pro správu a publikaci terminologického slovníku NCEZ**
- P8.2 - Doporučení SW nástrojů pro správu a publikaci ostatních standardů a terminologií pro účely elektronizace zdravotnictví

## 2 P8.1 – Doporučení SW nástrojů pro správu a publikaci terminologického slovníku NCeZ

Tento dokument obsahuje první z uvedených produktů balíku práce WPE, produkt **P8.1 - Doporučení SW nástrojů pro správu a publikaci terminologického slovníku NCeZ**.

### 2.1 Vazba na další produkty balíku práce

Vstupem pro tento balík práce je produkt P10 - „Metodika správy terminologie elektronického zdravotnictví“ balíku práce WPF, a produktu P7 - „Analýza zahraničních vzorů a návrh metodiky správy standardů“.

### 2.2 Zadání pro produkt P8.1

Produkt bude zpracován jako dokument MS Word obsahující tyto části:

- Seznam kritických požadavků na SW nástroje. Seznam bude vytvořen v následujících krocích:
  - Identifikace základních vztažených systémů, resp. oblastí elektronického zdravotnictví
  - Identifikace, analýza a prioritizace kritických požadavků na SW nástroje podporující procesy pro správu terminologického slovníku NCeZ, vycházející z výstupu balíku práce WPF, produktu P10 – „Metodika správy terminologie elektronického zdravotnictví“, včetně požadavků ekonomických.
  - Identifikace požadavků vztažených systémů na způsob provázání se systémem správy základních terminologií elektronického zdravotnictví.
- Posouzení existujících SW nástrojů v této oblasti, porovnání existujících a dostupných nástrojů z pohledu kritických a ekonomických požadavků. Součástí porovnání bude:
  - Popis způsobu hledání potenciálních kandidátů SW nástroje včetně metody pro výběr finálních kandidátů.
  - Analýza vybraných kandidátů na SW nástroje (analýza funkčních vlastností).
  - Evaluace vybraných kandidátů dle identifikovaných kritických a ekonomických požadavků.
- Doporučení pro výběr vhodného nástroje z hlediska aplikovatelnosti v podmínkách ČR a návrh organizačního zajištění.
  - Sumarizující porovnání nejvhodnějších kandidátů.
  - Doporučení SW nástroje pro implementaci v ČR.
  - Návrh organizačního zajištění implementace nástroje.
  - Harmonogram pro implementaci SW nástroje včetně implementačního průvodce.



## 2.3 Úvod do problematiky

V rámci aktivit NCeZ se předpokládá tvorba sémantických slovníků vytvářených pomocí metodiky popsané v balíčku WPF, produktu P10. Metodika vychází z dokumentu Koncepce sémantického slovníku pojmů [1] a propojené slovníky označuje společně pod názvem Sémantický slovník NCeZ. Publikace slovníků je založená na principech propojených dat<sup>1</sup>, která je určena zejména pro podporu kvalitní publikace otevřených dat. Proces tvorby slovníků se zakládá na standardech Sémantického webu jako jsou jazyky RDF, SKOS, OWL a na jazyku OntoUML pro tvorbu diagramů sémantických modelů. Každý slovník pozůstává, ze dvou částí – glosář a ontologický model. Glosář popisuje základní strukturu pojmosloví pomocí standardu SKOS. Ontologický model zachycuje význam pojmů a vztahů mezi pojmy glosáře. Je vytvářen pomocí diagramů v jazyce OntoUML, které jsou následně transformovány do jazyka OWL a publikovány jako propojená data spolu s glosářem.

Cílem dokumentu je analyzovat existující nástroje a nabídnout sadu nástrojů pro realizaci metodiky produktu P10, včetně popisu chybějících funkcí a návrhy na rozšíření těchto nástrojů.

## 2.4 Způsob hledání relevantních nástrojů

Pro analýzu a výběr relevantních nástrojů pro glosář byly použity výsledky rešerše SKOS nástrojů z roku 2017 [1] doplněny o nástroje doporučené expertem v této oblasti<sup>2</sup>. Pro selekci relevantních modelovacích nástrojů konceptuálního modelování byl použit seznam existujících OntoUML nástrojů zveřejněný na portále [ontouml.org](http://ontouml.org)<sup>3</sup>. Pro selekci relevantních nástrojů pro správu formálních ontologií byl použit seznam ontologických editorů ze stránek mezinárodního konsorcia W3C<sup>4</sup>.

## 2.5 Klasifikace nástrojů

Tato sekce slouží jako přehled různých typů nástrojů, které přímo nebo nepřímo souvisí s tvorbou slovníků. Pro účely analýzy nástrojů v rámci metodiky produktu P10 je výhodné typy nástrojů rozdělit dle rolí, které v metodice participují na<sup>5</sup>:

- *Nástroje pro správu glosáře* – je určený zejména pro doménového experta a slouží jako hlavní rozhraní pro komunikaci se slovníkovým expertem.
- *Nástroje pro správu konceptuálních modelů* – je určený zejména pro slovníkového experta a ontologického inženýra. Slouží jako hlavní rozhraní pro komunikaci s mezi těmito rolemi.
- *Nástroje pro správu formálních ontologií* – je určený výhradně pro ontologického inženýra.

Pro širší přehled typu nástrojů v kontextu tvorby slovníků se lze inspirovat klasifikací SKOS nástrojů [1], která byla doplněna a upravena pro účely této analýzy. I když konkrétní nástroje mohou být uvedeny ve vícero kategoriích v následujícím textu jsou uvedeny v každé kategorii jenom jednou, dle jejich hlavních zaměření:

- **Úložiště** – databáze pro ukládání a dotazování se nad RDF datami, jako např. Eclipse RDF4J, Ontotext GraphDB, Jena TDB, Blazegraph, Stardog, AllegroGraph, Apache Marmotta, Linked Media Framework, Mulgara Semantic Store.
- **Editory** – programy umožňující tvorbu a správu: a) *glosářů v jazyce SKOS* jako např.: iQvoc, Intelligent Topic Manager, Lexaurus editor, Open Metadata Registry, Poolparty, SKOSed, SKOSjs, SKOS Shuttle,

<sup>1</sup> Linked Data Principles (2009), dostupné online na <https://www.w3.org/DesignIssues/LinkedData.html>

<sup>2</sup> V rámci této aktivity byly prohledány i open-source projekty z webových služeb <http://github.com> a <https://bitbucket.org>.

<sup>3</sup> <https://ontouml.org/ontouml/tooling>

<sup>4</sup> [https://www.w3.org/wiki/Ontology\\_editors](https://www.w3.org/wiki/Ontology_editors)

<sup>5</sup> Pro jednoduchost neuvažujeme nástroje pro tvorbu specifikace slovníků, kde ve většině případů postačí jenom textový dokument.



Tematres, ThesaurusAPI, ThManager, TopBraid EVN, xTree, VocBench, Termlt; b) *formálních ontologií v jazyku OWL* jako např. Protege, TopBraid Composer, SWOOP, NeOn Toolkit; c) *konceptuálních modelů založených na jazyku OntoUML* jako např. OLED, Menthor, Modelio s OntoUML modulem.

- **Navigátory** – programy, které umožňují vizualizaci a exploraci dat vyjádřených pomocí formátů Sémantického webu, a to zejména v souvislosti s propojenými daty anebo vyjádřeny v SKOS formátu. Mezi tyto programy patří: Skosmos, ASKOSI, Callimachus, Finnish Ontology Library Service ONKI, Rhizomer, SKOS\_WS, SKOS Reader.
- **Extraktory** – programy které umožňují automatizovanou extrakci dat, nebo pojmů z dokumentů a jejich mapování na metadatová schémata jako je SKOS. Příklady těchto programů jsou: Helping Interdisciplinary Vocabulary Engineering (HIVE), Poolparty Semantic Search, Poolparty Extractor, SemanticTurkey, SKOSSY.
- **Konvertory** – programy kterých hlavní funkce je změna formátu dat (např. mezi jazyky OWL a SKOS): Model Futures SKOS Exporter, OWLtoSKOS, SKOS Play!, SKOS2OWL, Skosify, XL2XML.
- **Validátory** – nástroje pro evaluaci kvality: a) *SKOS slovníků* jako např. qSKOS, Poolparty SKOS Quality Checker; b) *OntoUML konceptuálních modelů* – existují jenom integrované v OntoUML editorech; c) *OWL formálních ontologií* jako např. Pellet OWL Reasoner Validator.

## 2.6 Požadavky na výběr nástrojů

Správa významů je časově náročná a vyžaduje adekvátní nástroje, které umožní

- identifikovat pojmy z existujících legislativních dokumentů,
- tyto významy katalogizovat a umožnit jejich správu a vývoj v čase,
- katalogizovat formální vztahy mezi významy a umožnit jejich správu a vývoj v čase, a
- umožnit jejich zveřejnění ve formě dostupné pro člověka (který tím např. zjistí, jaký význam a vazbu na legislativní pojmy má datová sada, kterou má k dispozici), i pro strojové zpracování (např. pro vyhledávače datových sad využívající formálních vztahů mezi významy pro získání relevantnějších výsledků).
- využít významy pro anotování digitálních zdrojů a jejich následné vyhledávání

## 2.7 Požadavky vztahů systému elektronického zdravotnictví

Základní vztahované systémy elektronického zdravotnictví, resp. oblastí elektronického zdravotnictví jsou:

- Systém NZIS (Národní zdravotnický systém),
- Katalog standardů,
- IDRR – Integrované datové rozhraní resortu,
- Základní terminologický slovník (slovníky),
- Infrastruktura elektronického zdravotnictví a všechny její komponenty,
- Služby elektronického zdravotnictví: Index zdravotnické dokumentace, Národní zdravotnický informační portál (NZIP), Autoritativní zdravotnické registry (ARPZS, ARZP, ARP), Registr mandátů a souhlasů, Systém zabezpečené komunikace.

Kompletní sémantické slovníky definují životní cyklus objektů, které dané pojmy reprezentují. Pro integraci těchto systémů je důležité, aby vztahové systémy definovali identitu těchto objektů respektující tohoto životního cyklu.

## 2.8 Analýza nástrojů

Tato sekce analyzuje vybrané nástroje pro správu glosáře, konceptuálního modelu a formálních ontologií.



## 2.8.1 Nástroje pro tvorbu glosáře

### PoolParty Semantic Suite

PoolParty Semantic Suite<sup>6</sup> je profesionální sada nástrojů pro organizaci znalostí a obsahu. Sada nástrojů je zaměřena především na návrh a správu slovníků založených na textovém korpusu a jeho správě. Slovník může být použit pro fasetové vyhledávání, nebo jako zdroj propojených dat. Podpora pokročilého odvozování znalostí (OWL, SWRL, SHACL) je slabá. Hlavními nástroji sady jsou PoolParty (Basic/Advanced/Enterprise) Server and PoolParty Semantic Integrator, který poskytuje služby jako např. správa taxonomií a thesaurů, tagování konceptů, správa propojených dat, správa ontologií, sémantické vyhledávání, dolování z textu a rozpoznávání entit, doporučování obsahu, datovou analytiku, integraci dat a vizualizaci dat. Sada nástrojů také umožňuje integraci se službou SharePoint, MarkLogic, Drupal a Confluence. Časově neomezená on-site licence pro PoolParty Enterprise Server je 62500 eur, nebo 3750 eur za měsíc pro cloudovou službu. Ostatní varianty licence jsou levnější.

### TopBraid Enterprise Data Governance

TopBraid Enterprise Data Governance (EDG)<sup>7</sup> je hlavní produkt společnosti TopQuadrant (navíc k nástroji TopBraid Composer). Je to složitý nástroj, který umožňuje modelovat SKOS slovníky, OWL ontologie, referenční datové sady a jiné. Systém podporuje řízení oprávnění a také podnikové vyhledávání digitálních zdrojů dat. Dále podporuje definici SHACL omezení a jejich ověřování. Systém se zdá být silný v správě dat a slovníku a integraci dat. Všechny operace uživatelského rozhraní (tedy operace vytváření/aktualizace/mazání) jsou zaznamenány a jsou dotazovatelné. Informace o ceně nejsou na webu k dispozici.

### Vocbench

Vocbench 3<sup>8</sup> je open-source manažer SKOS slovníků a OWL ontologií. Vychází z jeho předchůdce Vocbench 2<sup>9</sup>, který poskytoval neintuitivní a pomalé uživatelské rozhraní. Vocbench 3 je kompletní přepis nástroje, který je mnohem responsivnější. Nástroj poskytuje omezení přístupu založené na rolích, vícejazyčné projekty, schvalovací pracovní postup, historii úprav, generování formulářů, podporu importů Excel souborů. Editace OWL ontologií je jenom základní bez podpory odvozování.

### Termit

Termit<sup>10</sup> je open-source nástroj (s GPL v3.0 licenci<sup>11</sup>) pro systematickou správu odborných významových slovníků, modelování významu pojmů, a jejich využití pro sdílení a vyhledávání digitálních zdrojů organizací. Poskytuje funkce:

- vytváření a správu glosářů a pojmů,
- základní modelovací konstrukty OntoUML (např. přiřazování UFO typů k pojmům a jejich vztahům),
- publikaci glosářů formou propojených dat,
- správu digitálních zdrojů dat (např. HTML dokumenty, odkazy na webové stránky, nebo RDF datové zdroje),

<sup>6</sup> <https://www.poolparty.biz>

<sup>7</sup> <https://www.topquadrant.com/products/topbraid-enterprise-data-governance/>

<sup>8</sup> <http://vocbench.uniroma2.it/>

<sup>9</sup> Vývoj Vocbench 2 se zastavil začátkem roku 2017 viz. <https://bitbucket.org/art-uniroma2>

<sup>10</sup> <https://github.com/kbss-cvut/termit>

<sup>11</sup> <http://www.gnu.org/licenses/gpl.html>



- anotování digitálních zdrojů dat a obsahu pojmy glosáře (aktuálně je podporován jenom obsah HTML dokumentů),
- textovou analýzu obsahu HTML zdrojů (např. vyhledávání klíčových slov v textu, doporučení výskytů pojmů glosáře v textu),
- vyhledávání datových zdrojů a slovníků,
- fultextové a fasetové vyhledávání pojmů všech slovníků,
- analytické a statistické ukazatele pojmů a slovníků (např. statistiky pojmů v slovnících, statistiky výskytů pojmů v anotovaných zdrojích a jejich obsahu, statistiky výskytů jednotlivých UFO typů ve slovníku),
- export glosáře (CSV, Excel, RDF).

## 2.8.2 Nástroje pro tvorbu konceptuálního modelu

### OLED

OntoUML lightweight editor (OLED)<sup>12</sup> je prostředí založené na editore TinyUML určené pro vývoj, vyhodnocování a implementaci doménových ontologií založené na OntoUML. Nástroj poskytuje lehkou a integrovanou sadu funkcí pro ontologické inženýry, jako je například syntaktická a sémantická verifikace modelu, odvozování z modelu, vizuální simulace modelu, automatické rozpoznání sémantických antivzorů s návrhy jejich oprav, validace vztahů dle teorie části a celku, validace ontologických vzorů. OLED podporuje import a export modelů nástroje Sparx System Enterprise Architect verze 10, který je realizován pomocí OntoUML zásuvného modulu<sup>13</sup>. V roce 2015 byl přispěvateli OLEDu založen nový projekt – editor Menthor, který je popsán níže. Vývoj editoru OLED se tím zastavil.

### Menthor

Menthor<sup>14</sup> editor je platforma pro vytváření, ověřování a implementaci ontologií, která je někdy označována jako nástupce OLED editoru. Kromě modelování v OntoUML Menthor editor podporuje:

- deklarativní jazyk OCL pro vyjádření omezení UML modelů, pro které neexistuje notace v UML diagramech,
- generování kódu do RDF, OWL a UML,
- nástroje pro verifikaci modelu včetně vizuální simulace modelu, automatické rozpoznání sémantických antivzorů s návrhy jejich oprav, ověření syntaxe,
- nástroje dokumentace – export do SBVR<sup>15</sup>,
- integraci s dalšími nástroji pro modelování<sup>16</sup> včetně Sparx System Enterprise Architect.

### Modelio editor

Modelio<sup>17</sup> je modelovací prostředí s otevřeným zdrojovým kódem, podporující hlavní standardy včetně UML, BPMN, Archimate, MDA, SysML. Poskytuje následující služby:

- import a export XML (umožňuje výměnu UML2 modelů mezi různými nástroji)
- publikování modelu v HTML formátu
- generování Java kódu

<sup>12</sup> <https://github.com/nemo-ufes/ontouml-lightweight-editor>

<sup>13</sup> <https://code.google.com/archive/p/ontouml-lightweight-editor/wikis/EnterpriseArchitect.wiki>

<sup>14</sup> <https://github.com/MenthorTools/menthor-editor>

<sup>15</sup> <https://www.omg.org/spec/SBVR/1.0>

<sup>16</sup> <https://github.com/MenthorTools>

<sup>17</sup> <https://www.modelio.org/>





- skriptovací podpora v jazyku Jython

Pro modelování v OntoUML jazyku existuje OntoUML zásuvní modul<sup>18</sup>. Podporuje rozšíření jazyka OntoUML, které bylo použito pro tvorbu Sémantického slovníku veřejné správy (SGoV). Navíc umožňuje tyto modely exportovat do RDF ve formátu ttl<sup>19</sup>.

### 2.8.3 Nástroje pro tvorbu formálních ontologií

#### Protege

Protégé<sup>20</sup> je široce používaný open-source nástroj pro vývoj ontologie ve formátu OWL s možností využití SWRL<sup>21</sup> pravidel. Nástroj se aktivně vyvíjí a obsahuje mnoho užitečných zásuvných modulů (ontologické odvozování, validace, automatizovaná oprava chyb, dotazování, porovnávání, vizualizace, mapování databáze, dokumentace, import z aplikace Excel atd.).

#### Topbraid Composer

TopBraid Composer<sup>22</sup> je komerční IDE pro transformaci RDF dat a vývoj ontologií. Většina jeho konfigurace je provedena pomocí RDF (také reprezentace dotazů SPARQL v rámci SPIN notace), a proto jsou znalosti snadno refaktorovatelné, dotazovatelné a auditovatelné. TopBraid Composer v jeho edici Maestro podporuje validaci omezení SHACL<sup>23</sup>, vizualizace a provádění datových proudů plně reprezentovaných zastoupena v RDF. TopBraid Composer v edici Maestro stojí \$3450 ročně.

## 2.9 Sumarizace analýzy

Dle analýzy nástrojů lze uvažovat o třech alternativních sadách nástrojů – komerční řešení založené na PoolParty Semantic Suite, komerční řešení společnosti TopQuadrant a open-source řešení založené na nástrojích TermIt, Modelio, Protege. Detaily jsou popsány v následujících podsekcích.

### 2.9.1 Nástroje pro tvorbu glosáře

Pro vývoj glosáře se nabízí dvě komerční řešení (PoolParty Semantic Suite, TopBraid Enterprise Data Governance) a dvě open-source řešení (Vocbench, TermIt) s volnou licencí. Silné stránky PoolParty Semantic Suite je kompletní podpora tvorby slovníků s integrací pro různé systémy v rámci organizace a zejména propojení s textovými dokumenty. TopBraid Enterprise Data Governance oproti PoolParty Semantic Suite poskytuje však lepší možnosti, co se týče integrace na úrovni datových zdrojů a lepší možnosti transformace RDF dat, včetně jejich publikace. Navíc pro formalizaci ontologií lze použít integrovaný nástroj Topbraid Composer.

Vocbench oproti nástroji TermIt poskytuje omezení přístupu založené na rolích, vícejazyčné projekty, schvalovací pracovní postup a propracovanější historii úprav. Tyto funkce však nejsou kritické pro správu sémantických slovníků.

TermIt oproti Vocbench poskytuje lepší podporu pro tvorbu slovníků z obsahu dokumentů a anotaci datových zdrojů. V porovnání s ostatními nástroji TermIt nabízí plné řešení pro tvorbu glosářů kompatibilních se

<sup>18</sup> <https://store.modelio.org/resource/modules/ontouml.html>

<sup>19</sup> <https://www.w3.org/TR/turtle/>

<sup>20</sup> <https://protege.stanford.edu>

<sup>21</sup> Semantic Web Rule Language -- <https://www.w3.org/Submission/SWRL/>

<sup>22</sup> <https://www.topquadrant.com/products/topbraid-composer/>

<sup>23</sup> Shapes Constraint Language -- <https://www.w3.org/TR/shacl/>



Sémantickým slovníkem pojmů. Z kritických požadavků jde zejména o vytváření základních OntoUML konstruktů slovníků, fasetové vyhledávání v pojmech dle UFO typů a export glosáře kompatibilním se slovníkem SGoV.

### 2.9.2 Nástroje pro tvorbu ontologického model

Menthor v porovnání nástrojem Modelio poskytuje komplexní sadu pro verifikaci modelu včetně vizuální simulace modelu, automatické rozpoznání sémantických antivzorů s návrhy jejich oprav a ověření syntaxe. Neumožňuje však vyjádřit všechny konstruktů rozšířeného jazyka OntoUML pro účely SGoV kompatibilních slovníků a neumožňuje spravovat vícero slovníků, což je kritické pro jejich mapování.

Modelio oproti Mentoru obsahuje kritické požadavky pro tvorbu SGoV kompatibilních slovníků – podporu exportu v podobě glosáře a modelu (tedy formalizované ontologie) a podporu rozšířeného jazyka OntoUML. Chybějící možnosti validace modelu lze nahradit expertizou ontologického inženýra a formalizaci do ontologií.

### 2.9.3 Nástroje pro tvorbu formálních ontologií

Topbraid Composer je vhodný nástroj v případě použití nástroje pro tvorbu glosářů TopBraid Enterprise Data Governance. V ostatních případech je Protege vhodnější díky komplexnější podpoře v ontologickém odvozování, validaci a automatizované opravě chyb OWL ontologií s podporou SWRL pravidel.

## 2.10 Doporučení SW nástrojů pro implementaci v ČR

Dle analýzy existujících nástrojů byly navrženy tři sady nástrojů. Z těchto řešení jedině open-source sada nástrojů pokrývá kritické požadavky pro tvorbu sémantických slovníků. U komerčních sad nástrojů by použití vyžadovalo další vývoj a uzavřenost zdrojových kódů tuto situaci kompiluje. Lze však v případě potřeby uvažovat o postupném nahrazování vytvořené open-source sady nástrojů komerčními nástroji, jak bylo popsáno v souhrnné sekci analýzy nástrojů. Pro implementaci v ČR byla vybrána sada nástrojů znázorněna v následujícím obrázku:

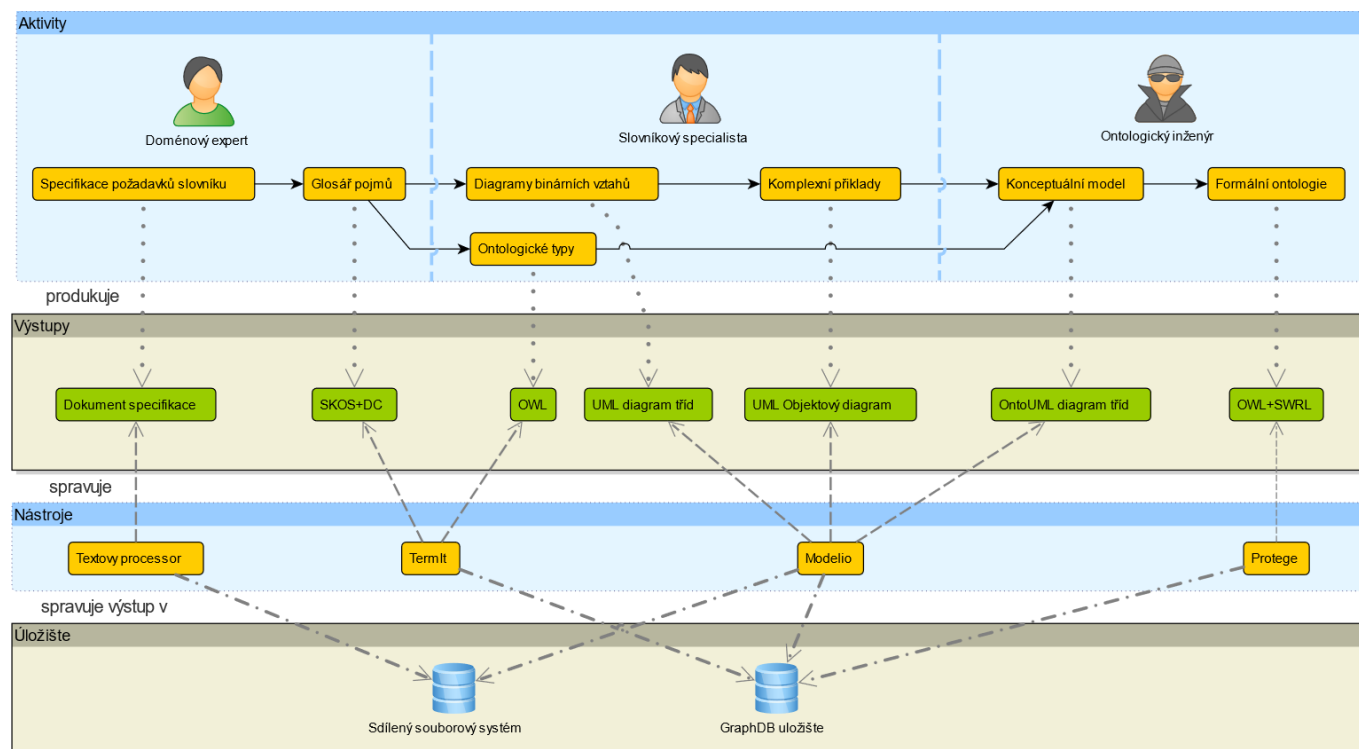


Diagram je rozdělen do čtyř vrstev:



- aktivity – jde o aktivity popsané v metodologii produktu P10. Ty jsou rozděleny dle implicitní přiřazení těchto aktivit k rolím.
- výstupy – jsou produkty jednotlivých aktivit.
- nástroje – spravují jednotlivé výstupy z aktivit.
- úložiště – slouží pro synchronizaci mezi nástroji a jednotlivé nástroje svoje výstupy v těchto úložištích spravují.

### 2.10.1 Nevýhody řešení a doporučení k další implementaci

Navržená sada nástrojů obsahuje nedostatky, které nejsou kritické pro tvorbu slovníků, avšak snižují jeho efektivitu. Pro zefektivnění práce lze uvažovat o následujícím rozšíření implementace uspořádaném dle priority:

- *synchronizace konceptuálního modelu a glosáře* – sada nástrojů podporuje synchronizaci modelu a glosáře v směru od modelu ke glosáři. Opačný směr je potřeba aktuálně spravovat manuálně.
- *rozšíření transformace do formální ontologie* – sada nástrojů umožňuje jenom základní transformaci konceptuálního modelu do formální ontologie. To neumožňuje zachytit některé modelovací chyby automaticky a přináší to více práce ontologickému inženýrovi.
- *omezení přístupu tvorby glosáře* – nástroj TermIt aktuálně neumožňuje omezení tvorby glosáře. Lze však uvažovat o omezení dle rolí, omezení práv jenom na konkrétní slovníky a přístup jenom pro čtení.

U prvních dvou bodů implementace kromě znalosti vývoje software vyžaduje i znalosti v oblasti ontologického inženýrství.

### 2.10.2 Návrh organizačního zajištění

Metodika P10 navrhuje tři typy slovníků s narůstajícími nároky na proces tvorby – jednoduchý slovník, formalizovaný slovník a kompletní sémantický slovník. Pro tvorbu formalizovaných slovníků je vhodné školení pro slovníkového specialistu v rozsahu 5 hodin. Předpokládá se, že efektivita tvorby formalizovaných a kompletních sémantických slovníků se bude zvyšovat díky nově vznikajícím slovníkům, které poskytují příklady modelování podobných termínů a také díky vybudování vrchní vrstvy abstraktních pojmů, které umožní rychlejší rozpoznání podstaty významu pojmu. Ze stejného důvodu je vhodné vytvořit pro pracovní skupinu startovací sadu slovníků různých typů (tedy příklady legislativních, agendových a datových slovníků) v jejich expertní doméně, nebo v doméně obecně známe.

### 2.10.3 Instalace doporučené sady nástrojů

Instalace doporučené sady nástrojů pozůstává z následujících kroků:

- Instalace úložiště GraphDB – instalační manuál je k dispozici na webové stránce <http://graphdb.ontotext.com/documentation/standard/installation.html>.
- Instalace nástroje TermIt – TermIt je webová aplikace běžící nad webovým serverem. Podrobný manuál včetně konfigurace je popsán <https://github.com/kbss-cvut/termit>. Nástroj je potřeba nakonfigurovat oproti uložišti vytvořeném v předcházejícím kroku. Nástroj obsahuje také skript pro transformaci modelů z nástroje Modelio do úložiště dat. Podrobný popis skriptu lze najít ve podsložce projektu „src/main/resources/transform-sgov“.
- Instalace nástroje Modelio – instalační manuál je k dispozici na webové stránce <https://www.modelio.org/downloads/download-modelio.html>.
- Instalace OntoUML modulu pro nástroj Modelio – instalační manuál lze najít na webové stránce <https://kbss.felk.cvut.cz/gitblit/summary/ontouml-modelio.git>. Po instalaci lze exportovat model pomocí kontextového menu konkrétního UML balíčku, který obsahuje



- Instalace nástroje Protege – instalační manuál je k dispozici na webové stránce [https://protegewiki.stanford.edu/wiki/Install\\_Protege5](https://protegewiki.stanford.edu/wiki/Install_Protege5).

## 2.11 Reference

[1] Koncepce sémantického slovníku pojmů. (2018). Vytvořeno v rámci Implementace strategií v oblasti otevřených dat II CZ.03.4.74/0.0/0.0/15\_025/0004172. Online na [https://opendata.gov.cz/media/dokumenty:sémantický-slovník-pojmů:c1v2d1\\_návrh\\_koncepce\\_sémantického\\_slovníku\\_pojmů.pdf](https://opendata.gov.cz/media/dokumenty:sémantický-slovník-pojmů:c1v2d1_návrh_koncepce_sémantického_slovníku_pojmů.pdf).

[2] Mochón, G., Méndez, E. M., & Bueno de la Fuente, G. (2017). 27 pawns ready for action: A multi-indicator methodology and evaluation of thesaurus management tools from a LOD perspective. *Library Hi Tech*, 35(1), 99-119.

[3] Jackson, M., Crouch, S., & Baxter, R. (2011). *Software evaluation: criteria-based assessment*. Software Sustainability Institute.

Web strategie: <http://www.nsez.cz>

Toto dílo podléhá licenci Creative Commons CC BY 4.0. Dílo je možné libovolně šířit a upravovat za předpokladu uvedení citace tohoto díla. Pro zobrazení podrobných licenčních podmínek navštivte <http://creativecommons.org/licenses/by/4.0/>. Licence se nevztahuje na použití loga Ministerstva zdravotnictví České republiky mimo reprodukci tohoto díla. Veškerá práva k logu jsou vyhrazena.

Citace dle ČSN ISO 690:2011:

MINISTERSTVO ZDRAVOTNICTVÍ ČESKÉ REPUBLIKY. *Doporučení SW nástrojů pro správu a publikaci terminologického slovníku NCeZ*. Verze 2.0. Praha, 2019. Licencováno pod CC BY 4.0, licenční podmínky dostupné z: <http://creativecommons.org/licenses/by/4.0/>.

