

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

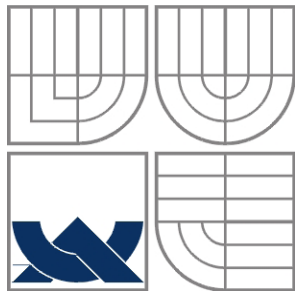
GENEALOGICKÁ SÉMANTICKÁ WIKI

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

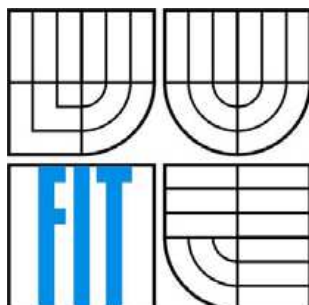
AUTOR PRÁCE
AUTHOR

Bc. JANA BRYCHOVÁ

BRNO 2009



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

GENEALOGICKÁ SÉMANTICKÁ WIKI

GENEALOGIC SEMANTIC WIKI

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. JANA BRYCHOVÁ

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. MAREK SCHMIDT

BRNO 2009

Abstrakt

Tato diplomová práce se zabývá možnostmi uložení genealogických dat do různých formátů a na základě výsledků navrhuje vhodný formát dat takový, aby byla následně možná jejich prezentace pomocí sémantického webu. V rámci této práce je implementována genealogická aplikace pro platformu KiWi. Tato aplikace umožňuje vizualizaci navrženého formátu pomocí technologie prefuse. V dokumentu jsou uvedeny základní a případně další potřebné informace ohledně základů technologií sémantického webu jako RDF, XML, ontologie, jazyk OWL. Dále dokument popisuje implementovanou aplikaci včetně použití .

Abstract

This thesis project is about possibilities of storing genealogic data in different formats and based on the results it suggests data format which can be used consequently as a source for the visualization by semantic web. In the scope of the project there is a genealogic application implemented for KiWi platform. This application enable visualization of designed format using prefuse technology. There are basic and aslo some other usefull information about the basic technologies of the semantic web like RDF, XM, ontology, OWL language stated in the document.

Klíčová slova

Genealogie, Sémantický web, RDF, GEDCOM, OPC, Family.Show, XML, OWL, Ontologie, KiWi, prefuse, JBoss Seam

Key words

Genealogy, Semantic web, RDF, GEDCOM, OPC, Family.Show, XML, OWL, Ontology, KiWi, prefuse, JBoss Seam

Prohlášení

Prohlašuji, že jsem vypracovala diplomovou práci samostatně pod vedením Ing. Marka Schmidta. Uvedla jsem všechny literární prameny a publikace, ze kterých jsem čerpala.

Poděkování

Chtěla bych touto cestou poděkovat mému diplomovému vedoucímu Marku Schmidtovi za odborné vedení, za seznámení s mnoha pro mě novými technologiemi, i za trpělivost při vedení.

Obsah

Úvod.....	7
Genealogie.....	8
Definice genealogie.....	8
Genealogie a práva.....	8
Genealogie a správnost dat.....	8
Vysvětlení pojmů z genealogie.....	8
Jedinec v genealogii.....	11
Možné vztahy v genealogii.....	11
Formát dat pro genealogii.....	12
Formát GEDCOM (GEnealogic Data COMMunication).....	12
Možnost použití formátu dle jeho obecných vlastností.....	12
Popis formátu.....	12
Ukázka formátu (zdroj [3]).....	12
Možnost použití formátu dle jeho technických vlastností.....	13
OPC formát aplikace Family.Show.....	13
Možnost použití formátu dle jeho obecných vlastností.....	14
Popis formátu.....	14
Ukázka formátu (vlastní, vytvořeno pomocí aplikace Family.Show).....	14
Možnost použití formátu dle jeho technických vlastností.....	16
Formát GEDCOM 6.0.....	16
Příklad formátu pro událost (zdroj [6]).....	16
Možnost využití formátu GEDCOM 6.0.....	17
Sémantický web.....	18
Definice Internetu.....	18
Definice Webu.....	18
Definice URI.....	18
Sémantický web.....	18
Historie sémantického webu.....	18
Sémantický web podle Berners-Leeho.....	19
Reprezentace znalostí.....	19
Ontologie.....	20
Agenti sémantického webu.....	20
Metadata.....	20
RDF a XML.....	20
Příklad RDF.....	21
Jazyk OWL.....	21
Návrh RDF pro genealogii.....	21
Úplně nový formát verus formát navazující na stávající.....	22
GEDCOM versus OPC.....	22
Genealogická data v RDF.....	22
Definice jmenných prostorů pro vlastnosti.....	22
Základní genealogická data.....	23
Data standardní údaje pro každé RDF v KiWi.....	24

Genealogická data vhodná k uchování	24
Další data, která je možné uchovávat	25
KiWi – Knowledge in a Wiki	27
Základní princip aplikace KiWi	27
První release aplikace Kiwi	27
KiWi-Vision	28
KiWi-System.....	28
Studie použití	28
Extensions.....	29
Grant pro KiWi projekt.....	29
Návrh aplikace implementované v rámci této diplomové práce	29
Obrazovky.....	29
Home – základní obrazovka, rozcestník.....	29
Create – obrazovka pro vytvoření nové osoby	30
Update – obrazovka pro změnu údajů o existující osobě	31
Display – zobrazení rodokmenu/stromu života	32
Algoritmus pro vizualizaci	33
Co má být zobrazeno a proč	33
Princip algoritmu pro výběr členů rodiny k zobrazení.....	33
Inicializace seznamů	34
Posloupnost přidávání osob v rámci generace.....	34
Posloupnost přidávání osob mezigeneračně	35
Pomocné funkce pro přidávání osob do seznamů	35
Vytvoření seznamů rodin	36
Algoritmus a Prefuse.....	36
Ad krok 1: Algoritmus pro nahrání dat do datových struktur nástroje perfuse	37
Ad krok 4: Vytvoření akcí pro zpracování dat.....	37
Použité technologie a nástroje	38
Java – programovací jazyk	38
Java Enterprise Edition.....	38
JBoss Seam	39
KiWi	39
Java Applet.....	39
Eclipse	40
Apache Ant	40
Prefuse – visualization toolkit.....	41
Přínosy, možná rozšíření a použití v praxi	41
Přínosy diplomové práce	41
Možná rozšíření diplomové práce.....	41
Možná rozšíření aplikace.....	42
Použití aplikace v praxi	42
Závěr	43
Literatura	44

Úvod

Mnoho lidí se přirozeně zajímá o svůj původ. Až do nedávné doby patřila genealogie spíše do rukou historiků a věnovala se převážně šlechtickým rodům. V současnosti je z ní pomocná věda historická, která je nejen koníčkem dostupným pro širokou veřejnost, ale má i praktické uplatnění. Vzhledem k možné automatizaci ve vyhledávání historických dat a hlavně jejich souvislostí se genealogické aplikace staly značně oblíbené. Dalším významným faktem je, že největším zdrojem dat dnešní doby je bezesporu Internet - a nejen to, na Internetu také najdeme webové aplikace, které jsou dostupné široké veřejnosti. Je tedy celkem logickým krokem umožnit webové vyhledávání genealogických dat na základě souvislostí.

Tato diplomová práce se proto zabývá vytvořením aplikace na základě zatím nepříliš rozšířené technologie sémantického webu, který umožňuje větší provázanost dat. Tato provázanost se silně uplatňuje v reprezentaci dat genealogických. Vzhledem k tomu, že o sémantickém webu se v současnosti mluví jako o budoucnosti webu, je jistě logickým krokem použití této technologie pro vytvoření takové aplikace.

Aplikace jako taková bude poskytovat možnosti zadávání, editace a zobrazování genealogických dat. V každém případě je zde ale ve hře jedna velice aktuální otázka, a to otázka ochrany osobních údajů.

Genealogie

Definice genealogie

Podle Slovníku cizích slov je genealogie vědou, která zkoumá vztah dvou souvisejících jevů nebo faktů jako rodokmen, rodopis, věda o rodové posloupnosti.

Podle Wikipedie[3] je genealogie definována jako pomocná historická věda, která zkoumá vztahy mezi lidskými jedinci, vyplývající z jejich společného rodového původu. Genealogie se zabývá buď studiem jednotlivých osobností, nebo sledováním proměn jednotlivých druhů vztahů. Soukromý zájem v této oblasti bývá označován jako rodopis.

V některých zdrojích se lze dočíst, že genealogie je pomocná věda historická, zabývající se šlechtickými rody. V dnešní době však nejsou v hledáčku historiků pouze šlechtivé rody, ale zabývají se veškerým obyvatelstvem, navíc se genealogie rozšířila i mimo historiky, většinou spíše jako koníček lidí, kteří se zajímají o své předky, o svůj rodokmen, svůj rodopis.

Genealogie a práva

Genealogie je věda, o jejíž poznatky se opírají právní instituce, které využívají její data například při rozhodování v dědických řízeních, ale i dalších případech týkajících se převážně rodinného práva.

Genealogie a správnost dat

Někdy dochází k falzifikaci dat, nebo se mohou objevit také smyšlená data. Tohoto se lidé dopouštějí například kvůli získání větší prestiže, popularity nebo k obohacení se neprávem o cizí dědictví. Tato diplomová práce se však nezabývá ověřováním pravosti dat, je tedy možné, že některá data uvedená v práci jsou nesprávná, ať už kvůli vysvětlení některých pojmů, nebo proto, že byly získány ze zdroje, u něhož není pravdivost ani nesprávnost údajů prokázána.

Vysvětlení pojmů z genealogie

Zdroje dat

Zdroje dat, nebo také prameny, jsou pro současnou genealogii především matriky. Před jejich zřízením se dají získat genealogická data hlavně na základě různých listin, případně úřední knihy, nebo také například kroniky. Dalším, hlavně v zahraničí populárním zdrojem dat, je sčítání lidu.

Matrika

Matrika (dnes již také registr, seznam, rejstřík) je veřejný úřední seznam, který v České republice shraňuje údaje o narozeních, sňatcích a úmrtích. V dnešní době je veden matrikáři na obecních úřadech.

Sčítání lidu

Sčítání lidu (také populační census nebo soupis obyvatelstva) je proces získávání údajů o jednotlivých příslušnících obyvatelstva. Organizátorem sčítání je zpravidla stát. Ne všechny shromážděvané údaje jsou použitelné pro genealogii.

Pokrevenství

Pojem pokrevenství označuje vztah mezi osobami, které pocházejí z jediného předka, důležitý je pak stupeň pokrevenství.

Příbuzenství

Příbuzenství, které je také nazýváno zešvagřením, jsou vztahy vzniklé manželstvím, tj. vztahy k pokrevním příbuzným partnera.

Generace

Česky též pokolení, definice dle [2]: rodový soubor osob, které mají společného otce. Jeho členy jsou synové a dcery jednoho otce.

Předek

Definice dle [2]: Předek - jinak též ascendent, je takový příbuzný, který je z předcházejícího pokolení nebo z pokoleních dále směrem do minulosti (otec - děd - praděd,...)

Potomek

Definice dle [2]: Potomek, nebo též descendent v přímé posloupnosti syn - vnuk - pravnuk, obecně znamená směr do budoucnosti

Po meči

Pojem po meči značí „z otcovi strany“, většinou je tento výraz používán pro potomky, pak se tedy jedná o potomky v otcovské linii.

Po přeslici

Pojem po přeslici je používán ve stejném smyslu jako pojem „Po meči“, pro označení potomků pocházejících z ženské linie rodu

Adopce

Adopce, nebo též osvojení, je přijetí cizí osoby za vlastní. Mezi osvojitelem a osvojence vzniká podobný právní poměr jako mezi rodiči a dětmi.

Rozrod

Definice dle [1]: Rozrod podává přehled o genealogických vazbách všech členů rodu (počínaje zakladatelem rodu) a rozčleněném do jednotlivých větví rodu. Je většinou zpracován na velké ploše

(papíru), osobní údaje se omezují na stručné vyjádření rozsahu života, na uvedení manželek každého předka formou křestních jmen, rokem svatby, příp. číselným označením rodovým znakem. Rozrod zaznamenává všechny potomky pocházející z jednoho společného páru - zakladatele rodu. Jsou v něm uvedeni všichni muži a ženy, kteří mají pokrevní vztah k zakladateli rodu.

Rozrod je vždy sestavou descendentní, tj. popis se ubírá směrem k následným, budoucím událostem, tj. potomkům.

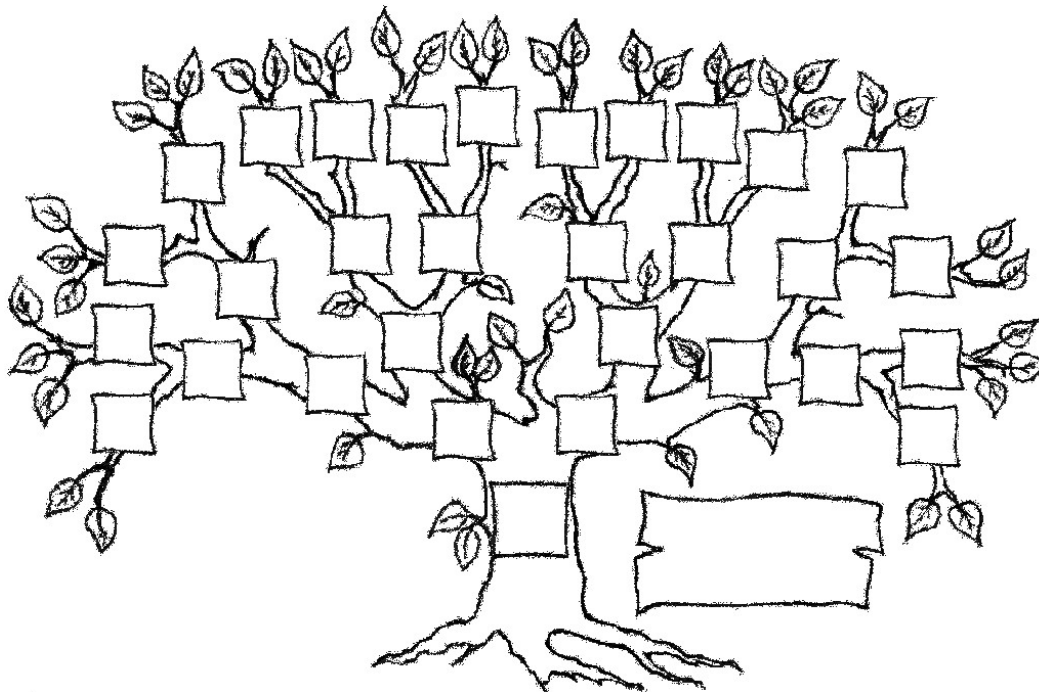
Vývod

Definice dle [1]: Vývod je druhem genealogického grafu, v němž jsou zachyceny předci po přeslici i po meči, kteří se v minulosti podíleli na zplození osoby, pro níž se tento vývod zpracovává. Je možné v něm zachytit genetické vlastnosti některých členů rodu (zdravotní, osobní apod.), sociální vývoj rodiny, vývoj profesionality v rodě (řemesla, zaměstnání), aj.

Vývod má charakter ascendentní, tj. postupuje do minulosti. Vyhledává všechny rodičovské páry. (Pozn. : manželství v tomto případě není podmínkou).

Rodokmen – strom života

Definice dle [1]: Zatímco u rozrodu a vývodu převažuje obsahová stránka a přehledné uspořádání údajů, u rodokmenu klademe důraz na výtvarné provedení při zachování historické pravdy. Bývá nazýván stromem života, znázorňující hlavní kmen a jednotlivé větve. Mívá charakter descendentní a zdůrazňuje pouze mužské členy rodu v celém jeho rozsahu. Bývá vypracován jako rozrod jedné větve rodu kmenu, ze kterého vyrůstají jednotlivé větve, tj. pokolení. Hodnota rodokmenu je především v jeho grafickém provedení a zpracování. U jednotlivých členů bývá uvedena délka života. K rodokmenu přikládáme osobní listy rodin a stručné životopisy jednotlivců.



Obrázek 1: Strom života/Rodokmen

Definice dle [2]: přehledný vývod, jak se některý rod po předcích vyvíjel tak, aby v něm byla patrná vzájemná závislost členů rodu. Rodokmen zaznamenává všechny syny a dcery pokud jsou manželskými potomky mužů pocházejících z manželství zakladatele rodu. Zakladatel rodu je nejstarší zjištěný přímý předek. Vnější znakem členů rodu je společné příjmení. Vnitřním znakem členů rodu je pokrevenství po meči. Do rodokmenu patří také: potomci neprovdaných dcer, potomci mužů, kterým byla povolena změna příjmení, potomci mužů, kteří přijali sňatkem příjmení nevěsty. Do rodokmenu nepatří: nemanželské děti synů, děti vdaných dcer - ani když rodina sňatkem rodičů přijala příjmení nevěstino, adoptivní děti, nemanželské děti do rodu vdaných žen. Rodokmen je základem historicky popisné genealogie (rodové kroniky). Osoby zařazované do rodokmenu se zapisují jménem. Příjmení se uvádí pouze u zakladatele rodu. Pokud došlo ke změně příjmení uvede se opět jen u prvního nabyvatele. Osoby se zařazují do rodokmenu podle pokolení (generací), v jednotlivých generacích podle rodin, v rodinách podle pořadí daného stáří - tj. dnem narození. Příslušníci jednotlivých rodin (rodiče a děti) se spojují svorkou.

Pro potřeby této diplomové práce budou do rodokmenu/stromu života zahrnuty vždy sourozenci a partneři dané osoby, z předků potom přímí předkové a jejich sourozenci, z potomků pak všichni přímí potomci a jejich partneři.

Rodopis

Pod pojmem rodopis se v dnešní době vnímá soukromý zájem v oblasti genealogie.

Jedinec v genealogii

Jedinec jako příslušník rodu může zastávat dvě pozice a to jako rodič (nebo ploditel) nebo zrozený (nebo zplozený). Je důležité, že každý jedinec může zaujímat obě pozice současně a to i k více dalším jedincům.

Možné vztahy v genealogii

Vztahy v genealogii jsou podstatnou součástí této práce, je jich potřeba pro navržení formátu dat. Je proto nutno uvažovat, že každý jedinec může být k dalším jedincům v následujících vztazích a to současně ve více než jednom. Uvedeny jsou pouze 2 základní vztahy, ostatní vztahy lze definovat na základě těchto.

Rodič => Zrozený

Každý jedinec může být rodič více zrozených. Jedinec ale také žádný takový vztah mít nemusí. Rodič také může zrozeného adoptovat.

Zrozený => Rodič

Zde je podstatné, že každý jedinec má 2 rodiče, otce a matku. Může se stát, že však nejsou známí. Může zde dojít k rozepřím, kdo je pokrevním otcem, neboť se může stát, že si otcovství nárokuje více jedinců. Vzhledem k tomu, že v tomto vztahu může dojít k adopci, budeme pro potřeby této práce uvažovat vztah zrozený => rodič následující možnosti: vlastní matka, nevlastní matka, vlastní otec, nevlastní otec.

Formát dat pro genealogii

Vzhledem k tomu, že základem každého sémantického webu je RDF formát, bude v rámci diplomové práce nutné takový formát dat vytvořit. O vlastnostech sémantického webu a formátu RDF budou širší informace v rámci následující kapitoly, v této kapitole si pouze uvedeme některé obecně používané formáty pro data genealogie, jejich struktur, výhody a nevýhody.

Formát GEDCOM (GE**nealogic Data **C**OMmunication)**

Asi nejčastěji se dá na webových stránkách dočíst o formátu genealogických dat GEDCOM. Tento formát je používám vícerymi již existujícími genealogickými aplikacemi, které mají usnadnit vytváření rodopisu. Někdy jsou veškerá data aplikace uložena přímo v tomto formátu, většina ostatních aplikací, které se alespoň trochu úspěšně prosadily, tento formát umí zpracovávat jako vstupní data nebo do něho umí data ze svého vlastního formátu vyexportovat.

Možnost použití formátu dle jeho obecných vlastností

Z toho hlediska, že většina aplikací tento formát podporuje, pokud s ním přímo nepracuje, bylo by vhodné, aby aplikace genealogických dat pro KiWi byla schopná data z formátu GEDCOM jak načítat, tak i zpětně do něj vlastní data vyexportovat.

Popis formátu

Tento formát je ve své podstatě definice formátu souboru, který se skládá ze tří částí a to Header, Records a Trailer. V části Header jsou uloženy převážně informace ohledně verze formátu a zdrojů informací. V části Records jsou pod sebou řazeny záznamy, tyto záznamy mohou reprezentovat jak jednu osobu, tak rodinu.

Pokud je záznamem jedna osoba, pak na jejím prvním řádku se nachází identifikátor jedince následovaný klíčovým slovem INDI, na dalších řádcích je pak nejdříve klíčové slovo, které udává jaká informace o konkrétním jedinci se za ním nachází, například jméno, pohlaví, datum narození, do které rodiny patří a podobně. V případě rodiny je uveden identifikátor, nikoliv jméno.

Pokud v záznamu nalezneme rodinu, pak na jejím prvním řádku je opět identifikátor, tentokrát následovaný klíčovým slovem FAM. Na dalších řádcích se pak nachází klíčová slova pro pozici v rámci rodiny následovaná identifikátorem člena, který tuto pozici zastává. Pozice bývají zpravidla manžel (HUSB), manželka (WIFE) a dítě (CHIL).

Sekce Trailer je ukazatelem konce datové struktury, obsahuje klíčové slovo TRLR.

Ukázka formátu (zdroj [3])

```
0 HEAD
1 SOUR Reunion
2 VERS V8.0
2 CORP Leister Productions
1 DEST Reunion
1 DATE 11 FEB 2006
```

```

1 FILE test
1 GEDC
2 VERS 5.5
1 CHAR MACINTOSH
0 @I1@ INDI
1 NAME Bob /Cox/
1 SEX M
1 FAMS @F1@
1 CHAN
2 DATE 11 FEB 2006
0 @I2@ INDI
1 NAME Joann /Para/
1 SEX F
1 FAMS @F1@
1 CHAN
2 DATE 11 FEB 2006
0 @I3@ INDI
1 NAME Bobby Jo /Cox/
1 SEX M
1 FAMC @F1@
1 CHAN
2 DATE 11 FEB 2006
0 @F1@ FAM
1 HUSB @I1@
1 WIFE @I2@
1 MARR
1 CHIL @I3@
0 TRLR

```

Možnost použití formátu dle jeho technických vlastností

Možnosti využití tohoto formátu jsou nízké, je zde pozitivní to, že pro záznam rodiny už se například odkazuje na členy pomocí identifikátorů. Bohužel například pro města (místa narození) se používá řetězec se jménem města, takže v tomto bodě pokud dojde k překlepu, dochází k problémům se spojením více lidí například narozených ve stejném městě. Tento formát je jen těžce použitelný, protože neposkytuje možnosti pro přidávání nových atributů. Pokud má být databáze široká a vyhledávání možné například i podle místa studia nebo bydliště, tak se stává formát GEDCOM nepoužitelný.

OPC formát aplikace Family.Show

OPC (Open Package Convention) formát souboru je používán v aplikaci Family.Show, která se v nedávné době objevila jako novinka v oblasti genealogických aplikací. Na této aplikaci spolupracuje skupina vývojářů známá pod jménem Vertigo společně s vývojáři společností Microsoft. Aplikace je v současné době ve verzi 3.0. Aplikace Family.Show je zaměřená jak na ukládání dat, tak i na jejich prezentaci, proto je pro tuto práci přínosem jak v oblasti formátu dat, tak i v oblasti jejich

reprezentace, kterou se práce bude zabývat v dalších kapitolách.

Možnost použití formátu dle jeho obecných vlastností

Pro tento formát je celkem výhodnou vlastností to, že je založen na struktuře souborů XML. To z něj činí, jak bude později vysvětleno, perspektivní podklad pro formát RDF souboru. Vzhledem k tomu, že aplikace Family.Show je Open Source, je k dispozici kód pro ukládání dat do tohoto formátu. Aplikace Family.Show také obsahuje konvertor z/do dříve popsaného formátu GEDCOM.

Popis formátu

Jeden soubor zde reprezentuje jednu rodinu, zatím má koncovku *.family. Jak již bylo zmíněno, je založený na formátu XML a je poněkud více popisný než formát GEDCOM. Jak již XML napovídá, používají se zde značky pro ohraničení údajů. Na začátku souboru je uvedena verze XML, ve které je soubor uložen, hned za verzí XML následuje značka pro rodinu, která v sobě jako atributy obsahuje některé informace, které jsou později využívány pro inicializaci zobrazení při načtení rodiny.

Značka <PeopleCollection> uvozuje a zakončuje seznam osob, které jsou do rodiny zahrnuty. Jednotlivé osoby jsou pak definovány značkou <Person>, která je vnořena do značky <PeopleCollection>, a mají svůj unikátní identifikátor jako atribut. Pro popis vlastností osoby jako je jméno, příjmení, pohlaví, datum a místo narození, datum a místo úmrtí pak slouží další zanořené značky. Mezi značky zanořené do značky pro osobu u staršího formátu patří také fotografie, která pokud existuje, tak uvádí relativní a absolutní cestu k souboru s obrázkem.

Na rozdíl od formátu GEDCOM nejsou vztahy mezi osobami vyjádřeny pomocí rodin, ale každá osoba má přímo u sebe uložen seznam vztahů. Slouží k tomu značka <Relationships>, která je také vnořena do značky <Person>, a do které jsou vnořeny další značky <Relationship>, které ale už v atributu uvádějí vztah, jaký je mezi danými osobami. O osobě, se kterou má popisovaný jedinec vztah, je pak vnořeno ještě několik málo informací jako je role ve vztahu, identifikátor osoby, celé jméno a další informace, uvádějící, zda je vztah přirozený, zda došlo k adopci a nebo například při partnerství/manželství může být vztah ukončený.

V příkladu uvedeném níže je vidět velká popisnost, je zde uvedena pouze jedna jediná osoba. Soubory s rodinami mohou tedy nabývat celkem mimořádných velikostí, na jednu osobu je průměrně přibližně 2000 znaků (ověřeno z vlastních zkušeností s touto aplikací). Záleží hlavně na množství vztahů, které k jsou k osobě uvedeny, množství osobních údajů je malé v porovnání s daty o vztazích.

Uvedena je starší verze formátu OPC, novější verze již předpokládá zahrnutí fotografií a textových dokumentů přímo ve formátu tak, aby mohla být všechna data pro celou rodinu uložena v jednom jediném souboru. Ve starší verzi jsou samostatně uloženy jak fotografie, tak dokumenty formátu RTF s životopisem osoby. Pro tento životopis nejsou definovány žádné formáty nebo šablony.

Ukázka formátu (vlastní, vytvořeno pomocí aplikace Family.Show)

```
<?xml version="1.0"?>
<Family xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" Current="2f0db722-bd95-4c7c-
```

```

b9ff-c9f09d1723d2" CurrentName="John Doe" FileVersion="1.0">
<PeopleCollection>
  <Person Id="2f0db722-bd95-4c7c-b9ff-c9f09d1723d2">
    <FirstName>John</FirstName>
    <LastName>Doe</LastName>
    <Gender>Male</Gender>
    <BirthDate>1954-07-23T00:00:00</BirthDate>
    <BirthPlace>London</BirthPlace>
    <DeathDate xsi:nil="true" />
    <IsLiving>true</IsLiving>
    <Photos />
    <Relationships>
      <Relationship xsi:type="ChildRelationship">
        <RelationshipType>Child</RelationshipType>
        <PersonId>1628b775-99c8-4203-b94a-120cd83bc9e8</PersonId>
        <PersonFullname>Kate Doe</PersonFullname>
        <ParentChildModifier>Natural</ParentChildModifier>
      </Relationship>
      <Relationship xsi:type="SpouseRelationship">
        <RelationshipType>Spouse</RelationshipType>
        <PersonId>6d3062b2-1a25-4abf-90d9-d396ffa3ad8c</PersonId>
        <PersonFullname>Jane Doe</PersonFullname>
        <SpouseModifier>Former</SpouseModifier>
        <MarriageDate>1979-01-01T00:00:00</MarriageDate>
        <DivorceDate>1992-01-01T00:00:00</DivorceDate>
      </Relationship>
      <Relationship xsi:type="ParentRelationship">
        <RelationshipType>Parent</RelationshipType>
        <PersonId>76b08f58-768b-4e96-bf92-a781746674d2</PersonId>
        <PersonFullname>Bill Doe</PersonFullname>
        <ParentChildModifier>Natural</ParentChildModifier>
      </Relationship>
      <Relationship xsi:type="ParentRelationship">
        <RelationshipType>Parent</RelationshipType>
        <PersonId>556dc5c9-b29f-480a-b9fe-e6debd34cb8</PersonId>
        <PersonFullname>Suzan Doe</PersonFullname>
        <ParentChildModifier>Natural</ParentChildModifier>
      </Relationship>
      <Relationship xsi:type="SiblingRelationship">
        <RelationshipType>Sibling</RelationshipType>
        <PersonId>468d7d3e-606c-4961-ac56-47a7328b2fab</PersonId>
        <PersonFullname>Jill Carter</PersonFullname>
      </Relationship>
    </Relationships>
    <HasParents>true</HasParents>
    <HasSpouse>true</HasSpouse>
  </Person>

```

```
</PeopleCollection>
</Family>
```

Možnost použití formátu dle jeho technických vlastností

Výhodou formátu je jistě struktura XML, která nejen že umožňuje rozšiřování o další atributy případně elementy, ale také umožňuje větší pochopitelnost dat, než tomu bylo u formátu GEDCOM. Možnou výhodou i nevýhodou je uložení obrazových a textových dat přímo do jednoho souboru. Seznam vztahů uvedený u každé osoby může být výhodou při částečné ztrátě dat, ale rozhodně je výhodná při vizualizaci, kdy není třeba zobrazovat všechna data o dalších osobách a tím urychlí zpracování a samotné zobrazení dat.

Formát GEDCOM 6.0

Formát GEDCOM 6.0, jinak též GEDCOM XML, vznikl na základě formátu GEDCOM, ale jak už název napovídá, využívá stejně jako OPC strukturu XML. Stejným rysem spolu s původním GEDCOM formátem je oddělení záznamů o rodině od záznamů o jednotlivci. Jako odlišnost od OPC formátu lze uvést, že kromě záznamů o rodině a jejích členech, se v souboru nachází i záznamy o událostech, kterými jsou například svatba, narození nebo křest dítěte.

V některých směrech, jako je například větší používání odkazů, je tento formát výhodnější než OPC, na druhou stranu jsou zde často ukládána data, která se mohou zdát redundantní. Například uvedení věku nevěsty v době svatby lze spočítat za předpokladu, že známe datum svatby a datum narození nevěsty. V příkladu níže je uvedena pouze část pro popis událostí, popis rodiny a jednotlivých členů je stejným stylem.

Příklad formátu pro událost (zdroj [6])

```
<!-- Event Records -->
<EventRec Id="EV001" Type="marriage" VitalType="marriage">
  <Participant>
    <Link Target="IndividualRec" Ref="IN001"/>
    <Role>husband</Role>
    <Age>26</Age>
  </Participant>
  <Participant>
    <Link Target="IndividualRec" Ref="IN002"/>
    <Role>wife</Role>
    <Age>21</Age>
  </Participant>
  <Date Calendar="Julian">ABT 7 NOV 1834</Date>
  <Place>
    <PlaceName>
      <PlacePart Type="town" Level="4">Cove</PlacePart>,
      <PlacePart Type="county" Level="3">Cache</PlacePart>,
      <PlacePart Type="state" Level="2">Utah</PlacePart>,
      <PlacePart Type="country" Level="1">USA</PlacePart>
```



```

        <!--
            This would print as Cove, Cache, Utah, USA.
            In the data stream, each comma is followed by a blank,
which can't be seen here.
            The line breaks that we have used for clarity are not in
the actual data stream.
        -->
        </PlaceName>
        <Coordinates>18.153N 178.150E</Coordinates>
        <PlaceNameVar Method="kana">. . .</PlaceNameVar>
        <PlaceNameVar Method="romanji">. . .</PlaceNameVar>
    </Place>
    <Religion>Reformed Christian</Religion>
    <ExternalID Type=". . ." Id=". . ."/>
    <Submitter>. . .</Submitter>
    <Note>. . .</Note>
    <Evidence>
        <Citation>
            <Link Target="SourceRec" Ref="SR001"/>
            <WhereInSource>File No. 7895-09, p. 23</WhereInSource>
            <WhenRecorded>10 June 1903</WhenRecorded>
            <Extract>Text extracted from the source.</Extract>
            <Note>Certified copy in possession of Larry T. Smith, Sandy,
Utah.</Note>
        </Citation>
    </Evidence>
    <Enrichment>. . .</Enrichment>
    <Changed>. . .</Changed>
</EventRec>

```

Možnost využití formátu GEDCOM 6.0

Využití formátu GEDCOM 6.0 jsou zhruba stejné jako využití formátu OPC díky struktuře XML. Rozdílem je zde hlavně to, že pro události, rodiny a členy existují 3 odlišné typy záznamů. Výhodou může být již zmiňovaná provázanost, která je ale nejvíce patrná pouze pro události, na které je odkazováno z dat pro jedince nebo pro rodinu. Další provázanost je viditelná mezi rodinou a jejími členy, na které se odkazuje. Toto ale na druhou stranu není nevýhodou formátu OPC, protože ten již na základě své definice odkaz mezi rodinou a členem nepotřebuje vyjádřit.

Sémantický web

Definice Internetu

Definice dle [3]: Internet je celosvětový systém navzájem propojených počítačových sítí („sítí sítí“), ve kterých mezi sebou počítače komunikují pomocí rodiny protokolů TCP/IP. Společným cílem všech lidí využívajících Internet je bezproblémová komunikace (výměna dat).

Nejnámější službou poskytovanou v rámci Internetu je WWW (kombinace textu, grafiky a multimédií propojených hypertextovými odkazy) a e-mail (elektronická pošta), avšak nalezneme v něm i desítky dalších. Laici někdy spojují pojmy WWW a Internet, i když WWW je jen jednou z mnoha služeb, které na Internetu nalezneme.

Definice Webu

Definice dle [3]: World Wide Web (WWW nebo také pouze zkráceně web), ve volném překladu „celosvětová pavučina“, je označení pro aplikace internetového protokolu HTTP. Je tím myšlena soustava propojených hypertextových dokumentů.

V češtině se slovo web často používá nejen pro označení celosvětové sítě dokumentů, ale také pro označení jednotlivé soustavy dokumentů dostupných na tomtéž webovém serveru nebo na téže internetové doméně nejnižšího stupně (internetové stránky).

Definice URI

Definice dle [3]: URI, celým názvem Uniform Resource Identifier – („jednotný identifikátor zdroje“) je řetězec znaků s definovanou strukturou, který slouží k přesné specifikaci zdroje informací (ve smyslu dokument nebo služba), hlavně za účelem jejich použití pomocí počítačové sítě, zejména Internetu.

URI je nejobecnější z několika příbuzných typů identifikátorů. URI může popisovat zdroj jak čistě z hlediska jeho identity (a neurčovat, kde je možno zdroj získat), tak čistě z hlediska toho, jak je možno zdroj nalézt (a nepopisovat jeho identitu), tak i obojí současně – přesnou identitu zdroje i jak je možno ho dosáhnout.

Oproti URI popisuje URL primárně způsob, jakým se lze ke zdroji dostat, naopak URN specifikuje zdroj jako takový a nesnaží se o návod k jeho dosažení. Hranice mezi těmito typy je však mírně mlhavá a zejména místo URL se často uvádí obecnější termín URI.

Sémantický web

Sémantický web je označován jako nástupce současného webu, někdy jako web 2.0. Za autora myšlenky sémantického webu je označován tentýž muž, který byl u zrodu původního webu. Tímto mužem je Tim Berners-Lee.

Historie sémantického webu

Sémantický web je myšlenka na poměry výpočetní technologie a informatiky ne zrovna nová. Je to již 8 let, kdy Tim Berners-Lee poprvé představil veřejnosti jeho myšlenku. Bylo to v květnu roku

2001 a to článkem [7] uvedeným v Scientific American. V tomto článku Tim Berners-Lee spolu se svými kolegy z W3C poukazuje na fakt, že internet a hlavně tedy jeho služba web je rostoucí směs dokumentů, které nejsou nijak organizovány, vzájemně jsou propojeny pouze v minimální míře a tvoří tak spíše chaos.

Sémantický web podle Berners-Leeho

Co si Tim Berners-Lee představuje pod pojmem sémantický web, přeloženo z článku [7]: Nová forma obsahu webu, která je srozumitelná/pochopitelná pro počítače, uvolní revoluci nových možností.

Dále je v článku uvedeno, že sémantický web by neměl být samostatnou součástí, ale měl by rozšířit současný web. Informace v něm by měly být lépe definované a to tak, aby usnadnily komunikaci mezi člověkem a počítačem. Prvním krokem vývoje sémantického webu by mělo být propojení se současným webem, což již v současné době probíhá. Do budoucna by pak toto propojení mělo posloužit jako významná nová funkcionalita, kdy stroje budou schopny lépe zpracovat informace, které nyní ne zcela úplně chápají a následně zobrazují.

Web zůstane stále decentralizovanou službou a i když nikdy neodstraní nekonzistence, kdy se stane, že hyperlink ukazuje na neexistující objekt, bude i nadále tato decentralizace umožňovat jeho exponenciální růst, což je důležitým rysem pro nástroj, který má být univerzálním a který má poskytovat nejširší spektrum informací.

Reprezentace znalostí

V sémantickém webu, pokud má správně fungovat, musí mít počítače přístup ke strukturovaným kolekcím informací a k množinám implikačních pravidel, která mohou používat pro vedení automatického zpracování.

Tradiční systémy pro reprezentaci znalostí byly doposud centralizované, což ale vyžadovalo přesně shodnou definici společných pojmů a centralizované systémy byly nuceny rychle růst jak do šířky, tak do hloubky, což způsobilo zpravidla jeho neříditelnost. Tyto systémy také často omezují otázky tak, aby systém zvládl odpovědět spolehlivě. Tim Berners-Lee v článku také poukazuje na Göedlovy věty o neúplnosti, kdy se tradiční systémy reprezentující znalosti snaží problémy plynoucí z těchto vět odstranit tím, že každý systém má množinu svých pravidel pro vyvozování závěrů o datech. Jako příklad uvádí genealogický systém operující nad databází rodiných stromů: v genealogii vždy platí pravidlo, že manželka strýce je teta. A i když jde vždy přenést data mezi systémy, pravidlo, které existuje v úplně jiné formě, většinou převedeno být nemůže.

Oproti tradičním systémům pro reprezentaci znalostí, vývojáři sémantického webu akceptují paradoxy a otázky, které nelze zodpovědět. Tato akceptace je cenou za dosažení univerzality. Proto je vytvořen jazyk pro pravidla tak výrazový, aby splnil potřeby a požadavky webu na široké uvažování.

Výzvou pro sémantický web je poskytnout takový jazyk, který vyjadřuje jak data, tak pravidla pro usuzování o datech a který umožňuje export pravidel z jakéhokoliv existujícího systému pro reprezentaci dat na web.

Ontologie

Prvním krokem při vytváření každého sémantického webu je konceptualizace dat. Ontologie je zpravidla nejčastějším nástrojem právě pro konceptualizaci dat. Bývá specifická pro určitý obor, pro určitou doménu. Ontologie jsou pak formalizované reprezentace určitých znalostí, dat, pojmů. Návrh reprezentace je postaven tak, aby data mohla být znovupoužita a dále spracovávána. Ontologie bývají často navrženy jako sítě nebo hierarchie.

Agenti sémantického webu

Jak již bylo dříve zmíněno, hlavní podstatou sémantického webu je uspořádání dat tak, aby mohla být dále zpracovávána a to co nejvíce automaticky, čili počítačem. Právě pro zpracování sémantických dat automatickým způsobem jsou navrženi agenti.

Tito softwaroví agenti jsou vlastně krátké programy nebo programové komponenty, které se starají o vyhledávání dat, případně jejich zpracování. Většinou by měli být k dispozici distribuovány stejně jako data tradičního webu. Agenti nepracují a nevyhledávají data sami od sebe, je nutné, aby byli pro nějaký vyhledávací účel pověřeni uživatelem.

Metadata

Definice dle [8]: Metadata mohou být stručně definována jako (strukturovaná) *data o datech*. Zachycují obsah, kontext a strukturu dat, která popisují. Síťová metadata, na která se zaměříme, jsou nejčastěji zapisována prostřednictvím XML, které svými vlastnostmi nejvíce odpovídá požadavkům na otevřenost, přenositelnost a interoperabilitu formátu pro výměnu a ukládání dat.

RDF a XML

Dvě velice důležité technologie, XML (eXtensible Markup Language) a RDF (Resource Description Framework) byly již při uvedení sémantického webu v roce 2001 uváděny jako dvě hlavní technologie, které mají pomoci sémantickému webu, aby dosáhl svého cíle, tedy aby počítač byl schopen chápat data i ze sémantického hlediska.

XML je technologie vhodná pro dosažení univerzálnosti a to především díky tomu, že lze vytvářet své vlastní značky, které označují a uvozují strany nebo sekce textu. Tyto značky jsou při reprezentaci dokumentu skryté, ale právě pro nástroj, který tyto data prezentuje, mají velký význam. Jsou proto zpracovávány skripty, při jejichž psaní ale musí být autorovi znám význam jednotlivých značek. XML soubor vlastně autorovi pomáhá vytvářet strukturovaný text, ale dalším osobám už neříká, co daná struktura znamená a co vyjadřuje.

Význam a sémantiku takovým značkám pak dodává právě RDF, které kóduje význam do množiny trojic, kde každá trojice je spíše jako podmět, přísudek a předmět v základní větě. Tyto trojice mohou být zapsány jako XML tagy (značky). V RDF, dokument tvrdí, že určitá věc (člověk, webová stránka nebo cokoliv jiného) má vlastnost (například „je sestrou“, „je autorem“, ...) s nějakou hodnotou (jiná osoba, jiná Webová stránka, kniha, ...). Tato struktura se jeví jako přirozený způsob jak popsat většinu dat, která jsou zpracovávána strojově. Podmět i předmět jsou každý jedinečně identifikován pomocí URI identifikátoru. Také slovesa mohou mít své URI, které umožní komukoliv definovat nový koncept, nové sloveso pouze tím, že definují URI pro toto sloveso někde na webu.

Lidský jazyk občas používá stejné výrazy pro různé věci, lidé toto rozpoznají, automatické zpracování ovšem ne. Pro přiblížení ve svém článku Tim Berners-Lee používá příklad, kdy najme klaunskou doručovací službu, aby doručila zákazníkům nafukovací balónek v den jejich narozenin. Bohužel pokud počítač vytáhne z databáze zákazníků, kam má být doručen účet, může se stát, že klaun místo klienta navštíví pouze poštovní schránku na poštovním úřadě. Tento problém může být vyřešen jednoduše tím, že budou použity různé URI pro každý specifický pojem, tedy v našem případě pro každý typ adresy. Potom lze jednoduše rozlišit mezi adresou, která je pro doručování, a mezi adresou ulice, kde zákazník skutečně bydlí.

Trojice RDF formují informace z webu do souvisejících věcí. Protože RDF používá URI pro zakódování této informace do dokumentu, URI zajišťuje, že pojmy nejsou pouze slova v dokumentu.

Příklad RDF

Převzato z [8]

```
<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns\#
xmlns:vcard=http://www.imc.org/vcard/3.0/
xmlns:dc=http://purl.org/dc/elements/1.1/>
  <rdf:Description
    about=http://www.nejakaadresaveformatu.html>
    <dc:Title>Titulek textu</dc:Title>
    <dc:creator rdf:resource=http://www.adresa/neco>
    <dc:date>2007-15-09</dc:date>
  </rdf:Description>
  <rdf:Description about= http://www.adresa/neco >
    <vcard:fn>Nejaka Karta</vcard:fn>
    <vcard:email>vas@email</vcard:email>
  </rdf:Description>
</rdf:RDF>
```

Jazyk OWL

Jazyk OWL (Web Ontology Language) byl navrhnout speciálně pro ontologie sémantického webu. Je tedy navržen pro použití aplikacemi, které potřebují zpracovávat obsah informací namísto pouhé prezentace informací lidem. OWL umožňuje větší strojovou interpretabilitu webového obsahu, než jak ji podporují XML nebo RDF tím, že nabízí přídatný slovník s formální sémantikou. Jazyk OWL má tři podmnožiny, které jsou sami o sobě jazyky, jsou jimi OWL Lite, OWL DL a OWL Full.

Návrh RDF pro genealogii

V předchozích kapitolách jsme si představili několik různých formátů dat pro genealogii a sémantický web. Pokud mají být genealogická data prána jako podklad pro zpracování sémantickým webem, je třeba navrhnout RDF formát pro genealogická data.

Úplně nový formát versus formát navazující na stávající

Jedním z prvních důležitých rozhodnutí je zda se bude navazovat na již existující genealogický formát, nebo zda bude lepší vytvořit si kompletně vlastní od začátku. Vzhledem k možnosti případné snazší konverze mezi již existujícími formáty, bude vhodnějším navázat na některý existující a pouze ho upravit pro potřeby sémantického webu. Počítá se také s tím, že navrhovaný formát se bude v průběhu implementace v menší míře přizpůsobovat a upravovat na základě nových poznatků z praxe při implementaci sémantického webu.

GEDCOM versus OPC

Pro prvotní návrh je nutné si vybrat hlavně mezi formáty GEDCOM, který je již delší dobu populární v různých aplikacích a určitě jeho XML verze může být základem pro RDF, a jeho druhým hlavním oponentem - formátem OPC aplikace Family.Show, který je také založen na XML.

Hlavním rozdílem mezi těmito dvěma formáty je, že GEDCOM rozlišuje data pro rodinu a data pro jednotlivce (potom vztahy mezi jedinci se musí odvozovat z rodiny), kdežto OPC má pouze data o jednotlivcích, kdy vztahy jedince k ostatním jsou uloženy přímo u jedince.

Snazší implementace se nabízí, pokud bude pouze jeden typ dat jako je tomu u OPC. Co se možnosti redundance dat týče, tak v OPC se můžou zdát data redundantní například v případě, že je v rodině více sourozenců. Pak pokud jich je například 5, tak u každého z nich jsou 4 záznamy vedoucí k ostatním sourozencům, tedy 20 záznamů. Na druhou stranu ve formátu GEDCOM je jedna osoba obsazena ve více rodinách podle role. Jedince můžeme najít nejprve jako dítě v rodině, později pak v partnerských vztazích, ať se již jednalo o manželství nebo pouhá partnerství, kde je v roli rodiče, pokud z partnerství vzešly děti, a zároveň jednoho z partnerů. V dnešní době není střídání partnerů až tak neobvyklou situací, a proto je opět možné, že jedna osoba se vyskytne i ve více jak dvou rodinách.

Vzhledem k tomu, že není žádný z výše uvedených formátů ani extrémně vhodný, ani extrémně nevhodný a vzhledem k tomu, že na základě OPC formátu již existuje moderní vizualizace pomocí klasické windows aplikace, přikláním se k formátu OPC, který se jistě neukázal jako úplně nevhodný pro vizualizaci vzhledem k uvedeným faktům.

Genealogická data v RDF

Pokud má RDF formát sloužit hlavně k ukládání genealogických dat, je třeba si stanovit, jaká hlavní data jsou pro genealogii nezbytná, jaká je vhodné ukládat a jakými se pravděpodobně není třeba vůbec zabývat. Na základě implementace pak je opět možné, že toto bude třeba přehodnotit.

Navrhovaný formát RDF je rozdělen do částí vždy tak, aby korespondující část kódu byla přiřazena u vlastnosti, kterou popisuje.

Definice jmenných prostorů pro vlastnosti

Pro formát uložených dat jsou použity některé vlastnosti definované v již existujících RDF, pro nově definovaná data je pak vytvořen nový genealogický jmenný prostor.

```
<rdf:RDF
```

```
xmlns:foaf="http://xmlns.com/foaf/0.1/"
```

```

xmlns:kiwi="http://www.kiwi-project.eu/kiwi/core/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
xmlns:gene="http://www.kiwi-project.eu/kiwi/genealogy/" >

```

Základní genealogická data

Uvažuje se, že pokud všichni lidé budou mít tato data vyplněna, pak je možné sestavit celý rodinný strom pouze na jejich základě. Dále jsou tu základní genealogická data o narození, úmrtí a pohlaví. V podstatě se tedy jedná o nejnütnější a nejdůležitější data:

- ID osoby

```
<rdf:Description rdf:about="URI_osoby">
```

- jméno – nutno užívat více křestních jmen a příjmení, dále name - jméno pro zobrazení aplikací (použito například v případě odkazu na danou osobu)

```
<foaf:firstName>Karolina</foaf:firstName>
```

```
<foaf:surname>Novakova</foaf:surname>
```

```
<foaf:name>Karolina Novakova</foaf:name>
```

- ID matky

```
<gene:mother rdf:resource="URI_matky"/>
```

- ID otce

```
<gene:father rdf:resource="URI_otce"/>
```

- Pohlaví – uvažuje se „female“ – ženské a „male“ - mužské

```
<foaf:gender>male</foaf:gender>
```

- Identifikátor, zda je osoba naživu či po smrti, tento identifikátor se může zdát irelevantní, pokud je vyplněno datum a místo úmrtí, ale pokud toto není známo, pak je vhodné tuto informaci uchovat jinak. Pro tuto vlastnost existují 3 možnosti: „alive“ – žijící, „dead“ – po smrti, „not-known“ – není známo (například u pohřešovaných osob)

```
<gene:living>dead</living>
```

- narození – datum + místo; v budoucnu možné rozšíření pro místo, lze uvažovat jako URI na místo, například jako vlastnost geo:populated_place

```
<gene:birthdate>24-03-1932</birthdate>
```

```
<gene:birthplace>Praha</birthplace>
```

- úmrtí – datum + místo; v budoucnu možné rozšíření pro místo, lze uvažovat jako URI na místo, například jako vlastnost <geo:populated_place>

```
<gene:deathdate>20-10-2005</deathdate>
```

```
<gene:deathplace>Praha</deathplace>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```

Data standardní údaje pro každé RDF v KiWi

Tato část obsahuje údaje jako ID, datum vytvoření, datum modifikace. Dále sem byla zahrnuta data o typu objektu, který je v RDF uchováván.

```
<kiwi:id>1441794</kiwi:id>
<kiwi:modifiedOn>2009-05-22T16:56:03+0200</kiwi:modifiedOn>
<kiwi:createdOn>2009-05-22T16:56:03+0200</kiwi:createdOn>
<rdf:type rdf:resource="http://www.kiwi-
project.eu/kiwi/core/ContentItem"/>
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
```

Genealogická data vhodná k uchování

Tyto data není nutné uchovávat pro sestavení rodinného stromu, ale slouží spíše jako křížové reference a doplňující informace k základním datům, která jsou potřeba pro provázanost stromu, v případě zařazení do RDF výše by byly vnořeny na stejné úrovni jako například rodiče:

- u otce a matky je vhodné označit případnou adopci včetně data adopce

```
<gene:adoptiveMother>"URI_matky"</adoptiveMother>
<gene:adoptiveFather>"URI_otce"</adoptivefather>
<gene:date>1944-06-14</date>
</parents_adoption>
```

- ID partnerů + k nim datum začátku, datum konce a informace, zda se jedná o manželství/partnerství

```
<gene:partner>"URI_partnera1"</partner>
<gene:marriageDate>1952-11-19</marriageDate>
<gene:divorceDate>1966-04-08</divorceDate>
<gene:partner>"URI_partnera2"</partner>
<gene:marriageDate>1970-06-05</marriageDate>
```

- ID sourozenců

```
<gene:siblink>"URI_sourozence1"</siblink>
<gene:siblink>"URI_sourozence2"</siblink>
```

- ID dětí + informace o případné adopci

```
<gene:child>"URI_ditete21"</child>
<gene:child>"URI_ditete21"</child>
<gene:adoptiveChild>"URI_ditete"</adoptiveChild>
<gene:adoptionDate>1955-07-13</adoptionDate>
```


Další data, která je možné uchovávat

Tyto data se neuvažují pro implementaci do RDF v rámci rozsahu diplomové práce, proto jsou řádky níže pouze možný návrh. Vzhledem k tomu, že se většinou jedná o životopisná data, je otázkou, zda tato data patří do genealogie, nebo zda by spíše bylo vhodné vytvořit nějaký jiný samostatný RDF formát například pro životopis a do něho tato data uložit. Pak by například stačilo ke každé osobě přidat do RDF jednu trojici odkazující na RDF s životopisnými daty této osoby. U zaměstnání případně u škol lze uvažovat databázi s právními institucemi, s jejímiž daty by mohla genealogie spolupracovat.

Uvažujeme-li, že každý člověk má otce a matku zaznamenány, pak je jednoduché vytvořit celý rodinný strom. Vzhledem k tomu, že taková data ale neexistují a dnešní genealogové, ať už profesionální nebo amatérští, vyhledávají možné souvislosti na základě dalších životopisných údajů, je vhodné uvažovat i následující:

- adresa bydliště – kompletní adresa bydliště jako URI, spolu s datem pobytu (od-do)

```
<gene:residence>" house>URI"</residence>
<gene:residenceDateFrom>1954-07-28</residenceDateFrom>
<gene:residenceDateTo>1966-08-18</residenceDateTo>
```
- škola – studovaná škola s URI školy, léty studia a navštěvovanými třídami/obory v jednotlivých letech

```
<gene:school>"URI_skoly"</school>
<gene:schoolLocation>"URI_mapa"</schoolLocation>
<gene:schoolYearFrom>1951-09-01</schoolYearFrom>
<gene:schoolYearTo>1952-08-31</schoolYearTo>
<gene:domain>"URI_oboru"</domain>
<gene:class>3.B</class>
```
- zaměstnání – zaměstnavatel (URI), adresa výkonu povolání (URI) a nejlépe i pozice

```
<gene:employments>
<gene:employer>
  <gene:institution>URI_zamestnavatele </institution>
  <geo:house>"URI_mapa"</house>
  <gene:dateFrom>01-09-1951</dateFrom>
  <gene:dateTo>31-08-1951</dateTo>
  <gene:position>Ucitelka</position>
</employer>
</employments>
```
- fotky – URI, případně s URI lokace pořízení snímku, datem pořízení snímku. Pro URI lokaci snímku není možné jasně určit, která konkrétní vlastnost by měla být použita. Celkem rozumným řešením se zdá být možnost uložení jakékoliv vlastnosti, ať už se jedná o obydlené místo (viz příklad), dům, ulici, park nebo třeba místo pod hladinou oceánu.

```
<gene:photo>"URI_fotografie"</photo>
```

```
<gene:populated_place>"URI_mapa"</populated_place>  
<gene:photo>"URI_fotografie2"</photo>  
<gene:populated_place>"URI_mapa2"</populated_place>
```

- životopis – URI dokumentu, kde se nachází textová podoba o průběhu života osoby, všechny další

```
<gene:curriculum>"URI_dokumentu"</curriculum>
```

KiWi – Knowledge in a Wiki

Podle webových stránek projektu KiWi [10] je tento projekt označen jako Kolaborativní správa znalostí, poháněná sémantickým webem. KiWi je projekt, který nabízí nový přístup ke správě znalostí a současně kombinuje filozofii wiki s inteligencí a metodami sémantického webu.



Tento projekt je realcí na současný trend, kdy různé wiki a sociální softwary revolučně změnily způsob, kterým vytváříme a distribuujeme znalosti. Sémantický web také začal přetvářet způsob, kterým udržujeme, objevujeme a sdílíme znalosti nad různými platformami.

Základní princip aplikace KiWi

Volně přeloženo z [11]: Technologiemi oplývající KiWi systém je na komponentách založené architektuře vybudován na JBoss Seam a Java EE technologiích. Hlavním principem je, že všechno je reprezentováno jako content item (položka obsahu, pro potřeby této práce bude i nadále používán anglický termín). Content item se vždy skládá z pro lidi čitelného obsahu, který je propojen s RDF vztahy a uživatelé jej mohou anotovat pomocí sémantických značek. Toto spojení umožňuje, aby byl content item sdílen mezi aplikacemi, například content item, který je vytvořen ve Wiki aplikaci může být v tu samou dobu zobrazován aplikací TagIT jako lokace na mapě a nebo jako blogový příspěvek v Blog aplikaci. Toto je umožněno díky flexibilnímu uspořádání poskytovaném technologií RDF. Částečně vyčnívajícím aspektem jádra KiWi systému je, že umožňuje použití RDF pro Java developery v mnohem příjemnější formě (již nikdy více není potřeba manuálně vytvářet a ukládat triple store). Tato příležitost se projevuje například poskytnutím flexibilních fasád (facade) pomocí rozhraní se speciální Java anotací, která mapuje Java metody na vlastnosti RDF.

První release aplikace Kiwi

Úplně první a zatím poslední oficiální Open Source release aplikace KiWi se uskutečnil v letošním roce, a to 27. února. Stalo se tak rok po začátku tohoto projektu. Podle článku [11] je jádro KiWi systému flexibilní platformou pro budování různých druhů semanticko-sociálních softwarových aplikací. V době release takovýmito aplikacemi byla Semantic Wiki a TagIT aplikace. KiWi po releasu poskytuje všechny základní služby potřebné pro takovéto aplikace, jako je editace, značkování, ukládání obsahu a asociovaná meta-data, má vlastní trojici (triple store), poskytuje transakce a verzování nad obsahem a meta-daty, spojuje otevřený datový server, dále je zde k dispozici mnoho dalších drobných možností, které mají v oblibě vývojáři semanticko-sociálních aplikací (jako služby usnadňující práci s ontologiemi nebo SKOS, atd.) Nakonec KiWi systém také má nyní celkem přehledné rozhraní, které se jen těžko hledá v aplikacích založených nad sémantickým webem.

KiWi-Vision

Příslib efektivního managementu znalostí měl podněty od průmyslu již více jak deset let ale v reálu dostupné systémy často nesplňují očekávání. KiWi projekt financovaný z EU je projektem, který kombinuje wiki metody pro kolaborativní vytváření obsahu s technologiemi sémantického webu, aby pozvedl management znalostí na vyšší úroveň.

Pro KiWi v současnosti existuje kupa článků ohledně vizí, které je možné díky této technologii dosáhnout. Jedná se o vize managementu znalostí pro Web 3.0. Vzhledem k tomu, že se zatím jedná pouze o vize, některé z nich ještě nejsou v současnosti v projektu obsaženy. Vzhledem k cílové skupině čtenářů, kteří mají být i netechnicky zaměřeni, je většina článků oproštěna o technické detaily.

Pro příklad těchto vizí je možné uvést například Wiki způsob: nikoli technologie ale filozofie; Management znalostí + Wiki filozofie + sémantický web = KiWi nebo zaměření na uživatele v článku Ultimátní podpora: Pět rozměrů „uživatelskosti“. Kompletní články je možno přečíst na stránkách projektu [10].

KiWi-System

Převzato z [10], volně přeloženo: Filozofie Wiki se neaplikuje pouze na aplikace typu wiki. Většina sociálních softwarových systémů jako blogy, stránky pro sdílení fotografií nebo platformy sociálních sítí sdílí stejné principy: přispívání všemi, jednoduše použitelné, jednoduché provázání informací, možnost vytváření verzí a webově založené.

KiWi-System se zaměřuje na prolomení hranic těchto systémů takovým způsobem, aby sloužil jako platforma pro implementaci a integraci mnoha různých druhů sociálně softwarových služeb. Má v úmyslu prolomit hranice informací tím, že dovolí uživatelům spojovat obsah novými způsoby, které jdou až za hranici uživatelských rozhraní, například pomocí sémantických anotací.

KiWi proto přebírá ideu wiki a posouvá ji na další úroveň tím, že vytváří nový druh sociální platformy, která umožňuje uživatelům sdílet a integrovat znalosti jednodušeji, přirozeněji a přiléhavěji. Také přizpůsobením obsahu a funkcionalit podle osobních požadavků.

Studie použití

Kromě vyvíjení jádra systému, KiWi projekt zahrnuje dvě studie použití pro podnikový management znalostí spolu s průmyslovými partnery. Navíc se KiWi zaměřuje na vývoj veřejného ukázkového použití, který ilustruje rysy KiWi systému v čistě sociálně softwarovém nastavení, například mimo celkem komplexní kontext podnikového řízení znalostí.

Extensions

V rámci projektu jsou také přidávány další rozšíření, které představují aplikace nad touto platformou. V prvním release KiWi projektu byly zahrnuty dvě takové aplikace, Wiki a TagIT. V novějších verzích se již objevuje i Blog. V rámci této diplomové práce je vytvořena aplikace, která je ve své podstatě dalším rozšířením. Pokud se mluví o spolupráci genealogické aplikace s jinými částmi systému, pak se hovoří právě o těchto existujících či vznikajících rozšířeních.

Grant pro KiWi projekt

Jak již bylo řečeno, projekt KiWi, kterého se účastní několik komerčních ale i nekomerčních institucí, byl podpořen grantem Evropské Unie, a to grantem European Community's Seventh Framework Programme (FP7/2007-2013), No. 211932.

Návrh aplikace implementované v rámci této diplomové práce

V rámci diplomové práce bude naimplementovaná genealogická webová aplikace pro platformu KiWi. Tato aplikace bude založena na navrženém RDF a tedy na sémantickém webu. Aplikace by měla umožňovat hlavně zadávání genealogických dat, samozřejmě jejich uložení, editaci a hlavně vizualizaci. Vzhledem k tomu, že v genealogii se jako výstupem poznatků o datech používá hlavně strom života neboli rodokmen, je hlavním požadovaným výstupem programu zobrazení právě tohoto stromu života. Je nutné ale mít k dispozici i datovou informaci o jedinci zahrnutém v tomto stromu života, takže je požadováno zobrazení podrobností i o každé zahrnuté osobě.

Obrazovky

Pro aplikaci jsou navrhovány 4 obrazovky. První základní obrazovka je určená jako rozcestník pro akce ke konkrétní osobě. Z tohoto rozcestníku jsou pak další obrazovky volány pomocí tlačítek. Za každým tlačítkem se skrývá nová obrazovka. Pro vytvoření nové osoby a zadání údajů o ní do databáze, slouží tlačítko Create person a následně zobrazená obrazovka Create.html. Pokud chceme existující osobu pouze upravit, pak toto lze učinit z obrazovky Update.html, která je vyvolána tlačítkem Update person ze základní obrazovky. Posledním tlačítkem na základní obrazovce, je Display family tree. Toto tlačítko vyvolá obrazovku Display.html, kde se zobrazí JavaApplet s rodokmenem dotyčné osoby.

Home – základní obrazovka, rozcestník

Jak již bylo řečeno, základní obrazovka reprezentovaná webovou stránkou Home.html slouží v zásadě jako rozcestník pro další funkcionalitu aplikace. Dále ji lze použít k zobrazení základních údajů, které jsou k dané osobě uložena v databázi.

Pokud je vybrána nějaká osoba (k výběru dochází například za pomoci extension Wiki pro KiWi, nebo v budoucnosti i za pomoci jiných aplikací), tak po přechodu na základní obrazovku pro extension Genealogie kliknutím na Genealogii z hlavního menu pro KiWi dojde k zobrazení detailů pro vybranou osobu. Jedná se o údaje jako jméno, příjmení, datum

narození, místo narození, pohlaví,... ale i seznam sourozenců, rodičů, dětí a partnerů. K dané vybrané osobě jsou k dispozici možnosti Editovat osobu (Update person), Zobrazit rodokmen (Display family tree). Dále je k dispozici možnost vytvoření nové osoby.

Pokud žádná osoba vybrána nebyla nebo vybraná aktuální položka není typu osoby, pak se na základní obrazovce zobrazí pouze informace o tomto faktu a zobrazí se tlačítko s možností zadání nové osoby do databáze (Create person).

The screenshot shows the KOWI Genealogy Application interface. At the top, there is a navigation bar with links for 'genealogy', 'TagIT', 'Inspector', 'Admin', 'Wiki', and 'Dashboard'. Below this, a 'History' bar shows 'StartPage - Martin Novák'. The main content area is titled 'Genealogy Application' and displays the profile for 'Martin Novák'. The profile information includes: First name: Martin, Last name: Novák, Gender: male, Status: alive, Date of birth: Tue Oct 16 00:00:00 CET 1951, Place of birth: Brno. There are sections for 'Partners' (Helena Malá), 'Siblings' (Jan Novák), and 'Children' (Michaela Nováková, Michal Novák). At the bottom of the profile, there are three buttons: 'Create new person', 'Update current person', and 'Display current person'. The footer of the page states 'Powered by KOWI - Knowledge in a Wiki (version 0.2-prerelease)'.

Create – obrazovka pro vytvoření nové osoby

Obrazovka reprezentovaná webovou stránkou Create.html slouží k zadávání nových osob do systému. Na této obrazovce se zadává:

- zobrazované jméno (Displayed name), které se používá ve většině výpisů
- jméno, příjmení osoby – First name, Last name
- pohlaví osoby – Gender, s možnostmi „Female“ (žena) nebo „Male“ (muž)
- informace o tom, zda je osoba ještě žijící – Living s možnostmi „Alive“ (žijící), „Dead“ (po smrti) „Not-known“ (není známo)
- datum a místo narození – Date of birth, Place of birth
- datum a místo úmrtí – Date of death, Place of death
- URI matky a URI otce – mother's URI, father's URI – reprezentující URI adresu na RDF rodičů; vzhledem k tomu, že políčka musí obsahovat přímo uri rodičů, je k dispozici možnost našeptávání existujících osob v databázi dle jména pro usnadnění výběru

Z genealogického hlediska není žádný ze zadávaných údajů povinný, neboť v historii často nejsou kompletní údaje známy, pokaždé může chybět nějaký jiný. Aby ale bylo zaručeno, že je alespoň jedno pole vždy vyplněné a při zobrazování od sebe byly jednotlivé osoby rozeznatelné, je nastaven údaj „zobrazované jméno“ jako povinný.

Obrazovka je ukončena tlačítky s možnostmi pro uložení/neuložení zadaných údajů a další navigaci. K dispozici jsou následující tlačítka:

- vytvoření osoby dle zadaných údajů a návrat na předchozí stránku (Create and back)
- vytvoření osoby dle zadaných údajů a přechod na obrazovku Update pro možnost další editace nebo doplnění zadaných údajů (Create and update)
- návrat zpět bez vytvoření nové osoby (Back)

The screenshot shows the 'Create new person' form in the KOWI Genealogy application. The form includes the following fields and options:

- Displayed name: [text input] eg.: John X. Doe
- First name: [text input] eg.: John
- Last name: [text input] Doe
- Gender: Female Male
- Alive status: Alive Dead Not-known
- Date of birth: [text input]
- Place of birth: [text input] e.g. Brno, CZ
- Date of death: [text input]
- Place of death: [text input] e.g. Prague, CZ
- Mother URI: [text input]
- Father URI: [text input]
- Partner URI: [text input] [Add]

Buttons at the bottom of the form: [Create and Back], [Create and Update], [Back].

Footer: Powered by KOWI - Knowledge in a Wiki (version 0.2-prerelease).

Update – obrazovka pro změnu údajů o existující osobě

Tato obrazovka slouží ke změnám prováděným nad již zadanými údaji, případně vyplnění ještě nezadaných údajů. Při zobrazení stránky Update.html, která reprezentuje tuto obrazovku, se zobrazí stejný formulář pro vyplnění osobních údajů dané osoby, jaký se zobrazuje na obrazovce pro vytvoření osoby, avšak s tím rozdílem, že tentokrát jsou políčka předvyplněna podle toho, jaké údaje jsou již uloženy v databázi.

Obrazovka je opět ukončena tlačítky s možnostmi pro uložení/neuložení zadaných údajů a další navigaci. K dispozici jsou následující tlačítka:

- uložení údajů a návrat na předchozí stránku (Update and back)
- uložení údajů, ale setrvání na stránce (Update)

- návrat zpět bez uložení (Back)
- obnovení dat na stávající údaje v databázi (Back to saved)

The screenshot shows the KiWi Genealogy application interface. At the top, there is a navigation bar with links for 'genealogy', 'TagIT', 'Inspector', 'Admin', 'Wiki', and 'Dashboard'. Below this, a 'History' bar shows 'StartPage - Martin Novák'. The main content area is titled 'Genealogy application' and contains a 'Change of person' form. The form includes fields for 'Person to be changed is' (Martin Novák), 'Displayed name', 'First name', 'Last name', 'Gender' (Female/Male), 'Status' (Alive/Dead/Not-known), 'Date of birth', 'Place of birth', 'Date of death', 'Place of death', 'Mother URI', 'Father URI', and 'Partner URI'. There are also buttons for 'Update', 'Update and Back', 'Back', and 'Back to saved'. A list of related entries is shown below, including 'Helena Malá' with a 'delete' button. The footer indicates the application is powered by KiWi - Knowledge in a Wiki (version 0.2-prerelease).

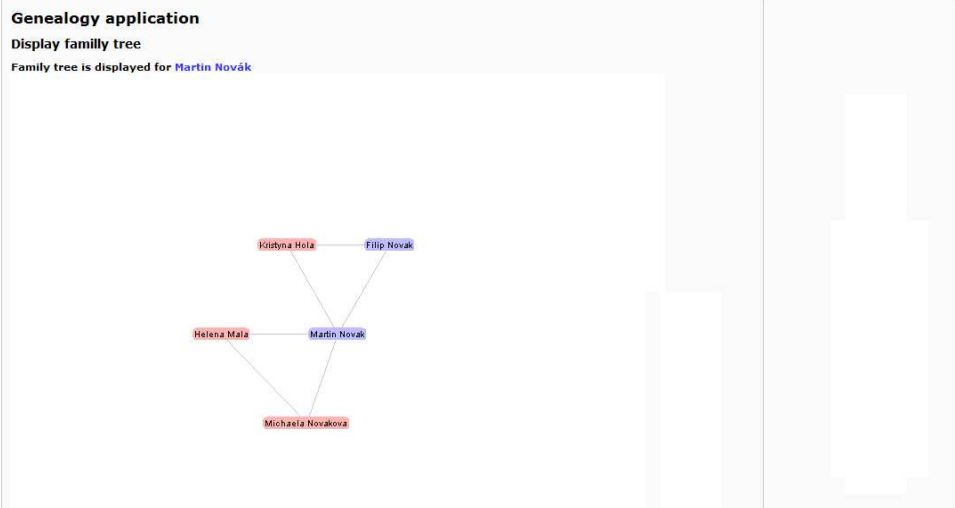
Display – zobrazení rodokmenu/stromu života

Stiskem tlačítka Display family tree z obrazovky Index se aplikace přesune na obrazovku Display, která je reprezentovaná webovou stránkou Display.html. Na této stránce dojde ke spuštění JavaAppletu, kde je k předem vybrané osobě zobrazen strom života.

Strom života je zobrazen defaultně pro pět generací, tj. generaci současnou k dané osobě, dvě generace předků a dvě generace potomků. Toto nastavení by mělo jít za běhu aplikace změnit pomocí tlačítek v horní části appletu.

V generaci současně pro danou osobu se zobrazuje daná osoba, její sourozenci a její partneři. V generacích předků se zobrazují vždy přímí předci, tj. rodiče dané osoby a dále rekurzivně rodiče rodičů atd. Ke každému z těchto přímých předků jsou pak zobrazeni také jejich sourozenci. U potomků jsou pak zobrazováni všichni přímí potomci, tj. vlastní děti, a dále rekurzivně děti těchto dětí. Dále je ke každému přímému potomkovi zobrazen partner.

Pokud daná osoba, pro kterou je rodinný strom zobrazován, nemá žádné potomky, případně počet generací potomků je menší než počet generací potomků požadovaných k zobrazení, budou zobrazeny pouze ty generace, které existují, tj. například pokud je někdo teprve dítětem a nemá vlastní potomky, pak při žádosti o zobrazení pěti generací dojde k tomu, že budou zobrazeny pouze tři, a to jedna současná a dvě generace obsahující předky.



Algoritmus pro vizualizaci

Co má být zobrazeno a proč

Pro zobrazení stromu života na obrazovce Display bylo nutno navrhnout algoritmus. Nejprve bylo potřeba ujasnit, které osoby budou zobrazovány, dále potom jaká informace má být zobrazena ke konkrétním osobám.

K vizualizaci jsou určeny vždy sourozenci vybrané osoby, přímí předci vybrané osoby a jejich sourozenci, přímí potomci vybrané osoby a jejich partneři. Zobrazení více osob by mohlo ztížit zobrazitelnost a tím i čitelnost pro uživatele, proto se zobrazení více osob nenavrhuje. Možnou změnou v tomto bodě je však počet zobrazovaných generací, který je defaultně nastaven na dvě generace předků, dvě generace potomků a jednu generaci, která je současnou pro aktuálně zobrazovanou osobu, tj. pro osobu, ke které daný strom života přísluší.

Ke každé osobě se bude zobrazovat Displayed name neboli zobrazované jméno, které je pro všechny osoby povinně zadáváno do databáze a proto se nemůže stát, že by někde chybělo. Dále bude u každé osoby rozlišeno, zda se jedná o muže nebo o ženu, neboť tato informace bývá zpravidla známá a proto často zadávána.

Princip algoritmu pro výběr členů rodiny k zobrazení

Princip celého navrhovaného algoritmu je ve vytvoření seznamu osob k zobrazení a jeho následnému zobrazení pomocí technologie Prefuse.

Inicializace seznamů

Na začátku algoritmu je nejprve inicializován takový počet seznamů, jaký počet generací má být zobrazován. Uvažuje se zde maximální počet generací, i kdyby některá z nich neměla v databázi uloženého žádného člena. Kontrola přítomnosti členů v každé generaci bude prováděna později před předáním seznamů k zobrazení pomocí prefuse. Také dojde k inicializaci seznamů pro uchování vztahů mezi jednotlivými členy. Těchto seznamů bude o jeden více než je generací. Dá se říci, že tyto seznamy budou reprezentovat jednotlivé rodiny, kde rodinou jsou myšleny rodiče + jejich děti, tj. nejbližší a nejuzší členy příbuzenstva.

Posloupnost přidávání osob v rámci generace

Přidávání osob předků do seznamů osob i rodin bude prováděno postupně tak, že ke každému členu budou nejprve přidáni do seznamu sourozenci včetně vlastního jedince a to tak, že pokud se jedná o ženu, budou vloženy nejprve sourozenci a potom daná žena, pokud se jedná o muže, bude přidán nejprve on a teprve za ním budou přidáni sourozenci. Tato posloupnost zajišťuje, že každý z partnerů/manželů bude mít své sourozence pouze na jedné straně a to tak, že mezi partnery nebudou zobrazeni žádní z nich, z čehož tedy vyplývá, že rodiče dětí budou vždy těsně u sebe a proto bude omezen počet křížení v grafu a i hrany grafu budou na délku minimální. I z toho hlediska, že v běžném rodokmenu se zobrazují pouze přímí předci a nikoliv jejich sourozenci, tak jejich zobrazení po stranách způsobí minimum problémů v čitelnosti.

U současné generace budou v závislosti na pohlaví zobrazované osoby přidány nejprve sourozenci, pak vlastní osoba a pak její partner (případně partneri, pokud jich bylo více) v případě, že se jedná o ženu. V opačném případě, tedy jedná-li se o muže, pak bude pořadí také opačné, tj. nejprve partnerka(y), potom aktuálně zobrazovaná osoba a potom teprve sourozenci. Toto pořadí bylo opět stanoveno na základě předpokladu, že tak dojde k nejmenším ztrátám na čitelnosti celého výsledného výstupu.

U potomků pak je jisté, že se budou zobrazovat všichni sourozenci, tito sourozenci ale na rozdíl od předků a i od současné generace mají stejnou váhu, tedy nejsou pouze nějakou navíc dodanou informací, ale jsou základní zobrazovanou informací. V tomto případě je dodávanou informací seznam partnerů těchto potomků. Takže se prochází vždy seznam sourozenců a pokud je osoba ženou, pak je ihned přidána do seznamu, následně za ní jsou přidáni její partneři a potom se přechází ke zpracování dalšího ze sourozenců. Pokud se jedná o muže, pak jsou nejprve přidány partnerky a následně teprve tento muž. Po té se opět přechází ke zpracování dalšího sourozence.

Výsledná posloupnost by tedy měla být následující:

Manželčiny sourozenci, manželka, manžel, manželovi sourozenci.

V případě více partnerů pak:

Manželčiny sourozenci 1, manželka 1, manželčiny sourozenci 2, manželka 2, manžel, manželovi sourozenci.

Posloupnost přidávání osob mezigeneračně

Jako první zpracovávanou generací je vždy současná generace k dané zobrazované osobě. Potom dochází ke zpracování generací předků a následně teprve generací potomků.

Po zpracování kompletně celé první generace se zpracovává nejprve matka zobrazované osoby a to tak, že se přidají nejprve její sourozenci, pak ona sama do téže generace a dojde k vyvolání zpracování matčiny matky stejným způsobem rekurzivně, dokud není dosažen počet generací, které se mají zpracovávat. Potom se postupně začnou zpracovávat otcové a to tak, že se přidají do seznamu své generace a za nimi jejich sourozenci a pak se přejde ke zpracování jejich matky, pokud ještě má být její generace zpracovávána.

Postup pro dvě generace předků by byl tedy následující:

Současná generace, matčiny sourozenci, matka, babiččiny (matka matky) sourozenci, babička, dědeček (otec matky), sourozenci dědečka, otec, otcovi sourozenci, babiččiny (matka otce) sourozenci, babička, dědeček (otec otce), dědečkovi sourozenci.

Pro zpracování generací potomků pak bude docházet k postupnému zpracování od nejstarší generace po nejmladší. Nejprve budou zpracovány všechny děti osoby, pro kterou je strom života zobrazován, a to způsobem popsaným dříve v kapitole o zpracovávání osob v rámci jedné generace. Poté se přejde ke zpracování další časově mladší generace. Zde budou zpracovávány děti potomků z předchozí generace a to od leva do prava postupně. Vzhledem k tomu, že by zde měli být zobrazeni pouze vlastní přímí potomci, je třeba vynechat ze zobrazení takové potomky, které má partner přímého potomka z jiného partnerství.

Posloupnost při zpracování potomků je tedy například následující (vzhledem k tomu, že počet potomků se různí na rozdíl od počtu rodičů a prarodičů, je zde uveden příklad se třemi dětmi, které mají každý dvě vlastní děti):

V první generaci potomků: partnerka(y) syna 1, syn 1, dcera 2, partner(ři) dcery 2, partnerka(y) syna 3, syn 3; v druhé generaci potomků: partnerka(y) vnuka 1 (syn syna 1), vnuk 1, vnučka 1 (dcera syna 1), partner(ři) vnučky 1, partnerka(y) vnuka 2 (syn syna 2), vnuk 2, vnučka 2 (dcera syna 2), partner(ři) vnučky 2, partnerka(y) vnuka 3 (syn syna 3), vnuk 3, vnučka 3 (dcera syna 3), partner(ři) vnučky 3.

Pomocné funkce pro přidávání osob do seznamů

Vzhledem k tomu, že v databázi jsou uloženy pouze vazby na matku, otce a partnery, pak je třeba doimplementovat některé pomocné funkce ke získání dalších členů rodiny. Je proto naimplementována funkce pro získání seznamu sourozenců a dále také funkce pro získání seznamů dětí. Tento seznam je vrácen nezávisle na tom, kdo je druhým rodičem dětí, neboť je v praxi běžné, že jedna osoba má děti s více partnery. Také u seznamu sourozenců je bráno v úvahu, že někteří mohou být částečně nevlastní, tj. mají se zobrazovanou osobou pouze jednoho společného rodiče. Pokud existují takto nevlastní sourozenci, pak by měla být jejich posloupnost v seznamu následující:

Zpracovávaná osoba, pokud je mužem, vlastní sourozenci, nevlastní sourozenci ze strany matky, nevlastní sourozenci ze strany otce

nebo

nevlastní sourozenci ze strany matky, nevlastní sourozenci ze strany otce, vlastní sourozenci, zpracovávaná osoba pokud je ženou.

Vytvoření seznamů rodin

Seznam rodin ve své podstatě je vklíněn vždy mezi 2 seznamy generací. Je proto doplňován vždy, když se zpracovávají členové z obou těchto generací. Každá rodina v sobě v podstatě zahrnuje matku, otce a seznam dětí a ve výsledném zobrazení pomocí prefuse bude použita pro vytvoření uzle grafu takového, kde se všichni členové dané rodiny setkají.

Algoritmus a Prefuse

Při sestavování aplikace v prefuse se používá většinou následující obecný vzor[12]:

1. Nahrání dat, která mají být vizualizována, do datových struktur, které jsou definovány přímo v prefuse. Zde se využívá hlavně definovaných datových struktur tabulka, graf nebo strom. Tyto data jsou nahrána buď načtením ze souboru (například XML, CSV...), nahráním z databáze a nebo pomocí vlastního datového zdroje.
2. Vytvoření vizualizace, která mapuje nahraná data do tzv. Vizuelní abstrakce. Tabulky, grafy nebo stromy jsou přidány do vizualizace a je jim dán jedinečný název datové skupiny pro pozdější referenci.
3. Vytvoření tzv. RendererFactory, neboli továrny na vykreslování, a její zaregistrování do existující vizualizace. Tato továrna je zodpovědná za přiřazení vykreslovacích technik k vizualizovaným položkám.
4. Vytvoření série akcí (Action) pro zpracovávání dat. Tyto akce operují nad vizuelní abstrakcí a mohou zahrnovat cokoli, pro uvedení příkladů tyto operace zahrnují například nastavení polohy, barvy, velikosti a tvaru vizualizovaných položek nebo animace těchto vlastností mezi různými konfiguracemi. Instance těchto akcí mohou být seskupovány do seznamů akcí (ActionList) pro provedení různých zpracovávajících úloh. Akce, které jsou vyvolány přímo, jsou přidávány do vizualizace a jsou pojmenovány unikátním jménem, pomocí kterého jsou později referencovány.
5. V tomto kroku probíhá inicializace jednoho nebo více zobrazení (Display) pro zobrazení a manipulaci s vizuelními položkami. Dále je zde zahrnuta specifikace interaktivního chování, čehož je dosaženo přidáním ovládacích prvků (Control) k zobrazení. Vyhledávání a filtrování nad datovými položkami může být přidáno pomocí dynamického připojení dotazů.

Vzhledem k tomu, že ne všechny kroky jsou stejně rozsáhlé a u některých by bylo možné uvést maximálně jeden řádek kódu, jsou níže rozvinuty pouze ty kroky, které si zaslouží větší pozornost a jejichž implementace byla v něčem kreativní.

Ad krok 1: Algoritmus pro nahrání dat do datových struktur nástroje perfuse

Pro vizualizaci genealogických dat je vhodné zobrazovat graf. Stromová struktura není příliš vhodná, neboť se v ní předpokládá, že každý uzel má pouze jeden rodičovský uzel, což v praxi příliš nevyhovuje. Tabulková data také příliš nepasují na představu genealogického stromu života, takže nakonec pro zobrazení tohoto stromu života bylo použito datové struktury grafu, který je definován v knihovně perfuse.

Datová struktura grafu je složena ze dvou tabulek. Z tabulky uzlů (nodes) a z tabulky hran grafu (edges). Na začátku je tedy nutno vytvořit tyto tabulky. Pro jejich vytvoření byla použita následující, i když ne jediná, možnost, jak je vytvořit. Při běhu aplikace se vytvoří nejprve tabulka uzlů a to na základě seznamu osob z generací, jejichž vytvoření bylo popsáno dříve v této kapitole.

Tabulka uzlů je inicializována nejprve jako prázdná s nula řádky a nula sloupci. Po inicializaci je vytvořena struktura sloupců a to přidáním sloupců pro content item dané osoby, dále zobrazované jméno a pohlaví osoby reprezentované tímto content item. Dále jsou přidány dva sloupce typu celých čísel, kde jeden reprezentuje generaci, ze které daná osoba je, druhý pak reprezentuje její pozici v seznamu na základě pořadí v generaci. Poté jsou procházeny všechny generace a všechny osoby v nich jsou transformovány na řádky v tabulce. Každý řádek má svůj identifikátor, který je vytvořen automaticky při ukládání jako nejmenší číslo, které ještě není v tabulce použito. Pokud není s daty v tabulce příliš operováno (editace, mazání), pak je identifikátor stejný jako pořadí přidávaných řádků.

Tabulka hran je inicializována podobně jako tabulka uzlů, tj. nejdříve jako prázdná tabulka bez řádků a bez sloupců. Sloupce jsou zde následně přidány 3 a to jako zdroj hrany (from), cíl hrany (to) a typ hrany. Pro zdroj a cíl hrany reprezentuje číslo řádku v tabulce uzlů, typ hrany pak vztah mezi těmito dvěma osobami a to takovým způsobem, že například vztah matka (řádek 1) – dcera (řádek 2) je reprezentován jako (from = 1, to = 2, type = „mother“) nebo (from = 2, to = 1, type = „daughter“) . Je snaha uchovávat pouze typy „mother“, „father“, „partner“, které by měli postačovat pro vytvoření všech hran. Pro přidání řádků do této tabulky jsou pak použity seznamy rodin. Pro každou rodinu je přidána hrana mezi matkou a otcem typu partner, potom mezi všemi osobami ze seznamu dětí jsou vytvořeny dvě hrany, jedna k matce, jedna k otci.

Při inicializaci grafu jako objektu, je voláno vytvoření nového objektu tak, že jsou použity tabulky s uzly, nastaveno, že se nejedná o orientovaný graf a dále jsou zadány názvy sloupce zdroje a sloupec cíle z tabulky hran. Poslední parametr při vytvoření tabulky udává na jaký sloupec z tabulky uzlů se odkazují sloupce pro zdroj a cíl z tabulky hran. V tomto případě je nechána defaultní hodnota a to znamená, že je použito číslo řádku.

Ad krok 4: Vytvoření akcí pro zpracování dat

Tyto akce by měly zahrnovat hlavně nastavení vlastností zobrazovaným uzlům a hranám. Proto je naimplementováno vytvoření palety barev pro zobrazování uzlů – a to odstín modré pro zobrazování mužů a odstín červené pro zobrazování uzlů reprezentující ženy. Tato paleta je tedy použita pro vytvoření akce FillColor pro vybarvení uzlů. Dále je navržena černá barva pro akci TextColor pro uzly. Tato funkce by se měla postarat o zobrazení textů uzlů černou

barvou. Poslední implementovanou akcí je obarvení hran grafu. Barva je nastavena na odstín šedé a je volána u akce StrokeColor. Všechny tři akce jsou pak přidány do jediného seznamu akcí, který je přidán do vizualizace.

Použité technologie a nástroje

Zde je uveden seznam hlavních technologií a nástrojů, které byly použity pro vývoj aplikace v rámci této diplomové práce. Jejich volba byla částečně dána tím, že aplikace byla vyvíjena pro platformu KiWi, takže asi hlavní otázkou je, proč byla vybrána právě tato platforma.

K výběru platformy lze říct, že se jedná v současné době o běžící projekt, což na jednu stranu sebou nese nevýhody toho, že ještě ne vše funguje zcela bezchybně. Na druhou stranu je to moderní platforma, která se svižně dále rozvíjí. Pro její potenciální rozvoj hovoří i to, že je jí udělen grant od organizace European Community's Seventh Framework Programme. Dalším pozitivem, které hovoří pro tuto aplikaci, je i to, že se jedná o platformu určenou pro vývoj sémanticko-sociálních aplikací, které jsou v současnosti moderní a tak říkájíc „v kurzu“. Jedním z dalších důvodů, proč byla tato technologie vybrána, je skutečnost, že na rozvoji této platformy se nezanedbatelnou měrou podílí i naše fakulta.

Java – programovací jazyk

Dle [3] (volný překlad): Java je programovací jazyk původně vyvinutý ve společnosti Sun Microsystems a vydaný v roce 1995 jako základní komponenta Java platformy od téže společnosti. Tento jazyk přebírá mnoho syntaxe z C a C++, ale má jednodušší objektový model a méně nízkoúrovňových možností. Java aplikace jsou typicky kompilovány do bytecodeu a mohou tak být spuštěny na kterékoliv Java virtuálním stroji (Java virtual machine - JVM) nehledě na architekturu počítače.

Originální implementace Java překladače, virtuálních strojů a tříd knihoven byl rozvinut od roku 1995. V květnu roku 2007 už pak v souladu se specifikacemi Java Community Process, Sun umožnil použití většiny jejich Java technologií jako free software pod licencí GNU (General Public Licence).

Tím se tato technologie stala výhodnou pro použití v rámci projektu KiWi. Vzhledem k tomu, že JBoss Seam staví na Javě, bylo nanejvýš vhodné použít tento programovací jazyk také k implementaci rozšíření pro KiWi, které bylo vytvořeno v rámci této diplomové práce.

Java Enterprise Edition

Java platforma, Enterprise Edition nebo zkráceně Java EE je široce používaná platforma pro serverové programování v Java programovacím jazyce. Java platforma Enterprise Edition se liší od Java Standard Edition Platform (Java SE) hlavně v dalších přidaných knihovnách, které poskytují funkcionalitu pro nasazení chyby tolerující, distribuovaný, vícevrstvý Java software, široce založený na modulárních komponentách běžících na splikačním serveru.

Tento objektově orientovaný programovací jazyk je základem pro JBoss Seam technologie. Proto bylo třeba jej využívat v rámci kódu, kde se Seam a Java prolínají, ale byl také použit pro tvorbu vlastních tříd, které se často Seamu dotýkají jen okrajově.

Pro implementaci aplikace pro tuto diplomovou práci bylo použito JRE i JDK verze 1.6.0.

JBoss Seam

JBoss Seam je celkem nový mocný aplikační framework pro budování nové generace webovských 2.0 aplikací. Tento nástroj sjednocuje a integruje technologie jako je asynchronní JavaScript a XML (AJAX), Java Server Faces (JSF), Enterprise Java Beans (EJB3), Java Portlets a Business Process Management (BPM).

Seam byl navržen od základů tak, aby eliminoval komplexitu na aplikační a API úrovni. Dovoluje vývojářům sestavit komplexní webové aplikace jednoduše anotováním Plain Old Java Objects (POJOs), komponováním UI widgets a částečně také XML. Jednoduchost Seamu v budoucnu umožní snazší integraci s JBoss Enterprise Service Bus (ESB) a Java Business Integration (JBI).

Tento nástroj byl vybrán pro implementaci platformy KiWi, a proto byl využit při implementaci aplikace pro tuto diplomovou práci. Proč byl JBoss Seam vybrán jako nástroj pro implementaci KiWi?

Jistě hlavním důvodem bylo to, že tato platforma je určená pro webové 2.0 aplikace, které jsou tedy dle označení sémantické, což je platforma KiWi také. Jistě k tomu vedli i některé další důvody, které jsou uvedeny na oficiálních stránkách viz [13]. Zde je možné se dočíst více o výhodách jako je zpětná kompatibilita s J2EE, nebo například že JBoss Seam je prvním aplikačním frameworkem EJB 3.0. Dále je zde uvedeno, že tato technologie umožňuje jednoduchý způsob, jak implementovat AJAX aplikace, které jsou také částečně využity v platformě KiWi. Jistě pro KiWi podstatným důvodem je revoluční přístup v řízení stavu. Dalším ale neposledním důvodem je jistě snadné integrační testování, které je díky POJOs možné už na unit úrovni, tj. na úrovni nejmenší testovatelné jednotky.

Pro implementaci aplikace pro diplomovou práci byla použita verze JBoss 4.2.3 GA a JBoss-Seam 2.1.1 GA.

KiWi

Jak již bylo řečeno, KiWi je nově vznikající platforma pro vývoj sémanticko-sociálních aplikací. Má potenciál do budoucna díky vzrůstajícímu trendu právě tímto způsobem zaměřených aplikací. Celá platforma je implementována na technologiích Java a JBoss Seam.

Vývoj aplikace pro potřeby této diplomové práce proběhl nad verzí zdrojových kódů shrnutých pod pojmenováním Revision 1679.

Java Applet

Java applet je takový applet, který je zpřístupněn uživatelům ve formě Java bytecode. Java applet je spuštěn internetovým prohlížečem za použití Java Virtual Machine (JVM) nebo pomocí aplikace AppletViewer společnosti Sun, což je nástroj na testování appletů

samostatně. Java applety byly uvedeny již v první verzi programovacího jazyka Java v roce 1995. Java applety jsou obvykle psány právě v programovacím jazyce Java, ale mohou být psány i v jiných jazycích, které jdou zkompilovat na Java bytecode. Příkladem takového jazyka je třeba Python, ale pro tento diplomový projekt je nutno podotknout, že i Java ve spolupráci s JBoss Seam lze přeložit jako applet.

Applety jsou používány pro poskytnutí interaktivních vlastností webovým aplikacím a to hlavně takových vlastností, které není možné interpretovat pomocí HTML. Další obrovskou výhodou appletu je, že bytecode je nezávislý na operačním systému. Tyto vlastnosti a hlavně fakt, že je potřeba vytvořit interaktivní webovou aplikaci, vedli k rozhodnutí použít Java applet pro vizualizaci.

Eclipse

Eclipse je vícejazyčná platforma pro vývoj softwaru obsahující IDE a plug-in system pro možnost jejího rozšíření. Primárně je tato platforma napsána v jazyce Java a je používána pro vývoj aplikací právě v tomto jazyce a také díky prostředkům jako plug-in je používána pro programování v dalších jazycích jako C, C++, COBOL, Python, Perl, PHP a další.

Tato platforma je defaultně určená pro Java developery, proto také obsahuje Java Development Tools (JDT). Eclipse byl vydán pod licencí Eclipse Public License a tím se tato platforma řadí mezi open source software.

Vzhledem k tomu, že je tato platforma určena hlavně pro implementaci v Javě a navíc není pro tuto aplikaci překážkou ani JBoss Seam a Ant, které jsou také pro implementaci použity, stala se tato platforma ideálním nástrojem pro implementaci aplikace. Pro vlastní implementaci byl použit Eclipse, verze 3.4.2.

Apache Ant

Podle [14] je Apache Ant je na Javě založený sestavovací (build) nástroj. Teoreticky je to nástroj stejného druhu jako Make, ale bez vrásek, které Make má.

I přesto, že existuje již řada jiných build nástrojů jako make, gnumake, nmake, jam atd., tak vzhledem k tomu, že tyto všechny nástroje mají svoje limity, které nevyhovovaly originálnímu autorovi Ant nástroje, a to hlavně v případě, že se jednalo o software vytvářený pomocí více platform. Nástroje jako make jsou neodmyslitelně založené na shellu, takže vyhodnocují množinu závislostí, potom zpracovávají příkazy podobné příkazům shellu. Toto umožňuje vytvořit jakýkoliv program pro operační systém, na kterém pracujete, ale již to limituje vývojáře pouze na jeden konkrétní operační systém nebo typ operačního systému.

Ant toto řeší tak, že místo modelu, který obsahuje příkazy shellu, používá třídy v jazyce Java. Místo psaní shellovských příkazů jsou vytvářeny konfigurační soubory na základě XML volající cílový strom, kde jsou prováděny různé operace. Každá operace běží za pomoci objektu, který implementuje konkrétní interface pro danou operaci.

Ant nemá takovou vyjadřovací schopnost jako shell, ale na druhou stranu zde existuje možnost použít příkaz z shellu, což je ale opět závislé na operačním systému.

Vzhledem k tomu, že KiWi je aplikace sestavovaná nad množstvím platform, je jistě pochopitelné, že pro sestavování je používán právě nástroj Ant.

Prefuse – visualization toolkit

Prefuse[12] – vizualizační nástroj, je sada softwarových nástrojů pro vytváření vysoce interaktivních datových vizualizací. Originální Prefuse toolkit nabízí vizualizační framework pro programovací jazyk Java, Prefuse Flare toolkit pak poskytuje vizualizační a animační nástroje pro ActionScript a Adobe Flash Player. Vzhledem k tomu, že v této práci je vizualizace prováděna pomocí JavaAppletu, je jistě zřejmé, že je použit Prefuse toolkit.

Prefuse porporuje širokou sadu možností pro modelování dat, vizualizaci a interakci. Poskytuje optimalizovanou datovou strukturu pro tabulky, grafy a stromy, množství technik pro uspořádání a vizuální kódování, a podporuje animace, dynamické řady, integrované vyhledávání a konektivitu na databázi.

Prefuse je napsán v Javě, za použití Java 2D grafické knihovny a je jednoduše integrovatelný do Java Swing aplikací nebo webových appletů. Prefuse je licencován pod licencí BSD a může tak být volně použit jak pro komerční tak pro nekomerční účely.

Pro implementaci byla použita knihovna Prefuse-Beta, release z 20.10.2007.

Přínosy, možná rozšíření a použití v praxi

Vzhledem k tomu, že se tato diplomová práce nezaobírá všemi možnostmi do detailů, je možné na ni navázat a některé její přínosy rozvinout, případně jiné nově přidat.

Přínosy diplomové práce

Mezi hlavní přínosy této diplomové práce lze jistě zahrnout vytvoření nové genealogické aplikace, vytvoření RDF pro genealogická data, obohacení sémantického prostředí KiWi o další extension.

Nová genealogická aplikace má proti většině současných tu výhodu, že je webová a navíc je vybudovaná na základech sémantického webu, což ji předurčuje k možnosti použití do budoucnosti, kdy má sémantický web pomalu nahrazovat současný web.

Pokud by se do budoucna nepoužívala aplikace, je možné z ní použít alespoň navrhovaný RDF formát pro genealogická data, neboť tento formát již bude součástí platformy KiWi, která umožňuje tvorbu dalších sémantických aplikací.

Do platformy KiWi, která je v současné době stále ve vývoji, bylo dodáno nové rozšíření (extension), které nejen umožňuje být použité samo o sobě jako již zmiňovaná aplikace, ale také umožňuje jiné sémantické aplikaci nad touto platformou spolupracovat s tímto novým rozšířením a získávat tak možná potřebná data například při použití dotazování jazykem OWL.

Možná rozšíření diplomové práce

Tato diplomová práce se svým rozsahem některých úzce souvisejících problémů dotýká pouze okrajově, na tyto problémy lze ale jistě navázat a rozvinout je v další práci.

V diplomové práci není například zahrnuta a do větší míry dořešena problematika ochrany osobních údajů.

Dále zde nejsou uvedeny žádné teoretické možnosti, jak je možno dohledávat genealogická data v existujících datech po celém Internetu, nejen v datech, která byla zadána přímo prostřednictvím této aplikace.

Lze také doimplementovat použití OWL jazyka pro dotazování nad současnými daty a to například i ve spolupráci s dalšími sémantickými aplikacemi.

Vzhledem k tomu, že základním stavebním kamenem celého stromu života je jedinec, je možné rozšířit i RDF o další životopisné údaje, které mohou dále sloužit k vyhledávání pomocí OWL.

Možná rozšíření aplikace

Aplikace jako taková v současné době neobsahuje import nebo export jiných genealogických dat. Určitě by proto bylo přínosem, kdyby byla takováto možnost naimplementována. Minimálně možnost převodu z a do formátu GEDCOM by byla jistě využívána.

Je také možno zahrnout do aplikace již zmiňovaný problém ochrany osobních dat a zobrazovat pouze data, pro která má určitá osoba oprávnění. Případně bude třeba ještě naimplementovat funkce na správu těchto oprávnění.

K aplikaci lze doimplementovat možnost vyhledávat a dotazovat se pomocí OWL jazyka. Toto lze propojit i s jinými existujícími i nově přidanými aplikacemi v rámci KiWi, případně i jiných webových aplikací, pokud k nim bude doimplementováno patřičné rozhraní.

Pro vizualizaci genealogických dat lze naimplementovat různé možnosti, například přidat zobrazování dalších vztahů ve stromu života kromě existujících, nebo kompletně jiná vizualizace ať už s pomocí stávající nebo úplně jiné technologie.

Vzhledem k požadavkům na intuitivní ovládání aplikace a hlavně grafická rozhraní, je možné do existujícího JavaAppletu doimplementovat možnosti přidání a editace osob přímo prostřednictvím appletu. Applet by pak byl rozhraním pro kompletně celou aplikaci, nejen pro zobrazení stromu života.

Použití aplikace v praxi

Pro vytvořenou aplikaci se dá v praxi najít několik uplatnění. Je ale nutno uvažovat nad tím, že aplikace jako taková není určená primárně pro konkrétní možné použití, ale že pro některé případy by bylo třeba aplikaci dotvořit tak, aby lépe vyhovovala svému použití a aby poskytovala i další nezbytnou funkcionalitu.

Jedním z případů, kdy by bylo nutno aplikaci rozšířit, je například použití aplikace jako databáze osob, jako matriku pro úřady. Pro tento účel by bylo třeba doplnit RDF o některé další údaje, případně dotvořit nové RDF, na které by se původní RDF odkazovalo v jedné z trojic. Takovýto systém by pak mohli používat například nejen státní instituce jako je matrika, sociální úřad, policie, ale také například právníci při dohledávání pozustalých při řešení dědických vyrovnání a sporů.

Pokud by byla aplikace obohacena hlavně o možnosti jazyka OWL a dohledávání tak nových vztahů mezi jednotlivci, určitě by našla široké uplatnění mezi historiky, kteří se rádi zabývají otázkami typu „Mohla daná osoba potkat tuto konkrétní osobu?“ nebo například odhalit ještě neodhalené vazby mezi panovníky.

V praktickém použití aplikace nelze opomenout ani nadšené fanoušky genealogie, kteří rádi vytvářejí své vlastní rodokmeny. V takovém případě by byl tento nástroj skvělým úložištěm jejich dat a nástrojem pro jejich vizualizaci. Pokud by byla dořešena otázka problému ochrany osobních údajů, lze si snadno představit, že na základě údajů jednoho sběratele, mohou požádat strýčka z pátého kolene, aby doplnil tu část rodokmenu, kterou nejlépe zná on, a tak by bylo možné sestavit mnohem větší rodokmeny, což bývá cílem všech, pro něž se stala genealogie koníčkem.

Závěr

V průběhu diplomové práce jsem se seznámila s technologiemi jako Seam, JBoss, KiWi, Prefuse a spojila je s dalšími metodami jako je sémantický web, RDF. Tyto prostředky jsem spolu s některými dalšími známějšími zkombinovala do webové aplikace pro vizualizaci genealogických dat. Práce se neobešla bez problémů s technologiemi, které jsou stále ve vývoji, například v samotné platformě KiWi jsem našla několik problémů znemožňujících implementaci, tyto chyby však byly odstraněny za spolupráce s diplomovým vedoucím.

V dokumentu diplomové práce je uvedeno, které části těchto technologií byly použity, jakým způsobem byly zkombinovány, proč byly zkombinovány takovým způsobem a co tato kombinace přinesla.

Kvůli pochopení vizualizovaných dat byla zahrnuta kapitola, která pojednává o genealogii, proto, že právě genealogická data jsou vizualizována pomocí naimplementované aplikace pro sémantický web.

Je nutno uvést, že nejen genealogická data jako taková byla vstupem pro implementaci vizualizace, ale že na základě genealogických pojmů a poznatků získaných prozkoumáním konkrétních existujících genealogických formátů jako GEDCOM, GEDCOM XML nebo OPC byl navržen konkrétní RDF pro reprezentaci genealogických dat. Tento formát je používán ve vytvořené genealogické aplikaci pro sémantický web na platformě KiWi. V budoucnu je jistě vhodným krokem její rozšíření o další funkcionalitu jako například importování dat z jiných databází nebo jejich export

Přínosem této diplomové práce bylo osvojení si některých technik jak pro vizualizaci, tak pro sémantický web. Avšak jako hlavní přínos, na který mohou navázat i jiní, byla naimplementována aplikace pro zadávání, úpravu a vizualizaci genealogických dat.

Literatura

- [1] Urbánek Antonín. *Základní pojmy heraldiky a genealogie*. Dokument dostupný na URL http://www.zsru.cz/Her_gen/zakl_pojmy.htm
- [2] Ilčík Václav. *Základní pojmy genealogie*. Dokument dostupný na URL <http://www.ilcik.cz/genealogie/zacatecnik/pojmy/index.htm>
- [3] Wikipedie, Otevřená encyklopedie. Použito více dokumentů ohledně genealogie, ontologie, RDF, sémantického webu. Dokumenty dostupné na <http://cs.wikipedia.org/wiki/>
- [4] Česká genealogická a heraldická společnost v Praze. *Genealogie aneb rodopis*. Dokument dostupný na URL <http://www.genealogie.cz/genealogie/>
- [5] CodePlex, Open Source Community. *Family.Show project*, 2009. Dokument dostupný na URL <http://www.vertigo.com/familyshow.aspx>
- [6] Askren Jay. *The Semantic Web for Family History*, 18. srpna 1997. Dokument dostupný na URL <http://jay.askren.net/Projects/SemWeb/>
- [7] Tim Berners-Lee, James Hendler, Ora Lassila. *The Semantic Web*, květen 2001. Dokument dostupný na URL <http://www.sciam.com/article.cfm?id=the-semantic-web>
- [8] Matulík Petr, Pitner Tomáš. *Sémantický web a jeho technologie*, únor 2004. Dokument dostupný na URL <http://www.ics.muni.cz/zpravodaj/articles/296.html>
- [9] Gál Radim. *Sémantický web 1. díl, úvod*, 11. prosince 2007. Dokument dostupný na URL <http://www.owebu.cz/internet/vypis.php?clanek=1793>
- [10] Kolektiv autorů. *Collaborative Knowledge Management, powered by the Semantic Web*. Dokument dostupný na URL <http://www.kiwi-project.eu/>
- [11] Beiträge von Wastl. *First KiWi Open Source Release!*, 27. února 2009. Dokument dostupný na URL <http://www.schaffert.eu/2009/02/27/first-kiwi-open-source-release/>
- [12] Kolektiv autorů. *The prefuse vizualization toolkit*, 24. ledna 2009. Dokument dostupný na URL <http://prefuse.org/>
- [13] Kolektiv autorů. *JBoss Seam*, 2009, Dokument dostupný na URL <http://www.jboss.com/products/seam/>
- [14] Kolektiv autorů. *Apache Ant*, 14. května 2009 Dokument dostupný na URL <http://ant.apache.org/index.html>