

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství



Bakalářská práce

Objektová databáze historických dat

Jiří Bernas

© 2011 ČZU v Praze

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Jiří Bernas

obor Informatika

Vedoucí katedry Vám ve smyslu Studijního a zkušebního řádu ČZU v Praze čl. 16 určuje tuto bakalářskou práci.

Název tématu: **Objektová databáze historických dat**

Struktura bakalářské práce:

1. Úvod
2. Cíl práce a metodika
3. Literární rešerše
4. Identifikace potřeb cílové skupiny
5. Identifikace dat, které budou v systému zachyceny
6. Vytvoření objektového modelu struktury dat
7. Návrh historiografického systému
8. Vytvoření funkčního prototypu systému s využitím volně dostupného SW a s možností webového přístupu
9. Závěr
10. Seznam literatury
11. Přílohy

Rozsah původní zprávy: 30 - 40 stran

Seznam odborné literatury:

Arlow, Jim. UML 2 a unifikovaný proces vývoje aplikací : objektivě orientovaná analýza a návrh prakticky. 2., aktualizované a doplněné vydání. Brno : Computer Press, 2007. 567 s. ISBN 978-80-251-1503-9

Merunka, Vojtěch. Objektové modelování. 1. vydání. Praha: Alfa 2008, 197 s. ISBN 978-80-87197-04-2

Young, Michael J. XML krok za krokem. 2. vydání. Brno: Computer Press, 2006. 471 s.+ 1 CD-ROM. ISBN 80-251-1070-2

MARC 21 : bibliografický formát. 1. české vydání. Praha: Národní knihovna České republiky, 2003. 2 sv. ISBN 80-7050-427-7

Vedoucí bakalářské práce: **doc. Ing. Vojtěch Merunka, Ph.D.**

Termín odevzdání bakalářské práce: duben 2010

Tran v

.....
Vedoucí katedry



Jan

.....
Děkan

V Praze dne: 19.11.2008

Čestné prohlášení

Prohlašuji, že svou bakalářskou práci "Objektová databáze historických dat" jsem vypracoval(a) samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor(ka) uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil(a) autorská práva třetích osob.

V Praze dne 30. 11. 2011

Poděkování

Rád(a) bych touto cestou poděkoval(a) doc. Ing. Vojtěchu Merunkovi, Ph. D. za odborné vedení a cenné podněty.

Objektová databáze historických dat

Object database of historical data

Souhrn

Práce se zabývá návrhem databáze pro evidenci historických dat na základě objektového modelování. Jedná se o problematiku, která je v současnosti řešena převážně na principu relačních databází. Téma práce v sobě spojuje všechny úrovně návrhu informačního systému od definice cíle, přes analýzu, návrh a realizaci systému. Cílem bylo vytvořit funkční vzorek objektové databáze. Analytická část práce byla věnována identifikaci zdrojů informací a hledání minimálního společného protokolu. Na základě uvedených kroků pak byl vytvořen funkční vzorek systému.

Summary

The work addresses the design database for recording historical data based on object modeling. This issue is currently addressed through relational databases. Topic of study combines all levels of information system design from the definition of objectives, through analysis, design and implementation of the system. The aim was to create a functional object database sample. The analytical part of the work was devoted to identifying resources and finding the basic protocol. Based on the above steps, then the sample was created by functional system.

Klíčová slova: informační systém, objektové modelování, databáze

Keywords: information system, object modeling, databases

Obsah

1	Úvod.....	4
1.1	Obecně o práci historika a důležitosti poznávání historie.....	4
1.2	Význam ICT při této práci.....	4
1.3	Důvod volby tématu práce.....	4
2	Cíl práce.....	5
3	Metodika.....	6
4	Literární řešerše.....	8
4.1	Objektové modelování.....	8
4.2	Popisné standardy.....	8
4.2.1	MARC 21.....	8
4.2.2	EAD.....	8
4.2.3	Standard pro ukládání a zasílání archivních pomůcek druhu inventář a dílčí inventář v digitální podobě.....	9
4.2.4	Dublin Core.....	9
5	Identifikace cílové skupiny a jejích potřeb.....	11
5.1	Cílová skupina.....	11
5.2	Potřeby cílové skupiny.....	11
6	Identifikace objektů, které budou v systému zachyceny a definice tříd.....	12
6.1	Téma.....	12
6.2	Osoba.....	13
6.3	Instituce.....	14
6.4	Místo.....	14
6.5	Pramen.....	15
6.6	Literatura.....	19
6.7	Další údaj.....	20
6.8	Minimální protokol.....	21
7	Vytvoření funkčního prototypu systému s využitím volně dostupného SW a s možností webového přístupu.....	22
7.1	Realizace tříd.....	22
7.2	Realizace kolekcí.....	22
7.3	Umístění funkčního vzorku.....	22
8	Závěr.....	23
9	Použitá literatura.....	24

Seznam tabulek

Tab 1 Seznam atributů a metod třídy téma.....	13
Tab 2 Seznam atributů a metod třídy osoba.....	13
Tab 3 Seznam atributů a metod třídy instituce.....	14
Tab 4 Seznam atributů objektu místo.....	14
Tab 5 Seznam atributů objektu pramen.....	18
Tab 6 Srovnání popisu knihy v MARC21.....	19
Tab 7 Seznam atributů objektu literatura.....	20
Tab 8 Seznam atributů objektu další.....	21
Tab 9 Minimální protokol.....	21

1 Úvod

1.1 Obecně o práci historika a důležitosti poznávání historie.

Americký filosof George Santyana ve svém díle Život rozumu napsal „Ti, kdo si nepamatují minulost, jsou odsouzeni k tomu, si ji opakovat.“ Tento výrok lze aplikovat jak na celé národy nebo země, tak i na jednotlivé lokality, kraje, obce, města, menší komunity či jednotlivce. Historie je zdrojem poznání, proč došlo k některým událostem, co bylo jejich příčinou a jaké byly následky.

Pro vědecké zkoumání historie je v první řadě důležité shromáždit ověřené informace. Teprve na jejich základě je možno popsat historickou událost a analyzovat její příčiny a následky.

1.2 Význam ICT při této práci

Odhlédneme-li od základního využití výpočetní techniky, kdy textový editor nahrazuje psací stroj a adresář s naskenovanými dokumenty šanon s kopiemi, poskytují nám informační a komunikační technologie další podstatně širší možnosti využití. V první řadě se jedná o on-line zpřístupnění základních informací o pramenech a literatuře paměťovými institucemi. Do této kategorie lze zařadit knihovní katalogy a informace o archivních fondech a muzejních sbírkách. V druhé úrovni jsou pak základní informace doplňovány podrobnějším popisem. Zde je možno zmínit například zpřístupnění informací o obsahu knihy a on-line zpřístupnění archivních pomůcek. Třetí úroveň pak reprezentují on-line přístupné digitalizované obrazy knih, archiválií či muzejních předmětů.

Je nutno upozornit, že on-line přístupem se zde rozumí vzdálený přístup prostřednictvím internetu, který dává zájemci (badateli) možnost získat informace bez nutnosti navštívit instituci tyto informace poskytující. Takový přístup může být zcela neomezený, ale může být rovněž podmíněný a časově omezený.

1.3 Důvod volby tématu práce

Další oblastí pro uplatnění informačních a komunikačních technologií je agregace dat z mnohdy rozdílných informačních systémů jednotlivých paměťových institucí.

Existuje možnost relativně snadné agregace dat z takových rozdílných systémů? Mohou informační a komunikační technologie nabídnout historiku více možností než jen textový editor a úložiště? Snaha najít odpověď na tyto otázky byla důvodem pro volbu tohoto tématu práce.

2 Cíl práce

Cílem práce je vytvoření objektového modelu historických dat a ukázka jeho praktického použití. Součástí řešení je nalezení společných popisných polí používaných paměťovými institucemi k popisu objektů. Ukázka praktického využití bude realizována prostřednictvím funkčního vzorku systému, který historiku umožní evidovat a zpracovávat informace potřebné k jeho práci. Navržený systém bude zpracován za použití open-source produktů a bude dostupný dálkovým přístupem.

3 Metodika

Při návrhu systému bylo nejprve nutno identifikovat cílovou skupinu uživatelů a definovat její potřeby. Na základě potřeb cílové skupiny pak byly identifikovány objekty, které bylo nutno v systému zachytit. S ohledem na skutečnost, že prvky popisu jednotlivých objektů se liší, vyvstala potřeba nalézt jejich průnik a sestavit jednotný protokol. Zároveň však bylo nutné zajistit, aby v systému zůstaly zachyceny prvky popisu, které nejsou součástí uvedeného průniku.

Identifikace objektů, jejich popisných polí a sestavení jednotného protokolu se následně staly podkladem pro definici objektového modelu. Při definici objektového modelu bylo využito aplikace Daskalos.

Po definici objektového modelu bylo možné provést strukturovanou analýzu a strukturovaný návrh funkčního vzorku jeho realizaci za pomoci aplikace MetaCaseEdit. Vzhledem ke skutečnosti, kdy je jedním z parametrů funkčního vzorku dálkový přístup a s ohledem na fakt, že nebyl nalezen vhodný webhosting, byl funkční vzorek realizován za pomoci relační databáze MySQL a skriptovacího jazyka PHP.

I. Literární rešerše

4 Literární rešerše

4.1 Objektové modelování

Naprostá většina dostupné literatury zabývající se objektovým modelováním se zabývá spíše jeho praktickým využitím při návrhu informačních systémů. Do této kategorie můžeme zařadit i knihu UML2 a unifikovaný proces vývoje aplikací¹. Jednou z mála publikací věnující se čistě objektovému modelování je práce V. Merunky Objektové modelování².

4.2 Popisné standardy

Jedním z cílů práce je nalezení společného protokolu pro popis objektů. Zároveň je nutno konstatovat, že hlavními zdroji historických dat jsou historické prameny a literatura.

Největšího pokroku při standardizaci popisu objektů dosahují v současnosti knihovny, které využívají standardů rodiny MARC, zejména standardu MARC21³.

Popis historických pramenů v archivech dosud úrovně knihoven nedosahuje, byť i zde je patrná tendence k mezinárodní standardizaci. Jako nejvýznamnější standard v této oblasti je třeba uvést EAD⁴. Tento standard se však dosud neprosadil v českém prostředí, a proto je potřeba zohlednit česká specifika popisu objektů v archivech. Při definici protokolu pramenů byl proto zohledněn Standard pro ukládání a zasílání archivních pomůcek druhu inventář a dílčí inventář v digitální podobě⁵.

Ačkoli objekty uložené v archivech a knihovnách budou tvořit naprostou většinu zdrojů historických dat, nelze pominout zdroje mimo tyto instituce. Pokud charakter takových zdrojů odpovídá objektů popisovaným archivními nebo knihovnickými standardy je možno tyto použít. V opačném případě je třeba vytvořit vlastní popisné schéma, což však není dle mého názoru příliš vhodné a preferuji spíše využít standardu Dublin Core⁶.

4.2.1 MARC 21

Formát MARC21 patří do skupiny formátů MARC (MACHINE-Readable Cataloging). Jak již název napovídá, jedná se o strojově čitelný záznam obsahu, který je vytvářen na základě jiných standardů, například AACR2. Jednotlivá popisná pole jsou zaznamenána v textové podobě za pomoci tagů. Katalogy některých knihoven umožňují zobrazení informací o knize v podobě tagovaného záznamu v MARC 21, což vytváří možnost získávat snadno informace do systému.

4.2.2 EAD

Formát EAD (Encoded Archival Description) je obdobně jak MARC 21 technickým formátem sloužícím pro záznam obsahu. V tomto případě se jedná o archiválie a obsah popisných polí je definován standardem ISAD. Na rozdíl od standardu MARC 21 nejsou

1 Arlow, Jim. UML 2 a unifikovaný proces vývoje aplikací : objektově orientovaná analýza a návrh prakticky. 2., aktualizované a doplněné vydání. Brno : Computer Press, 2007. 567 s. ISBN 978-80-251-1503-9

2 Merunka, Vojtěch. Objektové modelování. 1. vydání. Praha: Alfa 2008, 197 s. ISBN 978-80-87197-04-2

3 <http://www.loc.gov/marc/bibliographic/>

4 <http://www.loc.gov/ead/>

5 <http://www.mvcr.cz/clanek/archivni-standardy.aspx>

6 <http://dublincore.org/>

informace zaznamenány prostřednictvím vlastní definice formátu, ale je použito jazyka XML.

4.2.3 Standard pro ukládání a zasílání archivních pomůcek druhu inventář a dílčí inventář v digitální podobě

Standard pro ukládání a zasílání archivních pomůcek druhu inventář a dílčí inventář v digitální podobě je prvním pokusem o univerzální strojově čitelný záznam popisu archiválií v českých archivech. Standard využívá pro záznam informací jazyka XML obdobně jako standard EAD, s nímž ovšem není kompatibilní. Zároveň je třeba konstatovat, že nezahrnuje všechny možné úrovně popisu archiválií.

4.2.4 Dublin Core

Dublin Core je standardem vyvinutým původně pro popis elektronických zdrojů. Ve své základní podobě zahrnuje 15 nepovinných prvků. Tyto prvky mohou být následně rozšířeny za pomoci kvalifikátoru (kvalifikovaný Dublin Core). Pro svou jednoduchost, rozšiřitelnost a modifikovatelnost je Dublin Core využíván i pro popis jiných než elektronických zdrojů. V rámci této práce byl využit při definici popisu objektů mimo archivy a knihovny a při definici jednotného protokolu.

II. Projekt

5 Identifikace cílové skupiny a jejích potřeb

5.1 Cílová skupina

Historii lze studovat na mnoha úrovních. Může jednat o rozsáhlé a podrobné zkoumání historických událostí, jejich zařazení do kontextu dějin států, kontinentů či světa. Může však jít i o užší zkoumání historie konkrétního místa či oboru nebo života určité osoby. Speciálním typem historického bádání se pak zabývá genealogie.

Studium historie je také možno rozdělit podle toho, zda se jedná o bádání profesionálního historika či laického badatel.

Pro účely této práce však není rozdělení historiků na profesionály a laiky podstatné. Potřeby jsou v obou případech obdobné.

Cílovou skupinou lze definovat jako historiky zabývající se konkrétním tématem.

5.2 Potřeby cílové skupiny

Výchozím předpokladem historikova zkoumání je téma. Ke zvolenému tématu pak historik shromažďuje informace, které následně analyzuje provádí syntézu. Téma je obecný pojem, který může zahrnovat studium konkrétní události (vznik Československé republiky), časové období (napoleonské války), historii konkrétního místa (dějiny města Prahy), studium života osobnosti (život a dílo Františka Palackého) apod. Informace shromažďované ke konkrétnímu tématu mohou být získávány z literatury, pramenů a vlastního výzkumu.

Navrhovaný model proto musí umět zaznamenat informace o

- a) tématu,
- b) souvisejících pramenech (archivní prameny, orální historie apod.),
- c) související literatuře (knihy, články, monografie),
- d) zeměpisných místech,
- e) osobách (osobnosti, autoři literatury),
- f) institucích.

Důležitá je zde provázanost jednotlivých informací. V případě pramenů a literatury musí být možné zaznamenat, které instituce či osoby je vytvořili či spravují. U zeměpisných míst je nutné zachytit vazbu na literaturu a témata, která se danému místu věnují. Obdobně je tomu i u osobností.

6 Identifikace objektů, které budou v systému zachyceny a definice tříd

Identifikované objekty byly při tvorbě objektového modelu použity jako podklady při tvorbě tříd. V závěru kapitoly byl vytvořen výchozí protokol a na jeho základě byly zpětně upraveny názvy atributů a metody. Zároveň byly ty z atributů, které nejsou bezpodmínečně potřebné zrušeny a změněny na objekty *další*.

6.1 Téma

Téma, tedy definice předmětu zkoumání, je výchozím objektem modelu.

V určitých případech může být vhodné téma rozdělit na několik dílčích témat. Při zkoumání historické události můžeme například nejprve samostatně zkoumat příčiny, vlastní průběh události a následky a teprve na základě těchto dílčích výzkumů zpracovat téma v plném rozsahu. Obdobně je možno postupovat i u jiných témat. Nastane-li situace, kdy je třeba téma rozdělit na dílčí témata, je nutno zachytit vazbu dílčích témat na téma hlavní.

K tématu, případně jeho dílčím tématům, je třeba následně shromáždit informace o souvisejících tématech, osobách, místech, pramenech a literatuře. Pro tento účel je každému objektu třídy *téma* definována samostatná kolekce *podklady*. Osoby, místa, prameny a literatura jsou dostatečně konkrétní a tvoří proto objekty samostatných tříd. Související témata jsou odlišná od dílčího tématu. Nejsou předmětem aktuálního historikova zkoumání, ale do určité míry s ním souvisejí. Například při studiu života a díla Kryštofa Haranta z Polžic a Bezdruzic je souvisejícím tématem poprava na Staroměstském náměstí v Praze dne 21. června 1621.

Každé zkoumání je zpravidla zakončeno nějakým výstupem, respektive výstupy. Tím může být článek, kniha či monografie (objekty *literatura*) nebo výstava, přednáška či seminář (objekt *další*).

V systému, který využívá více uživatelů je žádoucí zachytit rovněž informaci o stavu zpracování tématu, neboť téma může být zařazeno mezi zdroje informací jiného tématu. Obdobná situace nastává v případě, kdy se historik věnuje několika tématům.

Stav lze zachytit a) proměnnou jejíž hodnotu definuje uživatel nebo b) prostřednictvím metody. Řešení prostřednictvím proměnné umožňuje přesnější klasifikaci stavu zkoumání, byť z praktického hlediska bude výhodnější preferovat výběr z číselníku před volným popisem. Řešení za pomoci metody je méně spolehlivé, neboť vyžaduje definici podmínky na jejímž základě bude stav určen. Lze například definovat, že téma, jenž má výstup je tématem zpracovaným. Nicméně je zřejmé, že v rámci zpracování tématu mohou vzniknout výstupy, aniž by téma bylo uzavřeno. Například je možné, že v průběhu zkoumání proběhne přednáška, bude vydán článek, ale zpracování tématu bude uzavřeno až vydáním shrnující publikace.

Na základě výše uvedeného byly identifikovány následující atributy a metody třídy *Téma*:

název atributu	popis
název	Stručný název tématu
popis	Podrobný popis tématu
datum od	Dolní časové vymezení tématu
datum do	Horní časové vymezení tématu
nadřazený objekt	Nadřazené téma. Objekt třídy <i>Téma</i> .
podklady	Zdroje informací k tématu, Množina objektů obsahujících informace k tématu; zdroje informací
autoři	Osoby podílející se na zpracování tématu. Množina objektů třídy <i>Osoba</i>
stav	Stav zpracování tématu. Možnými stavy jsou: plánované, rozpracované, uzavřené.
další údaje	Množina objektů třídy <i>Další údaje</i>

Tab 1 Seznam atributů a metod třídy *téma*

6.2 Osoba

Objekt zachycuje údaje o fyzické osobě. Může se jednat například o autory pracující na určitém tématu, o autory literatury nebo o osobnosti. Nelze vyloučit možnost, že jedna a tatáž osoba bude v historických pramenech vystupovat pod různými jmény. Může jít o zkomolení jména, pseudonym či přezdívku. V systému se tak mohou objevit dva objekty týkající se jedné osoby. Z tohoto důvodu byl mezi atributy zařazen atribut *nadřazený*, který umožní navázat různé varianty jmen na jeden hlavní objekt.

Na základě výše uvedeného byly definovány následující atributy, metody a kolekce třídy *osoby*:

název atributu	popis
jméno	Jméno osoby
příjmení	Příjmení osoby
tituly	Tituly osoby
popis	Podrobnější informace k osobě
datum od	Datum narození
místo do	Místo narození. Objekt třídy <i>Místo</i> .
datum do	Datum úmrtí.
místo do	Místo úmrtí. Objekt třídy <i>Místo</i> .
nadřazený objekt	Nadřazená osoba. Objekt třídy <i>Osoba</i> .
další_údaje	Kolekce objektů třídy <i>další</i> obsahující <i>další údaje</i> .
název	Metoda, která vrátí příjmení, jméno a tituly jako jednu položku.

Tab 2 Seznam atributů a metod třídy *osoba*

6.3 Instituce

Objekt zachycuje údaje o institucích. Může se jednat o historické instituce, archivy, knihovny apod. Obdobně jako v případě fyzických osob nelze ani zde vyloučit možnost, že jedna a tatáž instituce nebude vystupovat pod různými názvy. Může jít o zkomolení jména, pseudonym či přezdívku. V systému se tak mohou objevit dva objekty týkající se jedné osoby. Z tohoto důvodu byl mezi atributy zařazen atribut *nadřazený*, odkazující na objekt obsahující aktuální název instituce nebo na její poslední platný název.

Na základě výše uvedeného byly definovány následující atributy, metody a kolekce třídy *Instituce*:

název atributu	popis
název	Název instituce
popis	Podrobný popis
datum od	Datum založení
místo od	Původní sídlo
datum do	Datum zániku
místo do	Poslední sídlo
nadřazený objekt	Nadřazená instituce. Objekt třídy <i>Instituce</i> .
další údaje	Kolekce objektů třídy další obsahující další údaje.

Tab 3 Seznam atributů a metod třídy instituce

6.4 Místo

Objekt zachycuje údaje o zeměpisné lokalitě, například o státu, městě či obci. Obdobně jako v případě fyzických osob nelze ani zde vyloučit možnost, že jedno místo nebude vystupovat pod různými názvy. V rámci České republiky jde o velmi pravděpodobnou možnost a místa zde mohou mít české, německé či polské označení. Pominout nelze ani změny dané politickými poměry. Z tohoto důvodu byl mezi atributy zařazen atribut *nadřazený*, odkazující na objekt reprezentující aktuální nebo poslední platný název.

Na základě výše uvedeného byly identifikovány následující atributy a metody třídy *Místo*:

název atributu	popis
název	Název místa
popis	Podrobný popis
datum od	Odkdy se název používal
datum do	Dokdy se název používal
nadřazený objekt	Odkaz na aktuální název místa. Objekt třídy <i>Místo</i> .
aktuální název	Metoda, která z objektu nadřazený získá aktuální název.
další údaje	Kolekce objektů třídy další obsahující další údaje.

Tab 4 Seznam atributů objektu místo

6.5 Pramen

Tento objekt zachycuje informace o původních zdrojích informací, mezi které řadíme archiválie, vzpomínky pamětníků, kroniky. Variabilita pramenů je velmi vysoká. Při definici objektu *pramen* a jeho atributů a metod jsem jako výchozí použil strukturu pramenů v archivech, neboť je prakticky nejsložitější a je pravděpodobné, že objekt definovaný na jejím základě bude použitelný i pro ostatní prameny.

Základní jednotkou rozdělení dokumentů v archivu je archivní soubor (fond nebo sbírka). Fond je vytvářen na základě provenienčního principu, což znamená, že ho tvoří dokumenty vzešlé z činnosti jednoho původce (úřadu, osoby). Sbírky jsou naopak tvořeny tematicky na základě pertinenciálního principu. Každý archiv spravuje svou množinu fondů a sbírek. Vnitřně se fondy nebo sbírky mohou dělit na dílčí části, přičemž hloubka úrovně hierarchického popisu není omezena. V praxi se většinou jedná maximálně o tři úrovně popisu. Specifikem archivního popisu je skutečnost, že jsou zpravidla popisovány skupiny dokumentů, nikoli jednotliviny. K popisu jednotlivin se přistupuje v případě významných archiválií či mimořádně významných nebo naopak velmi malých fondů. Mezi archiválie popisované na úroveň jednotliviny patří listiny, mapy, plakáty, fotografie či filmy. Hromadně je pak popisován zejména materiál úřední provenience. V praxi se může historik setkat se situací, kdy pro svou práci potřebuje konkrétní dokument, který však může identifikovat jen jako součást skupiny dokumentů.

Historický pramen se zpravidla vyskytuje v jednom exempláři. Toto nemusí platit například u novodobých úředních dokumentů, které se mohou vyskytovat v multiplikátech. I zde však mohou být původně stejné multiplikáty opatřeny kancelářskými poznámkami, a proto je můžeme chápat i jako rozdílné dokumenty.

Na základě výše uvedeného je třeba konstatovat, že se u pramenů můžeme setkat s následujícími vlastnostmi, s nimiž se návrh modelu musí vyrovnat:

1. Fond má hierarchickou strukturu,
2. pramenem může být celý fond,
3. pramenem může být část fondu,
4. pramenem může být jeden dokument,
5. multiplikáty dokumentu se mohou vyskytovat v různých fondech a v různých archivech.

Jedním z řešení proto bylo definovat samostatné objekty pro fond, část fondu a jednotlivý dokument a tyto posléze zařadit do hierarchické struktury. Druhou možností bylo nalezení shodných prvků popisu a definovat pro fond, část fondu a jednotlivý dokument jeden objekt, respektive třídu objektů.

V archivní praxi se pro popis archivního fondu používají tato pole evidenčního listu NAD⁷:

- a) číslo archivu nebo kulturně vědecké instituce z číselníku archivů, který archivní soubor vede ve své základní evidenci; neuvádí se u archiválií uvedených v písmenu b),
- b) číslo z číselníku archivů, pod kterým jsou evidovány archiválie uložené mimo archivy a kulturně vědecké instituce v příslušném archivu,
- c) údaje o vlastnících nebo držitelích archiválií podle § 18 odst. 3 zákona, jde-li o archiválie uložené mimo archivy a kulturně vědecké instituce,
- d) číslo evidenčního listu, které je jedinečné a neměnné v příslušném archivu nebo kulturně vědecké instituci,
- e) název archivního souboru,
- f) evidenční status archivního souboru, vyjadřující vztah archivu nebo kulturně vědecké instituce, která vede archivní soubor v základní evidenci, k evidovaným archiváliím,
- g) časový rozsah archiválií archivního souboru, a to způsobem uvedeným v § 4 odst. 1 písm. i),
- h) datum, kdy byl evidenční list vyplněn a podepsán, popřípadě datum poslední změny na evidenčním listu,
- i) kategorii archiválií v archivním souboru,
- j) údaje o přístupnosti archivního souboru pro nahlížení podle § 34, 37 a § 38 odst. 1 až 3 zákona,
- k) metráž archivního souboru, a to způsobem uvedeným v § 4 odst. 1 písm. j),
- l) stav zachování archivního souboru v rozsahu úplnost, fyzický stav, u poškozených archivních souborů i charakter poškození,
- m) číslo skupiny tematické evidence a tematický popis archivního souboru, kde skupiny tematické evidence jsou obecné, významově nadřazené pojmy shrnující společné znaky určitého typu archivních souborů,
- n) místo vzniku a místo uložení archivního souboru,
- o) evidenční jednotky archivního souboru podle § 4 odst. 1 písm. k),
- p) archivní pomůcky archivního souboru podle § 7 odst. 2 písm. b), f), h) a i),
- q) údaje o tom, zda archivní soubor obsahuje archivní kulturní památku nebo národní kulturní památku, podle § 10 odst. 4,
- r) označení původce archivního souboru v původním jazyce, starší nebo jiné názvy původce s dobou, po kterou byl název užíván, a další údaje podle § 18 odst. 3 zákona,
- s) přírůstky a úbytky archivního souboru, kde se uvádí

⁷ Vyhláška č. 645/2004 Sb. , kterou se provádějí některá ustanovení zákona o archivnictví a spisové službě a o změně některých zákonů, ve znění vyhlášky č. 192/2009 Sb., čl. I, §1 Způsob vedení základní, druhotné a ústřední evidence Národního archivního dědictví

1. všechny vnější změny, v jejichž důsledku došlo ke změnám v popisovaném archivním souboru, a to v rozsahu: číslo vnější změny, datum vnější změny a číslo jednacích dokumentů o vnější změně,
 2. všechny vnitřní změny, a to v rozsahu: popis vnitřní změny a datum vnitřní změny,
 3. v případě evidence podle § 3 odst. 4 se uvádí základní popis přírůstku nebo úbytku, a to v rozsahu: charakteristika typu vnitřní změny nebo důvody vnější změny, metráž přírůstku nebo úbytku podle § 4 odst. 1 písm. j) a evidenční jednotky přírůstku nebo úbytku podle § 4 odst. 1 písm. k),
 4. údaj o převzetí archivního souboru do archivu nebo kulturně vědecké instituce v případě, nejsou-li předchozí údaje známé,
- t) údaje o výběru archiválií ve skartačním řízení nebo mimo skartační řízení,
- u) údaje o archiváliích náležejících do archivního souboru, uložených v jiných archivech nebo jiných kulturně vědeckých institucích nebo uložené mimo archivy a kulturně vědecké instituce,
- v) údaje o literatuře o archivním souboru pojednávající o obsahu a dějinách archivního souboru a edice z archivního souboru,
- w) informace o opatřeních podle § 27 a 32 zákona, jde-li o archivní soubory, které nejsou uloženy ve veřejném archivu,
- x) údaje podle § 18 odst. 3 zákona, datum uložení archiválií do archivu a údaje o případném omezení přístupnosti archivního souboru vyplývající ze smlouvy o úschově, jde-li o archivní soubory uložené v archivu na základě smlouvy o úschově,
- y) údaje podle § 18 odst. 3 zákona a datum uzavření smlouvy, jde-li o archivní soubory uložené na základě smlouvy mimo archiv nebo kulturně vědeckou instituci, která archiválie vede v základní evidenci,
- z) datum předání a místo uložení předaného archivního souboru a další údaje podle § 18 odst. 3 zákona, jde-li o archivní soubory fyzicky předané vlastníkům,
- aa) údaje o počtu kopií archiválií, které svými vlastnostmi a způsobem zpracování zaručují nejvyšší trvanlivost pro dlouhodobé uložení (dále jen "bezpečnostní kopie") a počtu kopií určených k uživatelské práci,
- bb) jméno zpracovatele evidenčního listu,
- cc) jméno operátora záznamu evidenčního listu, pokud je veden v digitální podobě,
- dd) značku archivního souboru, pokud je pro archivní soubor zavedena.

Z uvedených údajů mají hlediska historika zásadní význam položky:

- a) – c) pro identifikaci instituce či osoby, která archiválii spravuje,
- d) a e) pro identifikaci archivního fondu nebo sbírky,
- g) časový rozsah archiválií archivního souboru,
- p) archivní pomůcky, ze které je patrné, zda je fond zpracován a v jakém rozsahu,
- u) obsahující údaje o archiváliích náležejících do archivního fondu nebo sbírky.

Vnitřní členění fondu závisí na zpracovanosti fondu. Je-li fond částečně nebo zcela zpracován, existuje k němu alespoň jedna pomůcka. Ta se zpracovává k určité logické a uzavřené části fondu a lze ji tedy použít jako součást struktury fondu. O archivní pomůcce se zaznamenávají mimo jiné tyto informace:

- a) evidenční číslo pomůcky,
- b) fond k němuž byla zpracována,
- c) název pomůcky,
- d) časový rozsah zpracovaného materiálu.

V rámci archivní pomůcky je pak s různou mírou podrobnosti zachycena další struktura fondu. Na úrovni popisu logického celku, kterému jsou podřazeny další záznamy se jedná zpravidla pouze o název, eventuálně identifikátor. Na konečné úrovni popisu je pak zachycen druh identifikátoru (inventurní číslo, signatura), hodnota identifikátoru, název, časový rozsah, druh evidenční jednotky (kniha, karton, kartotéka ad.) a identifikátor evidenční jednotky.

V případě, že je fond nezpracovaný, má jeho struktura zpravidla pouze dvě úrovně: fond a popis na úrovni skupiny dokumentů. Popis na úrovni dokumentů v takovém případě zahrnuje druh identifikátoru (signatura, karton, kniha) a hodnotu identifikátoru.

Zohledníme-li fakt, že převážná část položek popisu archivního fondu není v případě systému pro správu historických dat zásadní, můžeme je definovat jako objekty *další*, čímž se nám velmi sníží počet atributů případného objektu a otevře se nám možnost definovat pro popis fondu, jeho části a dokumentu jeden objekt. Pokud bychom tak neučinili, museli bychom minimálně pro fond definovat samostatný objekt. Tím bychom sice měli všechny údaje součástí objektu, ale v případě zobrazení společně s objekty jiných tříd bychom informace stejně nezískali, neboť by nebyly součástí společného protokolu.

Na základě zvoleného řešení byly identifikovány následující atributy a metody třídy *Prameny*:

název atributu	popis
název	Název archivního souboru, pomůcky, dokumentu, skupiny dokumentů
popis	Případný podrobný popis
datum od	Dolní hranice časového rozsahu
datum do	Horní hranice časového rozsahu
nadřazený objekt	Nadřazený pramen. Objekt třídy <i>Pramen</i>
správce	Instituce nebo osoba, která pramen spravuje. Objekt třídy <i>Instituce</i> .
multiplikát	V případě multiplikátu je jeden z nich definován jako hlavní a ostatní se na něj odkazují prostřednictvím tohoto atributu. Objekt třídy <i>Pramen</i> .
další údaje	Kolekce objektů třídy <i>Další údaj</i> obsahující další údaje.

Tab 5 Seznam atributů objektu pramen

6.6 Literatura

Mimo studium pramenů je dalším zdrojem informací odborná literatura, která je výsledkem předchozího studia pramenů. Do této kategorie řadíme edice původních pramenů (například Berní rula, Codec diplomaticus Regni Bohemiae), odborné knihy, odborné články a monografie.

V případě definice objektu *literatura* je situace jednodušší. Jak již bylo zmíněno v kapitole literární rešerše, existuje v knihovnách vysoká míra standardizace popisu objektů. Příkladem může být srovnání záznamů získaných z katalogů Národní knihovny a Souborného katalogu⁸:

katalog Národní knihovny	Souborný katalog
LDR -----nam-a22-----a-4500	LAB -----nam 22----- a 4500
FMT BK	001 KN9912100000284908
001 nkc20061702201	005 20081126143519.0
003 CZ-PrNK	008 070205s2006---xr-----u-----cze-d
005 20070709112235.0	020 \$a80-86851-54-0 :\$c299.00 Kč
007 ta	040 ## \$a\$eAACR2\$91
008 061218s2006---xr-a---e-p---000-0-cze--	080 \$a004.652
015 a cnb001702201	080 \$a004.422.63
020 a 80-86851-54-0 (váz.) : c Kč 299,00	080 \$a681.3.062
035 a (OCoLC)137329694	080 \$a621.3
040 a ABA001 b cze	080 \$a004
072 7 a 004.4/.6 x Programování. Software 2	100 1# \$aMerunka, Vojtěch\$4aut
Konspekt 9 23	245 10 \$aDatové modelování /\$cVojtěch Merunka
072 7 a 37.016 x Učební osnovy. Vyučovací	250 \$a1. vyd.
předměty. Učebnice 2 Konspekt 9 22	260 \$aPraha : \$bAlfa,\$c2006
080 a 004.41 2 MRF	300 \$a177 s. :\$bil.
080 a 004.652 2 MRF	490 1# \$aInformatika studium
080 a 004.422.63 2 MRF	653 0# \$adatové modely\$adatové struktury\$aobjektově
080 a 004.41.045 2 MRF	orientované
080 a (075) 2 MRF	metody\$adatabáze\$aprogramování\$asoftwarové
1001 a Merunka, Vojtěch, d 1967- 7	inženýrství\$aobjektově orientované
xx0005022 4 aut	programo\$aučebnice\$amodely\$adata\$apočítače\$adatové\$a
24510 a Datové modelování / c Vojtěch	modelování\$apříklad\$atechniky\$asoftwaro\$asystémy\$aš
Merunka	koly\$avysoké\$aprojekty\$amatematika\$aazyky\$aSmalltal
250 a 1. vyd.	k\$aLambda\$aúlohy\$adotazy\$asymboly\$ainformace\$aanalýz
260 a Praha : b Alfa Publishing, c 2006	y\$apostupy
300 a 177 s. : b il. ; c 25 cm	830 #0 \$aInformatika studium
4901 a Informatika studium	910 \$a
504 a Obsahuje bibliografii	
65007 a softwarové inženýrství 2 czenas	
65007 a datové modely 2 czenas	
65007 a datové struktury 2 czenas	
65007 a objektově orientované metody 2	
czenas	
65009 a software engineering 2 eczenas	
65009 a data models 2 eczenas	
65009 a data structures 2 eczenas	
65009 a object-oriented methods 2 eczenas	
655 7 a učebnice 2 czenas	
655 9 a textbooks 2 eczenas	
830 0 a Informatika studium 0 s	
901 b 9788086851549 f 1. vyd. o 20070709	
9101 a ABA001	
9281 a Alfa Publishing	
9102 a BOA001 b 2-1186.973	
9103 a OLA001 b 1-184.789	
SYS 001702201	

Tab 6 Srovnání popisu knihy v MARC21

Přesto lze říci, že z hlediska navrhovaného systému není vhodné všechny položky popisu zahrnovat mezi atributy objektu. Situace je obdobná popisu archivního fondu. I v tomto případě bychom získali objekt s mnoha atributy, které by se však s ohledem na

8 <http://www.skat.cz>

společný protokol nezobrazovaly. I v případě zobrazení pouze objektů *literatura*, bylo obtížné přehledně zobrazit všechny atributy.

Vhodnějším řešením se proto jeví definice menšího množství atributů a označení podrobných údajů jako *další* objekty.

Na základě výše uvedeného byly identifikovány následující atributy a metody třídy *Literatura*:

název atributu	popis
název	Název knihy
datum od	Vročení tématu, dolní hranice
datum do	Vročení tématu, horní hranice
místo od	Místo vydání
nadřazený	Novější vydání knihy
autoři	Seznam autorů. Množina objektů třídy <i>Osoba</i>
správce	Instituce (knihovna) nebo osoba, která vlastní knihu. Množina objektů třídy <i>Instituce</i> .
další údaje	Kolekce objektů třídy <i>Další údaj</i> obsahující další údaje.

Tab 7 Seznam atributů objektu *literatura*

6.7 Další údaj

Ne vždy je možné při definici třídy zachytit všechny potřebné atributy. I přes pečlivý návrh se může snadno stát, že uživatel bude vyžadovat záznam informace, pro kterou nebude definován atribut. Jinou variantou je situace, kdy je počet atributů vysoký a zároveň nejsou takové atributy zásadní pro běžnou práci se systémem. Příkladem může být záznam knihy v knihovně. Standard MARC 21 obsahuje množství položek, z nichž je však v praxi užívána jen malá část. Pokud bychom chtěli umožnit uživateli zaznamenat všechna popisná pole MARC 21, bylo by třeba je definovat jako atributy, přesněji část jako atributy a část jako kolekce (opakovatelné položky). Takové řešení je krajně nepraktické, protože třída by obsahovala mnoho atributů, z velké části nevyplněných. Uvážíme-li, že takové dodatečné údaje jsou důležité svou informací, ale nejsou kritické pro fungování systému a pro vyhledávání, jeví se vhodné pro ně vytvořit samostatnou třídu a její relevantní instance pak zachytit u dalších objektů formou kolekce.

Při záznamu dalšího údaje je třeba zachytit název položky (údaje) a hodnotu. Název položky může vycházet z existujících konvencí nebo může být definován uživatelem. Z tohoto důvodu je vhodné zachytit rovněž informaci o původu názvu položky (vlastní, MARC 21, Dublin Core ad).

Vzhledem k univerzálnosti objektu nelze vyloužit požadavek na hierarchické členění dalších údajů a proto je definován atribut nadřazený.

Na základě výše uvedeného byly definovány následující atributy a metody třídy *Další údaj*:

název atributu	popis
zdroj	Identifikace pravidel, dle kterých je atribut definován (např. MARC 21).
název	Název atributu
hodnota	Hodnota atributu
nadřazený objekt	Nadřazené údaj. Objekt třídy <i>Další údaj</i> .
objekt	Metoda, která vrací seznam objektů, jehož se údaj týká.

Tab 8 Seznam atributů objektu další

6.8 Minimální protokol

Při sestavování atributů jednotlivých objektů byl kladen důraz na nalezení minimálního protokolu společného pro všechny objekty. Cílem navrhovaného systému není nahradit existující specializované evidence jednotlivých zdrojů informací. Proto byly v maximální míře potlačeny položku specifické pro konkrétní zdroje a pro jejich zaznamenání v systému byl vytvořen speciální objekt *Další údaje*. Níže uvedená tabulka uvádí minimální protokol společný všem objektům.

název atributu	popis
název	Stručný název tématu
popis	Podrobný popis tématu
datum od	Dolní časové vymezení
datum do	Horní časové vymezení

Tab 9 Minimální protokol

7 Vytvoření funkčního prototypu systému s využitím volně dostupného SW a s možností webového přístupu

Jedním z cílů práce je vytvoření funkčního vzorku systému za pomoci open source produktů. Tento funkční vzorek by měl ověřit možnost praktického využití objektového přístupu k datům. Základním parametrem systému byla dále možnost webového přístupu. Ačkoli v dnešní době existuje několik produktů umožňujících objektové zpracování dat, práce s nimi není nejsnadnější a nelze předpokládat, že by ji byl běžný uživatel (jakým historik je) ochoten absolvovat. Bylo sice možné realizovat funkční vzorek jako lokální aplikaci, ale tato možnost se jeví jako omezující. Realizace prostřednictvím webové aplikace umožní nejen ověřit výhody objektového přístupu k datům, ale i možnost realizace systému využívaného více uživateli. Zvolené řešení však naráží na již zmíněný problém nedostatečného rozšíření objektových databází. V současnosti neexistuje webhosting nabízející možnost použití objektové databáze. Funkční vzorek byl proto realizován za pomoci relační databáze MySQL a skriptovacího jazyka PHP.

Relační databáze je v tomto případě použita jen jako úložiště dat, k jejichž zpracování dochází na aplikační úrovni.

7.1 Realizace tříd

Množina tříd je definována v kapitole 7. Pro každou třídu byla vytvořena tabulka obsahující všechny instance třídy. Všechna pole tabulky jsou definována jako char. Výjimku tvoří pole popis, které je definováno jako text.

Definice atributů, tak jak jsou definovány v objektovém modelu, je zachycena v iniciačním skriptu aplikace.

7.2 Realizace kolekcí

Jak vyplývá z definice tříd, jsou některými atributy kolekce objektů. počet kolekcí v systému je proto proměnlivý. Z tohoto důvodu jsou v jedné tabulce zachyceny definice všech kolekcí. V další tabulce je pak zachycena příslušnost jednotlivých objektů ke kolekcím.

7.3 Umístění funkčního vzorku

Funkční vzorek je dostupný na adrese http://jirkab.euweb.cz/databaze_historickych_dat

8 Závěr

V rámci zpracování tématu vznikl objektový model historických dat a funkční vzorek systému pro jejich evidenci. V průběhu řešení se objevilo několik problémů. Za nejzávažnější z nich považuji diametrálně odlišný způsob zachycení údajů o pramenech, literatuře a dalších zdrojích informací. To se projevilo v definici protokolu společného pro všechny objekty, který obsahuje jen minimální množství údajů. Další komplikací (spornou otázkou?) je stále nedostatečná rozšířenost nástrojů založených na objektovém přístupu. Na základě toho bylo potřebné pro vytvoření funkčního vzorku použít tradičních nástrojů se slabou podporou objektového programování. V případě funkčního vzorku se nejedná o zásadní problém, ale pokud by se jednalo o rozsáhlejší systém s větším množstvím dat, bylo by pravděpodobně nutné zvolit jiný přístup.

Závěrem je však nutné konstatovat, že navzdory uvedeným problémům se objektový přístup k datům projevuje jako velmi přehledný a srozumitelný. Dosavadní přístup k evidenci dat založený na relačních databázích je ve světle zkušeností získaných při zpracování tématu možné považovat za příliš svazující. Lze tedy tvrdit, že při návrhu konkrétní evidence v zásadě uvažujeme objektově a teprve následně tyto úvahy překládáme do relačního jazyka. Právě v konfrontaci dosavadních zkušeností s objektovým přístupem spatřuji hlavní přínos, který jsem při zpracování tématu získal.

9 Použitá literatura

ARLOW, Jim. UML 2 a unifikovaný proces vývoje aplikací : objektově orientovaná analýza a návrh prakticky. 2., aktualizované a doplněné vydání. Brno : Computer Press, 2007. 567 s. ISBN 978-80-251-1503-9

MERUNKA, Vojtěch. Objektové modelování. 1. vydání. Praha: Alfa 2008, 197 s. ISBN 978-80-87197-04-2

MARC 21 : bibliografický formát. 1. české vydání. Praha: Národní knihovna České republiky, 2003. 2 sv. ISBN 80-7050-427-7

KOSEK Jiří, PHP – tvorba interaktivních internetových aplikací. 1. vydání. Praha : Grada, 1998. 492 s. ISBN 80-7169-373-1

PECINOVSKÝ Rudolf. Návrhové vzory. 1. vydání. Brno: Computer Press 2007. 523 s. ISBN 978-80-251-1582-4