

Sem vložte zadání Vaší práce.

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
KATEDRA SOFTWAREVÉHO INŽENÝRSTVÍ



Bakalářská práce

Encyklopedie datového skladu

Martin Čejka

Vedoucí práce: Ing. Stanislav Kuznetsov

11. května 2015

Poděkování

Děkuji svému vedoucímu Stanislavu Kuznetsovovi za odborné vedení a za pomoc. Děkuji i Lucovi de Ioannovi za sdílené informace a odborné rady v průběhu tvorby této práce. Dále vděčím svým blízkým Lence Rybákové, Jakubovi Krejčímu, Robertu Kotlářovi, Radimu Lengerovi, Vítu Královi a mým rodičům za podporu.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval(a) samostatně a že jsem uvedl(a) veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 46 odst. 6 tohoto zákona tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou, a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užit. Tyto osoby jsou oprávněny Dílo užit jakýmkoli způsobem, který nesnižuje hodnotu Díla, a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené. Každá osoba, která využije výše uvedenou licenci, se však zavazuje udělit ke každému dílu, které vznikne (byť jen zčásti) na základě Díla, úpravou Díla, spojením Díla s jiným dílem, zařazením Díla do díla souborného či zpracováním Díla (včetně překladu), licenci alespoň ve výše uvedeném rozsahu a zároveň zpřístupnit zdrojový kód takového díla alespoň srovnatelným způsobem a ve srovnatelném rozsahu, jako je zpřístupněn zdrojový kód Díla.

V Praze dne 11. května 2015

.....

České vysoké učení technické v Praze
Fakulta informačních technologií

© 2015 Martin Čejka. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí, je nezbytný souhlas autora.

Odkaz na tuto práci

ČEJKA, Martin. *Encyklopedie datového skladu*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2015.

Abstrakt

Na Českém vysokém učení technickém vzniká datový sklad, který zefektivní získávání informací a znalostí z univerzitních systémů. Pro sjednocení a zrychlení komunikace mezi vedením univerzity a týmem, který bude spravovat datový sklad, je třeba portál označovaný jako *Encyklopedie datového skladu*. Součástí mé práce je analýza a návrh tohoto portálu. Důležitou součástí návrhu je struktura podnikového a datového slovníku. Z provedeného průzkumu systémů pro řízení projektů jsem vybral ten nejvhodnější a vytvořil jsem na něm prototyp encyklopedie, který jsem nastavil a otestoval. Mé řešení je nasazeno na fakultní server a připraveno k použití.

Klíčová slova Datový sklad, encyklopedie datového skladu, Redmine, metadata, podnikový slovník

Abstract

At the Czech Technical University a data warehouse has been created, to access more effectively the information and knowledge from the university systems. In order to unite and accelerate the communication between the management of the University and the team providing the data warehouse, a portal called *Data warehouse Encyclopedia* is currently needed. The analysis and the design of the portal are included in my bachelor thesis. The structure of the business and data dictionary represents an important part of the design. Due to my research, I have chosen the most appropriate project management system, accordingly I have created the Encyclopedia prototype, having been configured and tested. The proposed solution is used on the server of the faculty and prepared to be used further.

Keywords Data warehouse, Data warehouse encyclopedia, Redmine, metadata, business dictionary

Obsah

Úvod	1
Cíl práce	1
I Teoretická část	3
1 Úvod do Business Intelligence	5
1.1 Úvod do datových skladů	6
1.2 Reporty	7
2 Metadata	9
2.1 Podniková encyklopedie	15
II Praktická část	17
3 Analýza	19
3.1 Současný stav	19
3.2 Úvod do analýzy	19
3.3 Požadavky	20
3.4 Modelování případu užití	21
4 Návrh	25
4.1 Rešerše systémů	26
4.2 Podoba slovníků	34
5 Realizace a testování	37
5.1 Slovník	37

5.2	Uchovávání soukromých informací	38
5.3	Nasazení	38
5.4	Testování	39
	Závěr	41
	Použité zdroje	43
A	Seznam použitých zkratk	45
B	Ukázka encyklopedie	47
C	Obsah přiloženého CD	51

Seznam obrázků

1.1	Schéma datového skladu	6
1.2	Report	7
2.1	Metadata v žádosti o práci	10
2.2	Dimenzionální model datového tržiště	11
2.3	Ukázka podnikové encyklopedie	16
3.1	Diagram aktivit	20
3.2	Diagram případu užití	22
4.1	Redmine	27
4.2	Wordpress	29
4.3	Trac	30
4.4	Taiga	31
4.5	Asana	32
4.6	OpenProject	33
4.7	Diagram návrhu slovníku	35
5.1	Diagram nasazení	39

Seznam tabulek

2.1	Návrh logické datové mapy část I.	12
2.2	Návrh logické datové mapy část II.	12
2.3	Návrh logické datové mapy část III.	12
3.1	Kontrola splnění funkčních požadavků	24

Úvod

Pro informované rozhodování na vrcholových pozicích ve firmách a organizacích je klíčová existence kvalitních a přehledných dat v podnikových zdrojích a jejich správná interpretace. V současné době je na Českém vysokém učení technickém dodávání informací pro vedení neorganizované a nepřehledné. Při používání aktuálně vyvíjeného datového skladu ČVUT bude třeba platforma, která zpřehlední komunikaci mezi vedením univerzity a technickou obsluhou datového skladu. Konkrétně bude tato takzvaná encyklopedie zaznamenávat požadavky na vytvoření reportů a bude centralizovaně zpřístupňovat vypracované reporty. Aby se mohla encyklopedie stát kvalitním zdrojem znalostí, bude obsahovat slovník pojmů pro správné pochopení informací a jednotnou komunikaci. V neposlední řadě bude i zdrojem informací pro technický tým datového skladu – bude obsahovat diagramy zdrojových databází, datového skladu a popisné informace. Téma jsem si zvolil, abych pomohl při tvorbě univerzitního datového skladu a představil nástroj pro vedení školy, čímž v konečném důsledku pomohu ke zkvalitnění výuky.

Cíl práce

V této práci nejprve nastíním význam a pozadí tohoto tématu a zadefinuji hlavní pojmy. Provedu analýzu a navrhnu podobu podnikového i datového slovníku. Dále provedu průzkum nabídky a vyzkouším bezplatné systémy použitelné jako encyklopedie. Nejlepší systém připravím dle návrhu, provedu pilotní nasazení na školní infrastrukturu a otestuji.

Část I
Teoretická část

Úvod do Business Intelligence

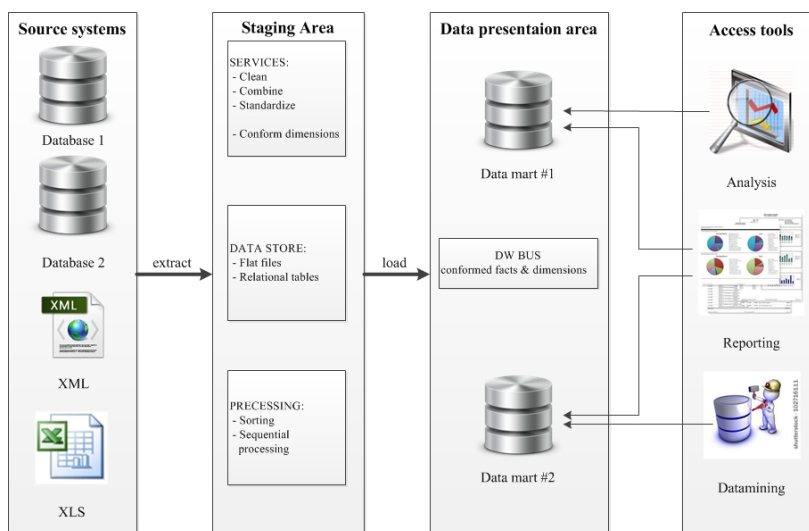
Jelikož téma datových skladů a jejich encyklopedií spadá pod obor Business Intelligence (BI), rozhodl jsem se začít definováním tohoto pojmu a jeho důležitosti:

Obecná definice BI zní podle [1] takto: Včasné dodání přesných a užitečných informací k efektivnímu rozhodování. BI zahrnuje postupy, nástroje a technologie k vyřešení kriticky důležitého úkolu - dodat informace k vedení organizace, firmy nebo např. univerzity tak, aby mohly sloužit jako podklad pro každodenní informované rozhodování a jako zpětná vazba při měření výsledků těchto rozhodnutí. Nejedná se pouze o fakta a čísla na papíře nebo na obrazovce. Tabulky, grafy a výpisy sice mohou být velice přesné a obsahovat obrovská množství dat, ale mluvit o nich v rámci BI můžeme až tehdy, dokáže-li jim čtenář (tedy vedení firmy, manažer, analytik, ředitel vládní nebo vědecké organizace..) rychle a snadno porozumět. Je zřejmé, že správné použití BI povede k lepšímu poznání trhu a zákazníků, předvídání trendů, zkoumání výsledků a tím pádem výrazně přispěje k zisku firmy nebo organizace a její konkurenceschopnosti. Proto je odvětví BI tak důležité a stoupá zájem o využívání jeho nástrojů.

Ačkoli k první zmínce o BI podle [2] došlo už v 19. století a potenciál vzniklý použitím výpočetní techniky byl spatřen už v roce 1958, k opravdovému rozvoji došlo až s nástupem výkonných počítačových systémů a specifického software koncem 90. let 20. století. Hlavním pilířem, na kterém stojí moderní BI jsou tzv. „datové sklady“, které popíši v další části.

1.1 Úvod do datových skladů

Následující informace jsem našel v [3]. Každá společnost si potřebuje ukládat data z různých zdrojů a vrstev (prodeje, zaměstnanci, dodavatelé, výroba, pobočky...) a databáze nemívají jednotný formát ani typ úložiště, obsahují mnoho nadbytečných nebo chybných dat a tím pádem se nehodí pro analýzu a získávání širokých znalostí. Pro potřeby BI je nutné načítat data z firemních databází do oblasti zvané *staging area*, tam se čistí a upravují a nakonec se nahrávají do úložiště ve sjednoceném formátu. Načtení, zpracování a uložení se nazývá *ETL¹ proces*. Upravená data se ukládají v datovém skladu, který slouží jako primární zdroj pro BI analýzy. V klasických transakčních databázích se data běžně mohou měnit a mazat, obsah datových skladů se může pouze číst a obvykle se nahrává v relativně velkých objemech najednou. Podle návrhu Ralpha Kimballa [4], jednoho z otců datových skladů, je datový sklad sjednocená množina datových tržišť (anglicky *data mart*). Každé datové tržiště obsahuje data z jedné části společnosti (např. registr zaměstnanců nebo prodeje), ale tato data mohou pocházet z více firemních zdrojů. Dle diplomové práce ing. Stanislava Kuznetsova [5] se na rozdíl od ukládání dat v normalizované podobě ve zdrojových databázích, data v datových tržištích ukládají podle dimenzionálního modelu, který je určen pro optimalizaci rozsáhlých dotazů v analytických nástrojích.

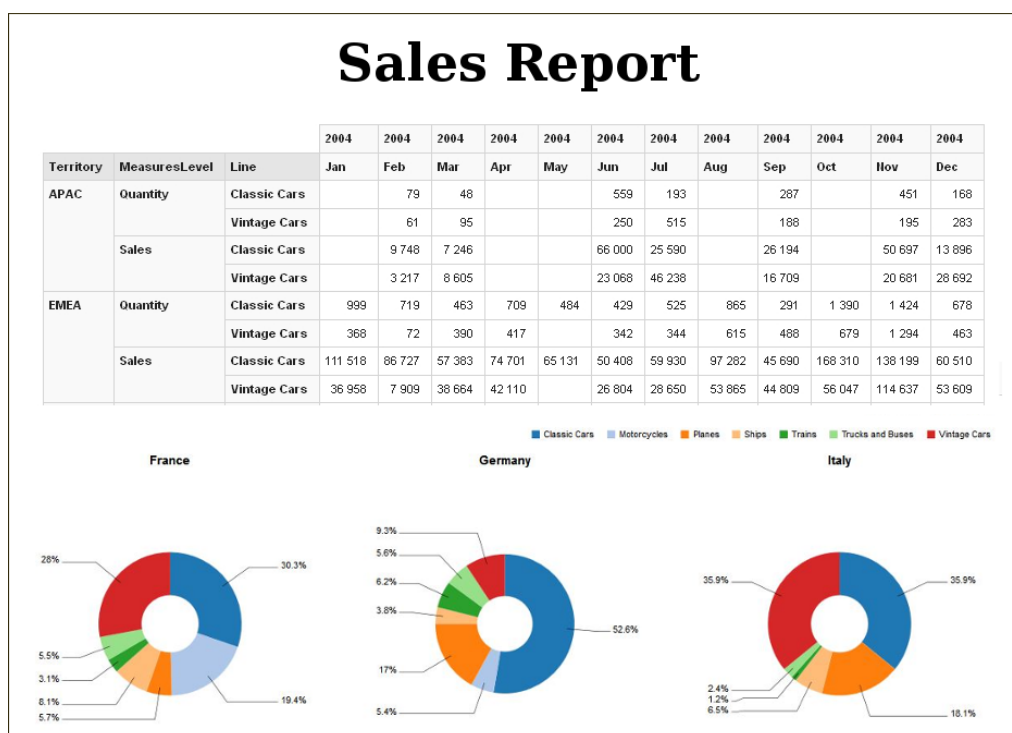


Obrázek 1.1: Schéma datového skladu dle R. Kimballa [5]

¹Extract, Transform, Load

1.2 Reporty

Dle [6] jsou data vytvořena, uložena, měněna, nahrána z mnoha různých systémů do datového skladu a jejich posledním krokem v business intelligence je jejich zobrazení koncovým uživatelům v přehledné, účelné a srozumitelné podobě. Slouží jako zdroj informací pro vedení a manažery, kteří nemají čas ani znalosti na práci s datovým skladem a musí se rychle a správně rozhodnout bez znalosti celého procesu. Tato vyobrazení se nazývají *reporty* a mají mnoho podob - tabulky, grafy i textové dokumenty. Reporty mohou být jak statické, tak interaktivní. Součástí nabídky softwaru datových skladů často bývají i nástroje ke tvorbě reportů, které komunikují přímo s datovým skladem. Získat správná data z mnoha firemních zdrojů, uložit je a vytvořit report je složitý proces s mnoha úskalími. Nicméně, k poslednímu kroku, k opravdovému získání znalostí, je třeba výsledky správně pochopit. Z tohoto důvodu nyní popíšeme metadata.



Obrázek 1.2: Ukázka reportu v aplikaci Saiku

Metadata

Jedna z mnoha definic zní takto: „Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information.” [7] Jinými slovy, metadata slouží k popisu dat a k jejich hlubšímu pochopení a interpretaci, poskytují kontext a souvislosti. Zároveň v nich může být zaznamenán zdroj, struktura nebo účel dat.[6]

„informace = data + metadata” [6]

Metadata jsou sám o sobě velice obecný pojem a mají mnoho podob a využití. Obrázek 2.1 zachycuje nevyplněný formulář - žádost o práci. Metadata, tedy popisky jednotlivých kolonek, jsou zvýrazněna. Je zřejmé, že bez těchto metadat by vyplněný dokument mohl obsahovat data, ale neměl by žádnou strukturu a šel by obtížně vyplnit nebo přečíst. Teprve metadata řeknou uživateli, co znamená obsah pole, jaký má povolený formát apod. Pro potřeby této práce představím dvě kategorie metadat:

2.0.1 Technická metadata

Technická metadata dle W. Inmona [8] existují v různých formách už od vzniku prvních počítačů. Technická metadata slouží technikům, programátorům a databázovým administrátorům a ti je používají k návrhu, vývoji a údržbě systémů a k dalším účelům. Na rozdíl od business metadat v další sekci se často vyskytují ve tvaru, který je pro nezasvěcené nepochopitelný. Technická metadata se často používají v komunikaci mezi počítačem a technickým uživatelem. (název sloupce v SQL² dotazu, název proměnné ve zdrojovém kódu...) Mezi nejnázornější příklady technických metadat patří:

²Structured Query Language

2. METADATA

**Acme Brick Works
Job Application Form**

Name		
Address		
Phone		Email
Education		
Degree?	University	GPA
Major		
Years attended		
Previous employer		date
Position		
Responsibility		
Salary		Bonus
Previous employer		date
Position		
Responsibility		
Salary		Bonus
Desired position		
Start date		
Referred by		
Interviewed by		

Obrázek 2.1: Metadata v žádosti o práci [8]

- název databázové tabulky
- rozvržení databázové tabulky
- název databázového sloupce
- formát databázové buňky
- omezení pro databázovou buňku

L. Beránek ve své diplomové práci [6] uvádí různé zdroje technických metadat. Ta výše zmíněná pochází z prostředí správy databázových systémů (DBMS³), tedy z databázových tabulek, které spravují velká množství podnikových dat. Systémy pro správu databází obsahují mnoho metadat popisujících tabulky, jejich vztahy i data. Dalším zdrojem mohou být UML⁴

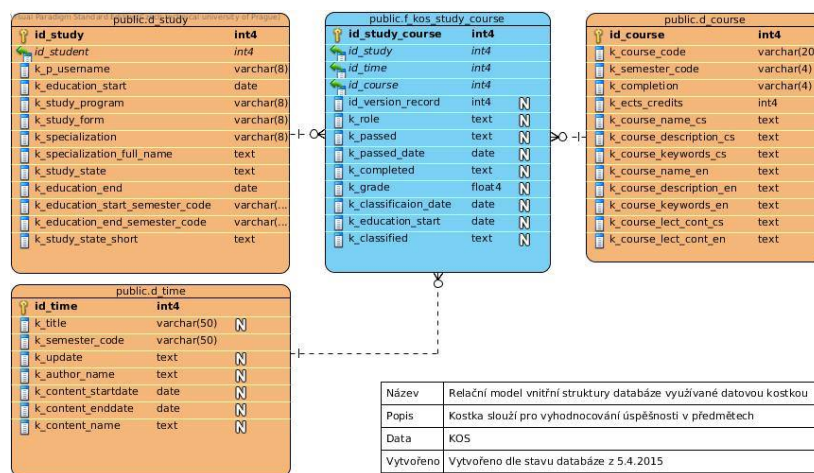
³Database Management System

⁴Unified Modeling Language

diagramy, což je skupina grafických notací, které napomáhají při návrhu systémů založených na objektově orientovaném programování. Metadata znázorňují samotné diagramy (class diagram, use case diagram, diagram komponent..) i další přidané údaje v nástrojích pro jejich tvorbu. Posledním příkladem mohou být ERD⁵, které se často používají pro vyjádření vztahů mezi objekty dané doménové oblasti při tvorbě fyzického a konceptuálního datového modelu. přestože se ERD mohou relativně snadno interpretovat do lidské řeči jako business metadata, původem a způsobem vytvoření se jedná o metadata technická.

2.0.1.1 Technická metadata v datovém skladu

Přímo v datovém skladu se můžeme často setkat s technickými metadaty, která popisují uložená data, jako například dimenzionální model datového tržiště na obrázku 2.2.



Obrázek 2.2: Dimenzionální model datového tržiště

Jak už jsem naznačil dříve v této práci, data jsou nahrávána ze zdrojových databází pro další zpracování. Nahrávána mohou být automaticky nebo manuálně, a to ve formě *exportů*. Export je tabulkový (např. excelový) soubor a jedná se o výsledek dotazu nad zdrojovou databází. V souboru jsou uvedena pouze data a technický název sloupce a Pro ETL tým je důležité vědět, co hodnoty ve sloupci představují, jaký mají datový typ a povolené hodnoty. Dále je vhodné namapovat na tento zdrojový atribut jeho ekvivalent v datovém skladu. Jakub Krejčí navrhl tabulky 2.1, 2.2 a 2.3 pro zachycení metadat ve zdrojové databázi, cílové databázi, tedy jak je vidět

⁵Entity Relation Diagram

na obrázku 2.2 a pro popsání transformací, kterými data prošla. Dále jsou v tabulce business metadata, kterým se budu věnovat v další části.

Zdrojový systém			
Databáze	Tabulka	Sloupec	Datový typ

Tabulka 2.1: Návrh části zdrojový systém v logické mapě. Zdroj: [9]

Cílový systém				
Tabulka	Sloupec	Datový typ	Typ tabulky	Typ historizace

Tabulka 2.2: Návrh části cílový systém v logické mapě. Zdroj: [9]

Business popis	
Obchodní název	Obchodní definice

Tabulka 2.3: Návrh části business popis v logické mapě. Zdroj: [9]

Tyto popisné informace je praktické ukládat někam, kde budou dostupné a přehledné. O jejich ukládání v datovém slovníku budu podrobněji psát v dalších částech této práce.

2.0.2 Business metadata

Business (podniková) metadata jsou dle [8] určená pro zaměstnance firmy, zejména manažery a vedení a používají se neustále při fungování podniku. Jsou psána v jazyce srozumitelném pro zaměstnance a poskytují datům kontext a význam. Podniková metadata vznikají při samotném počátku podnikání a často existují jen v myslích zaměstnanců, výrazné množství je zachyceno a předáváno ve formě dokumentů (nabídek, smluv..), obrázků, diagramů, tabulek, e-mailů, informačních systémů, databází nebo technologických nástrojů. Dle Lucy de Ioanny [10] se může jednat například o:

- podnikové označení údajů
- podniková definice údajů
- podniková aplikace zodpovědná za vytvoření nebo správu údaje
- platné hodnoty dle podnikových pravidel
- pravidla pro používání tohoto údaje
- obchodní pravidla, která souvisí s pojmem
- vzorce, podle kterých tento údaj vznikl

Jak jsem uvedl, jsou určena pro netechnické uživatele, ale své využití mají i pro analytiky a členy ETL týmu, kteří sklad spravují a tvoří reporty. Jak píše ve své bakalářské práci Jakub Krejčí [9], „Tento typ metadat je v rámci datového skladu nejdůležitější, protože přibližuje informace obsažené v datovém skladu běžnému uživateli. Bez business metadat by musel pracovat s daty z datového skladu pouze člen týmu vývojářů datového skladu, který rozumí architektuře zdrojových systémů, datového skladu a obchodnímu procesu společnosti.“

2.0.3 Metadata management

Následující kapitola je převzata z [8]. Pokud se zamyslíme nad širším použitím metadat, narazíme na problém, který demonstruji na příkladu. Mějme firemní report o provozu firmy za určitý časový úsek, například měsíc. Jak už jsem naznačil v dřívější sekci, bez metadat by se v reportu nacházela pouze čísla bez jakéhokoli označení a absolutně by jim nebylo rozumět. Nicméně, i když bude výpis popsán a budeme se moct dočíst, jaké tržby společnost měla, budou všichni zaměstnanci pracovat se stejnou definicí tržby? Kde ta definice bude zaznamenána? Patří či nepatří do tržeb následující položky?

- daně
- dopravné
- slevy
- poničené zboží
- provize
- vrácené zboží

Jak je vidět, nestačí mít validní data v databázích či datovém skladu, ale je nutné jim správně porozumět. Metadata se nacházejí všude a napříč odděleními mohou existovat různé definice téhož výrazu - mohou být zastaralé, neúplné nebo nesprávné. Jednotnost slovníků všech oddělení přispívá firemní efektivitě a hraje životně důležitou roli v použitelnosti datového skladu. Podobně jako se firemní data čistí, transformují a centralizují v datovém skladu, je velice vhodné sjednotit podniková metadata do jednoho úložiště. Navíc toto úložiště výrazně urychlí jejich hledání. Vytvořit takový repozitář může být složitá a dlouhodobá práce. Jeho správce se nazývá **Data steward**. Je to osoba ze strany podniku, tedy ne nutně technicky zaměřená, která zná firemní vztahy a procesy a je zodpovědná za správu a kvalitu

metadat. Je to pozice velice důležitá a je nutné vykonávat jí pečlivě, protože na kvalitě a přehlednosti metadat budou záviset mnozí zaměstnanci a manažeři.

Každý nástroj nebo součást technologické architektury má svá metadata, nazývána lokální metadata. Ta jsou obvykle technická a jsou často svázána s technologií, kterou podporují. Datové sklady mají svá metadata, ETL software má svá metadata, podnikové informační systémy mají svá metadata a tak dále. Každý z těchto nástrojů je v technických metadatach soběstačný, ale chybí jejich globální popis ze strany firmy. Účelem firemního úložiště metadat je obsáhnout lokální metadata a uvést je tak, že budou užitečné celé firmě. To je velký úkol i pro malou firmu a pro rozsáhlou nadnárodní organizaci může jít o závazek kolosálních rozměrů. Je třeba zpracovat enormní množství metadat a všechny definice udržovat aktuální a přesné.

2.0.4 Podnikový slovník

Velice vhodným způsobem k ukládání definic podnikových metadat je dle [8] centralizovaný podnikový slovník (Business dictionary). Ten sjednotí firemní jazyk a umožní nebo ulehčí dorozumívání a jednotnou interpretaci reportů. Elektronický (webový) slovník v ohromné míře urychlí hledání pojmů a ušetří čas zaměstnancům. V ideálním případě by měly slovník i jeho položky být přesné, aktuální, úplné, srozumitelné a přístupné. Dále by měly být jednotlivé pojmy strukturovány takto:

- **Název** pojmu, který chci popsat. Pod tímto názvem je pojem používán v rámci BI aplikací.
- **Zobecnění**, například: Asistent je zaměstnanec.
- **Funkce tohoto pojmu**, například: Kredity se používají k ohodnocení obtížnosti předmětu.
- **Vlastnosti**, obvykle s použitím slova „má,“ například: Bakalářská práce má vedoucího a téma.
- **Odvozené pojmy**, například: Přihláška ke studiu může být podaná, čekající na schválení, schválená.
- **Příbuzné pojmy** nikoli synonyma, ale další definice, které by mohly čtenáře zajímat. Například: Zkouška; kredity, průchodnost, body.
- **Konkrétní instance** obecného pojmu, například: Instancí předmětu je Úvod do lineární algebry.

- **Synonyma**, tedy slova stejného významu.
- **Zdroj** - dokument, slovník, manuál, smlouva... Pokud lze, uvést autora.
- **Příklady použití**
- **Datum** vytvoření a poslední úpravy

Podle bakalářské práce Jakuba Krejčího [9], která se zabývá přímo metadaty na datovém skladu, by slovník měl dále obsahovat:

- **Obchodní pravidla** - např. Student si může jeden předmět zapsat pouze dvakrát.
- **Problémy s kvalitou dat** - Jedná se především o definici možných nevyplněných polí (např. z historických důvodů nebyl daný atribut používán).
- **Řízení přístupu** - tedy které osoby nebo skupiny osob mohou přistupovat k datům v datovém skladu.

Dále by dle ISO 11179 měl název pojmu být vždy v jednotném čísle, definice by měla obsahovat co nejméně zkratek a neměla by obsahovat definice dalších pojmů. Definice těchto dalších pojmů by se měly nacházet ve slovníku přímo každá na své stránce.

2.1 Podniková encyklopedie

Informované rozhodování lidí ve vedoucích pozicích firem a organizací, ať už to jsou manažeři, řiditelé, analytici nebo rektori, závisí na přehledných zprávách (reportech) z datového skladu a jejich správné interpretaci. Protože oni obvykle nemají znalosti ani čas na tvorbu reportů, je třeba platforma, která zefektivní komunikaci mezi nimi a techniky, kteří se starají o datový sklad. [10] Tato platforma se nazývá Podniková Encyklopedie a dle [6] slouží zejména k těmto účelům:⁶

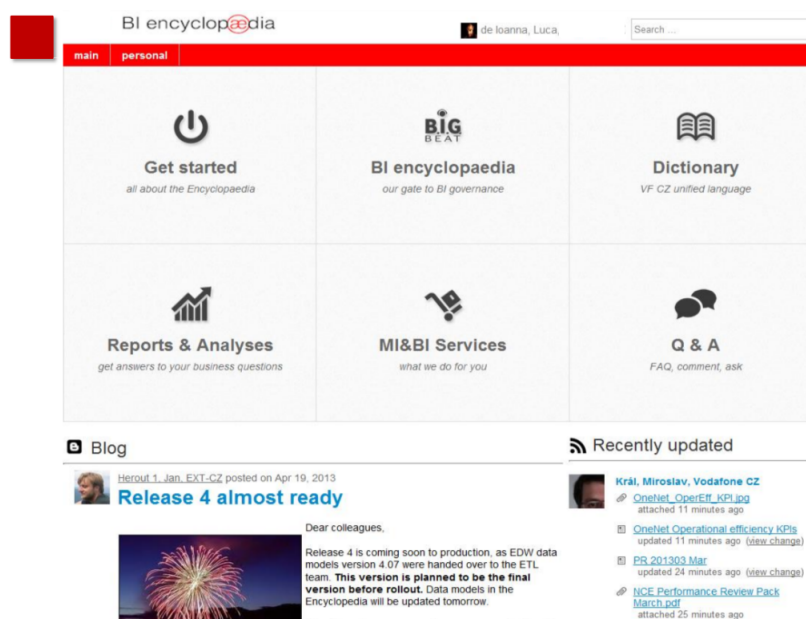
- Vkládání žádostí o reporty z datového skladu
- Katalog reportů - dokumentace, úložiště a správa reportů

⁶Toto je obecný návrh, v praxi může encyklopedie mít některé komponenty navíc nebo některé postrádat dle požadavků společnosti.

2. METADATA

- Podnikový (Business) slovník - Správa a zobrazení podnikových výrazů
- Datový slovník - prezentace datového slovníku, technických metadat, datových modelů, dokumentace zdrojových systémů
- Katalog služeb - evidence služeb napříč odděleními společnosti

Encyklopedie nabízí ty nejdůležitější informace pro rozhodování a vedení na jednom místě. Jakub Krejčí jí ve své práci [9] označuje i jako dokumentaci datového skladu. Dle studie uvedené v [8] je nejméně 50% hledání business informací neúspěšných, což má několik negativních následků. Zaměstnanec tím zbytečně ztratí čas a zaměstnavatel prodělá na ušlé příležitosti. Navíc, pokud se jedná o report, který už byl vypracován, ale není k nalezení, tvůrce reportů ztratí další čas znovuvytvořením toho reportu.



Obrázek 2.3: Hlavní okno podnikové encyklopedie jednoho z českých telefonních operátorů [10]

Část II
Praktická část

Analýza

3.1 Současný stav

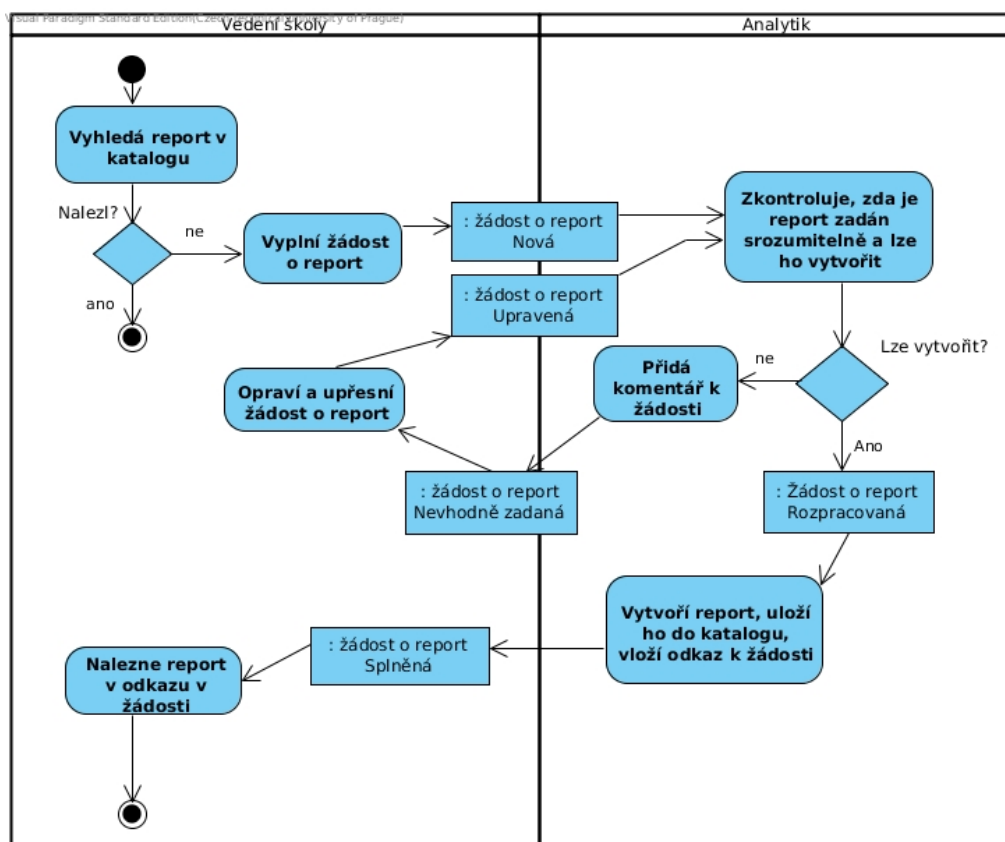
Současný stav získávání informací (reportů, statistik..) ze školních systémů pro informované rozhodování je málo organizovaný, nepřehledný a neefektivní. Žádosti i reporty se posílají přes elektronickou poštu a jsou často ukládány na soukromých počítačích. Pozdější hledání tohoto reportu může být problematické a rovněž jeho dostupnost pro ostatní je špatná. Definice pojmů se často nachází v útržkách informací v informačních systémech a v myslích lidí. Tyto předpoklady se projevují v pomalém a špatně informovaném rozhodování. Naprosto chybí centrální úložiště znalostí, jakým by měla být podniková encyklopedie. Na obrázku 3.1 jsem namodeloval diagram aktivity „Získání reportu“, jak by měla vypadat v budoucnu.

3.2 Úvod do analýzy

Celé téma podnikových encyklopedií je poměrně nové, a tak je problém získat kvalitní materiály o nejlepších způsobech návrhu a tvorby. Velkým zdrojem informací pro mne byla přednáška z magisterského předmětu „Enterprise data warehouses“ (MI-EDW) od Lucy de Ioanny [10]. Luca de Ioanna měl na starost vytvoření podnikové encyklopedie k datovému skladu jednoho z českých telefonních operátorů. Mnoho dalších poznatků mi sdělil v naší e-mailové konverzaci [11].

V České republice existují firmy, které vytvoří celý portál přesně pro konkrétní požadavky firmy, včetně integrace s databázemi a dalšími firemními systémy. Dá se předpokládat, že k vytvoření uceleného řešení pro velkou firmu je třeba tým odborníků, peníze na jejich platy a na softwarové

3. ANALÝZA



Obrázek 3.1: Diagram aktivit - žádání o report

licence. Celá tvorba může trvat několik měsíců v závislosti na velikosti firmy.

3.3 Požadavky

Na základě typických rysů podnikové encyklopedie jsem vytvořil následující požadavky:

3.3.1 Funkční

1. Encyklopedie obsahuje podnikový a datový slovník, do kterého lze přidávat definice podnikových a datových výrazů.⁷

⁷Cílem této práce je pouze navrhnout a vytvořit prostředí pro slovníky, plnit a spravovat je budou osoby ze strany univerzity.

2. Encyklopedie má rozhraní pro zažádání o nový report. Rozhraní obsahuje název a popis žádosti, lze nastavit datum, dokdy je nutné report vypracovat. Lze nastavit důležitost a osobu, která má report vytvořit.
3. Encyklopedie obsahuje katalog vypracovaných reportů.
4. Žádosti lze sledovat, upravovat a připojit k nim soubory.
5. Lze vyhledávat ve slovníku, žádostech o report i již vypracovaných reportech.
6. Encyklopedie umožňuje přihlašování a dle typu účtu zpřístupní některé funkce nebo obsah.

3.3.2 Nefunkční

1. Encyklopedie je přehledná.
2. Používání se lze snadno naučit.
3. Rozhraní je v českém jazyce.
4. Systém bude plně funkční na Firefox 30.0 [12] a Chrome 38.0 [13].
5. Kompatibilita s fakultní PostgreSQL databází.

3.4 Modelování případu užití

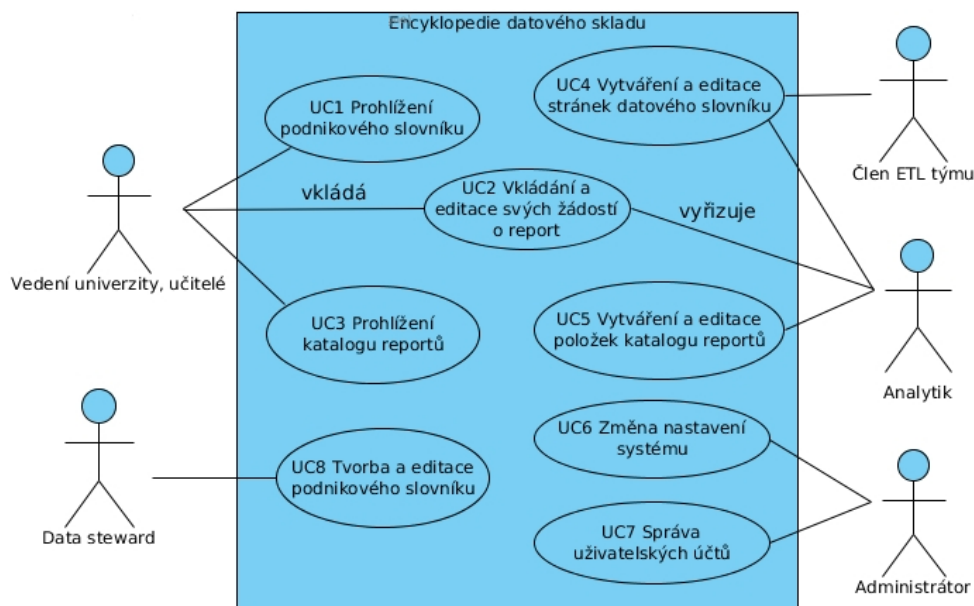
3.4.1 Seznam účastníků

1. **Vedení univerzity** - Rektor, děkanové, proděkanové, vedoucí kateder a další zaměstnanci školy. Encyklopedie jim bude sloužit jako zdroj znalostí - budou moci prohlížet centralizovaný katalog reportů a žádat o nové reporty, bude moci číst v podnikovém slovníku.
2. **Analytik** - Vyřizuje žádosti o report, provádí analýzy nad datovým skladem, komunikuje s vedením univerzity, přidává položky do katalogu reportů, může editovat datový slovník.
3. **Data steward** - Vytváří a edituje podnikový slovník, je zodpovědný za jeho korektnost a kvalitu.
4. **člen ETL týmu** - Stará se o nahrávání dat ze zdrojových systémů, spravuje data v datovém skladu. Může používat i upravovat datový slovník.

3. ANALÝZA

5. **Administrátor** - Spravuje celou encyklopedii, může měnit její nastavení a spravuje uživatelské účty.

Jedna osoba může vykonávat i více rolí najednou. Na základě seznamu účastníků a požadavků jsem vypracoval diagram případu užití, který je znázorněn na obrázku 3.2.



Obrázek 3.2: Diagram případu užití

3.4.2 Popis případů užití

UC1 Prohlížení podnikového slovníku - Uživatel smí vstoupit do podnikového slovníku a prohlížet si obsah jednotlivých stránek.

UC2 Vkládání a editace svých žádosti o report - Uživatel otevře formulář „žádost o report“. Vyplní název žádosti, co nejpodrobněji a nejpřesněji popíše, jaké informace by chtěl obdržet. Nepovinně může zadat, dokdy chce mít report hotový, kdo ho má vypracovat a zda je soukromý nebo veřejný.

UC3 Prohlížení katalogu reportů - Uživatel vstoupí do sekce uložených reportů, kde bude jejich seznam. U každé položky bude autor, datum vypracování, odkaz na soubor nebo na jiné úložiště, odkaz na žádost o report (bude-li dostupná). Uživatel může stáhnout a otevřít report. V případě jiného úložiště může být soubor dostupný jen pro některé uživatele. Pokud

zadá vypracovatele, tak ten dostane upozornění prostřednictvím e-mailu.

UC4 Vytváření a editace obsahu datového slovníku - Uživatel smí vstoupit do sekce datového slovníku, vytvářet nové položky, připojovat k nim soubory, psát obsah, prohlížet historii změn stránky, upravovat existující stránky.

UC5 Vytváření a editace položek katalogu reportů - Uživatel smí vstoupit do sekce uložených reportů a vložit novou položku. U každé položky bude autor, datum vypracování, odkaz na soubor nebo na jiné úložiště, odkaz na žádost o report (bude-li dostupná).

UC6 Změna nastavení systému - uživatel smí upravovat základní nastavení celého systému. Konkrétní možnosti budou záviset na použitém systému.

UC7 Správa uživatelských účtů - uživatel smí vytvářet uživatelské účty a měnit přístupová a jiná práva všem uživatelům.

UC8 Tvorba a editace podnikového slovníku - Uživatel vstoupí do sekce „podnikový slovník“ a může vytvářet, upravovat a mazat obsah stránek slovníku i celé stránky. Stránky slovníku bude tvořit podle formátu naznačeném v teoretické části.

Případy užití společné pro všechny uživatele:

UC9 Uživatel se přihlásí pomocí uživatelského jména a hesla. Dle jeho role se mu zpřístupní některé funkce portálu.

UC10 Uživatel vyhledá výraz pomocí textového pole. Vyhledávat bude možní ve slovníku, žádostech o report a v katalogu reportů.

Žádný z use casů jsem neshledal natolik nezřejmým, že by bylo třeba vytvořit jeho scénář.

3. ANALÝZA

Požadavky	Případy užití									
	UC1	UC2	UC3	UC4	UC5	UC6	UC7	UC8	UC9	UC10
F1	+			+				+		
F2		+								
F3			+		+					
F4		+								
F5									+	
F6										+

Tabulka 3.1: Kontrola splnění funkčních požadavků.

Návrh

Dle mého průzkumu neexistuje freeware systém, který by byl vytvořen přímo jako podniková encyklopedie. Nicméně tvořit takový portál přímo na míru požadavkům by nebylo efektivní. Nabízí se možnost použít systém, který bude splňovat požadavky, ačkoli je navržen pro obecnější nebo jiné, ale podobné řešení. Luca de Ioanna k tvorbě encyklopedie pro telefonního operátora použil komerční systémy Confluence (wiki, podpora týmové spolupráce) a Jira (issue tracker) od firmy Atlassian [10]. Protože jeho projekt byl úspěšný, rozhodl jsem se inspirovat jeho řešením a připravit podobný systém pro datový sklad ČVUT, nicméně, jak už jsem uvedl, pouze s použitím open source software. Popíší pojmy wiki a issue tracker, protože je budu dále používat:

Issue tracker - Systém, který dovoluje firmě zaznamenávat a sledovat „issue,” což může znamenat jak upozornění na nedostatek, tak otázku nebo žádost. Některé Issue trackery mohou být zaměřené na zadávání bugů při vývoji softwaru, ale obecně rozhraní pro vložení issue obsahuje podobné prvky. Obvykle lze nastavit název, podrobnosti, důležitost, datum, zodpovědnou osobu a připojit soubor. Vytvořené požadavky lze vylistovat a obvykle třídít, prohledávat, filtrovat a upravovat. Typicky dostává pověřená osoba e-mail při vytvoření issue a stejně tak dostane zadavatel upozornění, je-li issue splněn. Issue tracker má pro své obecné rozhraní velikou škálu použitelnosti, od helpdesku po testování software nebo, jako v případě podnikové encyklopedie, k zadávání žádostí o report. Zdroj: [14]

Wiki⁸ - Web složený z mnoha stránek, které může prohlížet velká sku-

⁸Nezaměňovat pojem wiki s Wikipedií, ta je pouze jednou z mnoha implementací systému wiki

pina uživatelů a rovněž editovat může početnější skupina lidí. Stránky mohou být jak úplně veřejné, tak i soukromé, například firemní. Nejčastěji se používají pro sdílení informací, dokumentů, pro komunikaci a spolupráci. Typický rys je odkazování neboli hyperlinkování mezi jednotlivými stránkami. Intuitivně a rychle se tak uživatel může navigovat skrz stránky a hledat informace. Obsah jednotlivých stránek se píše ve speciální syntaxi, která je ale velice lehká na naučení a používání. Samozřejmě lze ke stránkám vkládat obrázky a soubory a vyhledávat. Typicky se ke každé stránce automaticky ukládá historie změn a jejich autor. Zdroj: [15]. Wiki systém by byl ideální pro podnikový slovník - každý pojem by představovala jedna stránka a díky hypertextovým odkazům může čtenář rychle hledat definice.

4.1 Rešerše systémů

Alternativy zmíněných systémů jsem hledal na webu www.alternative.to, který umožňuje filtrovat software dle druhu licence, dále zobrazuje podrobnosti a popularitu dle uživatelů. Soustředil jsem se zejména na ty nejpopulárnější, které obvykle mají nejvíc funkcí, možnosti rozšíření a kvalitní podporu, ale zahrnul jsem do svého výběru i některé nové projekty. Mimo to jsem položil dotazy na webech stackoverflow.com a softwarerecs.stackexchange.com. Nadnesl jsem kritéria podnikové encyklopedie a požádal komunitu o doporučení. Většina odpovědí se shodovala se systémy doporučenými na webu www.alternative.to.

Wordpress, Trac, OpenProject a Redmine jsem zkoušel na svém osobním počítači. Využil jsem předkonfigurovaných freeware balíčků od vývojářské a hostingové společnosti Bitnami. Tyto balíčky, neboli stacky, obsahují zmiňovaný systém v aktuální nebo téměř aktuální verzi, Apache server a případně databázový server. Jejich výhoda je zřejmá - balíček se nainstaluje a celý systém je nakonfigurovaný a měl by správně fungovat bez další administrace a nastavování. Asanu a Taigu jsem vyzkoušel hostované na oficiálních serverech. Taigu sice lze stáhnout a hostovat na lokálním serveru, ale tato volba nepřináší žádné výhody, které bych pro potřeby testování využil.

4.1.1 Redmine

- Oficiální webová stránka: www.redmine.org
- Použitá verze: 3.0.2 [16]
- Licence: Open source

- Programovací jazyk: Ruby

Redmine je systém pro řízení projektů, který lze použít i pro softwarové projekty. Obsahuje wiki, issue tracker, prostředí pro diskuzní fóra, možnost integrace se systémem pro řízení verzí, ganttův diagram a další. Pro potřeby encyklopedie by se dala wiki použít jako slovník, dokumenty jako úložiště reportů a issue tracker na zadávání žádostí o report. Hlavní stránku lze personalizovat - například tam lze vložit přehled mnou zadaných úkolů, kalendář nebo dokumenty. Samozřejmostí je vyhledávání napříč všemi těmito moduly. Redmine vznikl již v roce 2006 a nabízí velikou uživatelskou podporu a mnoho bezplatných pluginů vytvořených uživateli. Bohužel velká část pluginů není kompatibilní s nejnovější verzí a je otázka, jestli a jak rychle budou aktualizovány. K Redminu se dá přistupovat i prostřednictvím mobilní aplikace. Lze nastavit role a práva a v nastavení lze přepnout celé rozhraní do českého jazyka. Vizuální vzhled vypadá poněkud stroze ve srovnání s ostatními systémy, ale toto téma se dá změnit opět na jedno z mnoha uživatelsky vytvořených. Drobný problém jsem zaznamenal s nastavením, protože je poněkud rozsáhlé a nepřehledné a konfigurování může být matoucí. Redmine splňuje všechny požadavky a byl by vhodným kandidátem pro encyklopedii.

The screenshot shows the Redmine interface for the project 'Encyklopedie Datového skladu ČVUT'. The main section is titled 'Úkoly' (Tasks). There are filter options for 'Stav' (Status) set to 'otevřený' (open) and 'Přidat filtr' (Add filter). Below the filters, there are action buttons: 'Použít' (Use), 'Smazat' (Delete), and 'Uložit' (Save). The main content is a table with the following data:

#	Fronta	Stav	Priorita	Předmět	Aktualizováno	Autor
3	Žádost o report	Nový	Vysoká	závěrečné práce	2015-04-10 20:01	Martin Čejka Admin
2	Žádost o report	Nový	Nizká	přijímací řízení	2015-04-10 19:59	Martin Čejka Admin
1	Žádost o report	Nový	Vysoká	průchodnost BI-LIN	2015-04-07 11:28	Martin Čejka Admin

At the bottom left, it says '(1-3/3)'. At the bottom right, there is a link: 'Také k dispozici: Atom | CSV | PDF'.

Obrázek 4.1: Redmine - Zadané žádosti

4.1.2 Wordpress

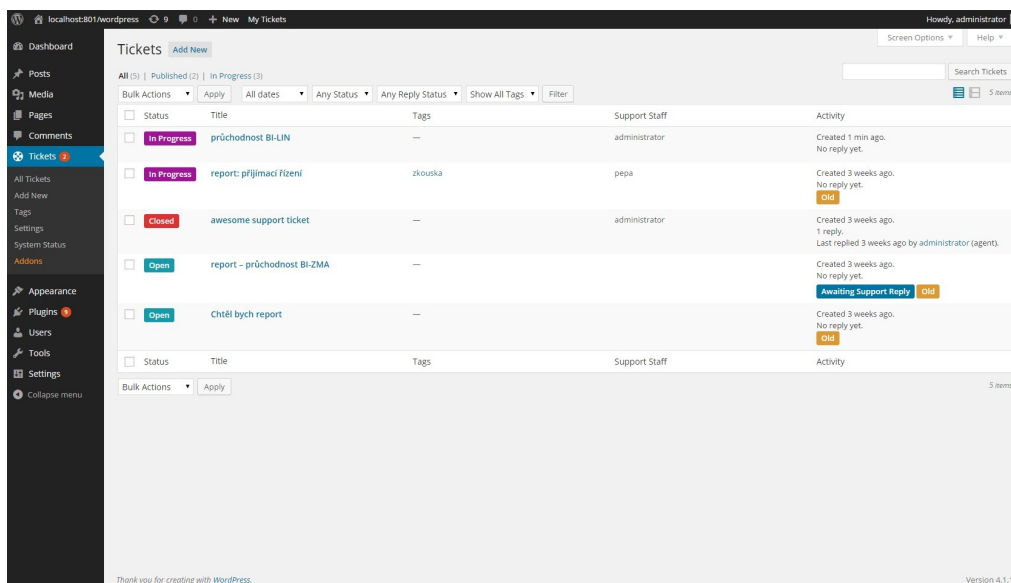
- Oficiální webová stránka: www.wordpress.com
- Použitá verze: 4.1.1 [17]
- Licence: Open source
- Programovací jazyk: PHP

Wordpress je široce používaný redakční systém. Sice vznikl jako blogovací nástroj, ale časem se vyvinul ve velice rozšiřitelnou platformu. Jedná se o CMS⁹ a své oblíbenosti dosáhl díky pestré nabídce pluginů, které se dají nainstalovat. Bohužel, většina pluginů byla vytvořena pro komerční účely a bez zaplacení nabízejí pouze omezenou funkčnost. Obsah Wordpressu se tvoří do tzv. stránek, což jsou nezávislé příspěvky obsahující naformátovaný text, obrázky, odkazy apod. Rovněž se mezi nimi dá vyhledávat text a lze je řadit do kategorií. Tyto stránky by mohly tvořit hesla slovníku. Celé prostředí vypadá moderně a to přispívá použitelnosti celého systému. Wordpress není sám o sobě určen pro týmovou spolupráci, a tak postrádá prostředí pro tvorbu a správu žádostí o reporty. Tato funkcionality se dá doplnit pluginy. Nalezl jsem a vyzkoušel těchto sedm různých pluginů:

- WP Support plus
- WP Ticket
- Awesome support
- Sola Support Ticket
- TaskFreak
- WP Project Manager
- WATS

Bez zacházení do podrobností mohu obecně říci, že některé byly špatně navrženy nebo obsahovaly výrazné bugy. Ty ostatní měly z důvodu neplacené licence omezenou funkčnost natolik, že se nedaly použít jako praktický issue tracker - např. chyběla možnost posílat e-mail při vložení žádosti. Z toho důvodu nelze použít Wordpress jako samostatnou podnikovou encyklopedii.

⁹Content Management System, Systém pro správu obsahu



Obrázek 4.2: Wordpress

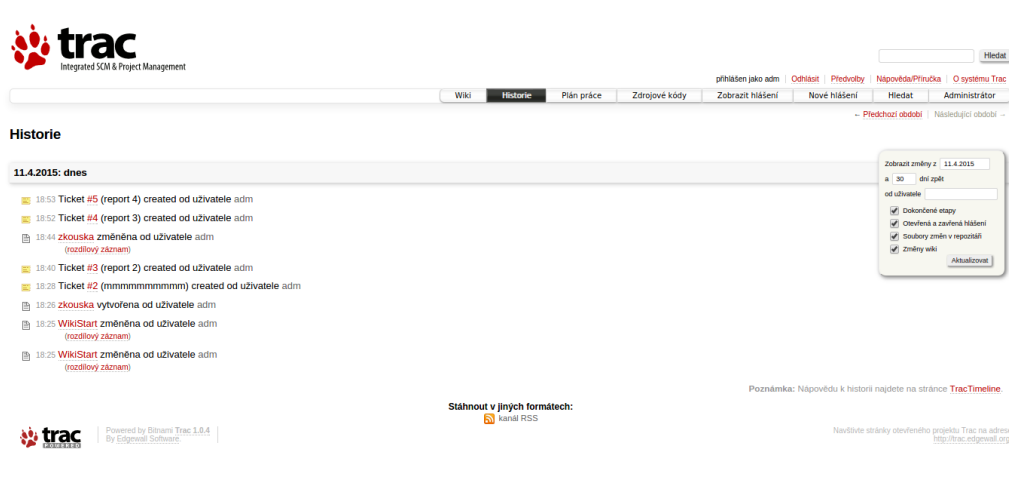
4.1.3 Trac

- Oficiální webová stránka: www.trac.edgewall.org
- Použitá verze: 1.0.4 [18]
- Licence: Open source
- Programovací jazyk: Python

Trac je wiki a issue tracking systém pro projekty, při kterých se vyvíjí software. Obsahuje i podporu pro systémy na řízení verzí. K systému lze přistupovat i přes mobilní aplikaci vytvořenou přímo pro Trac. Celé prostředí a funkcionality se velice podobá Redminu. Jazyk textů lze přepnout do češtiny a základní funkcionality lze rozšířit pluginy, například jde takto integrovat do Tracu LDAP autentizaci. Na webu track-hacks.org je k nalezení asi 200 pluginů vytvořených komunitou, včetně zajímavých rozšíření wiki a grafických témat. Při zobrazování zadaných žádostí o report je možno vyhledávat pomocí přednastavených SQL dotazů a vyfiltrovat tak výsledky. Samozřejmě lze fulltextově vyhledávat i nastavovat účtům různá oprávnění. Naneštěstí mnohá nastavení, například maximální velikost přílohy u reportu, nelze z nějakého důvodu nastavit přes webové rozhraní, ale je třeba editovat textový soubor s konfigurací. Trac by mohl být slušným kandidátem na encyklopedii, ale chybí mu modul podobný dokumentům,

4. NÁVRH

kteřé jsou v Redminu, OpenProjectu a Taize. Neexistuje místo, kam by se mohly přehledně ukládat reporty a žádný plugin tuto funkcionalitu nepřidává. Soubory se sice mohou připojovat k wiki stránkám a k položkám v issue trackeru, ale obě tyto varianty by byly velice nepraktické.



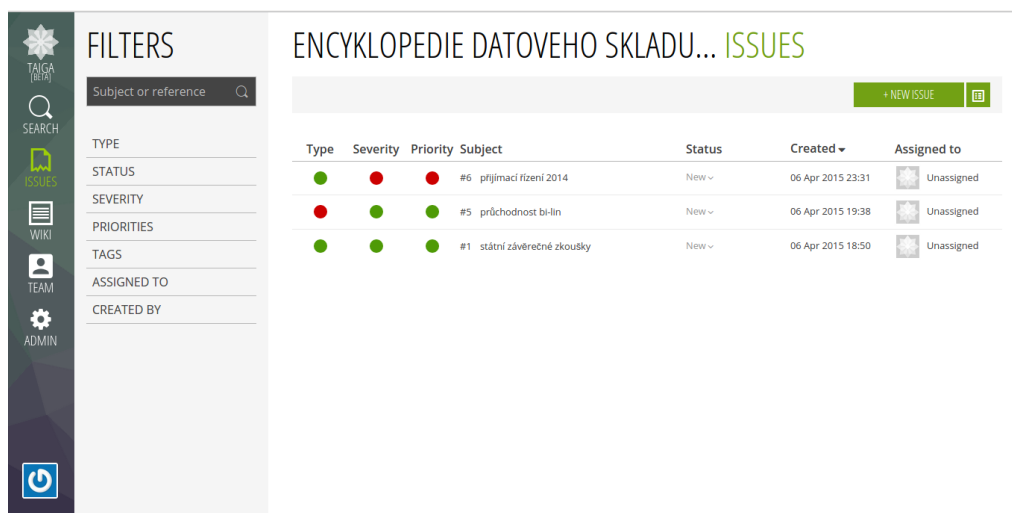
Obrázek 4.3: Trac - Historie akci

4.1.4 Taiga

- Oficiální webová stránka: www.taiga.io
- Použitá verze: Nelze zjistit
- Licence: Open source
- Programovací jazyk: Ruby

Taiga je moderní systém pro správu projektů, který se v době mé rešerše nacházel ve fázi beta testování. Obsahuje nástroje pro řízení agilního vývoje, ale i komponenty, které jsou důležité pro podnikovou encyklopedii - wiki a issue tracker. Navigace je intuitivní a celé prostředí vypadá velice moderně, dokonce bych řekl nejlépe ze všech testovaných systémů. Wiki modul vypadá typicky, ale má navíc užitečné menu s trvalými odkazy na přednastavené stránky z wiki. Issue tracker nabízí klasické možnosti, ale ke příspěvkům nelze vkládat soubory. Vytvořené reporty jde filtrovat dle mnoha kritérií a Taiga obsahuje i vyhledávání napříč wiki stránkami a issue trackerem. Největší problém je absence prostoru určenému k ukládání reportů. Jelikož k příspěvkům v issue trackeru nelze přikládat soubory, jediná možnost je

vytvářet pro každý report wiki stránku a k ní připojovat soubor s reportem, což ztěžuje jejich ukládání a vyhledávání. Další nevýhoda tkví v jazyce rozhraní - je celé v angličtině, což ubírá použitelnosti stránky, vzhledem k tomu, že všichni uživatelé nemusí mít angličtinu na příslušné úrovni. Celý systém je ve fázi vývoje, což má za následek malou nabídku pluginů a rozšíření, slabší podporu a větší riziko chyb a nedodělků.



Obrázek 4.4: Taiga - Issue tracker

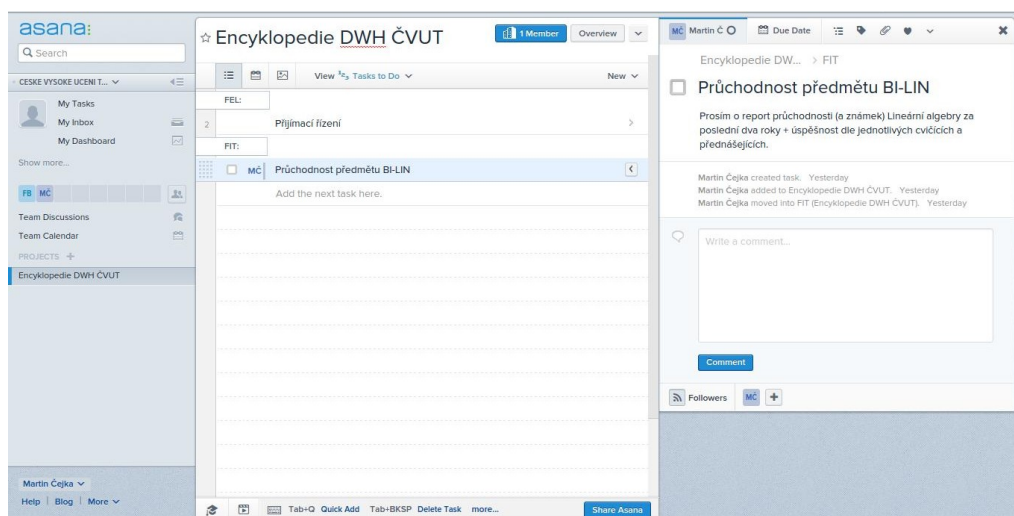
4.1.5 Asana

- Oficiální webová stránka: www.asana.com
- Použitá verze: Nelze zjistit
- Licence: Freemium
- Programovací jazyk: Javascript

Asana je systém pro týmovou spolupráci, který je na rozdíl od většiny mnou zkoušených systémů hostovaný na straně poskytovatele. Licencí se jedná o tzv. *freemium*, tedy zdarma, ale zpřístupňující další funkce při placení měsíčních poplatků. Asana je velmi přehledný a systém s mnoha funkcemi a dobře navrženým designem. Při zadávání úloh lze kromě standartních funkcí přidat tagy pro snadné vyhledávání a vytvářet podúlohy. Úlohy se mohou třídit do kategorií. Užitečným nástrojem v Asaně je týmová diskuze, kam lze rychle vkládat zprávy nebo klást otázky pro celý tým. I přes všechny

4. NÁVRH

klady zde chybí jakákoli možnost tvořit strukturovaný obsah pro podnikový slovník a Asana postrádá možnost instalace pluginů či rozšíření, které by mohly tuto funkčnost přidat. Protože jí nelze hostovat na vlastním serveru, nelze ani implementovat vlastní pluginy nebo další funkcionalitu. Dále z důvodu free licence může být na jeden projekt přiřazeno nanejvýš 15 uživatelů. Z těchto důvodů shledávám systém Asana nevhodným pro použití jako podniková encyklopedie.



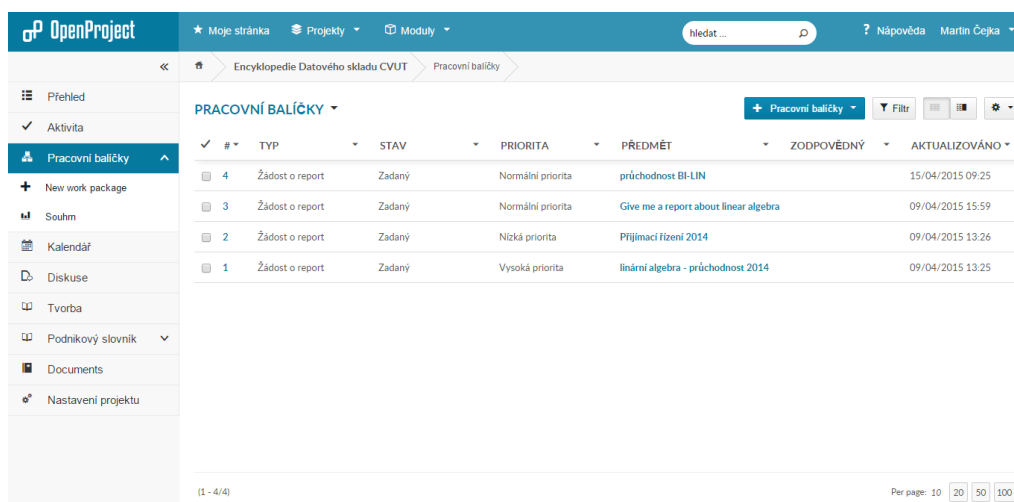
Obrázek 4.5: Hlavní obrazovka systému Asana

4.1.6 OpenProject

- Oficiální webová stránka: www.openproject.org
- Použitá verze: 4.0.9 [19]
- Licence: Open source
- Programovací jazyk: Ruby

OpenProject je systém pro efektivní řízení projektů spuštěný v roce 2012. Jedná se o tzv. „fork“ (nezávisle vyvíjenou odnož) systému Redmine, a tak není překvapivé, že má velice podobné funkce. Obsahuje issue tracker, wiki, dokumenty, kalendář podporu pro agilní vývoj, diskuze, meetingy a další nástroje pro řízení týmu a projektu. Lze nastavit role a práva a samozřejmě je vyhledávání napříč všemi těmito moduly. Pro potřeby encyklopedie by se dala wiki použít jako slovník, dokumenty jako úložiště reportů a issue tracker na zadávání žádostí o report. Hlavní stránku lze personalizovat -

například tam lze vložit přehled mnou zadaných úkolů, kalendář nebo dokumenty. OpenProject nabízí rozšiřitelnost pomocí pluginů, ale je jich výrazně než např. v Redmine. Grafické prostředí se dá přepnout do češtiny, ale místy jsem nacházel nepřeložená textová pole v angličtině. Ovládání je příjemné a intuitivní. Celkově by OpenProject mohl být velice dobrý kandidát na encyklopedii datového skladu.



#	TYP	STAV	PRIORITA	PŘEDMĚT	ZODPOVĚDNÝ	AKTUALIZOVÁNO
4	Žádost o report	Zadaný	Normální priorita	průchodnost BI-LIN		15/04/2015 09:25
3	Žádost o report	Zadaný	Normální priorita	Give me a report about linear algebra		09/04/2015 15:59
2	Žádost o report	Zadaný	Nizká priorita	Přijímací řízení 2014		09/04/2015 13:26
1	Žádost o report	Zadaný	Vysoká priorita	linární algebra - průchodnost 2014		09/04/2015 13:25

Obrázek 4.6: Zaznamenané žádosti o report v systému OpenProject

4.1.7 Volba systému

Ve spektru mnou hodnocených systémů se objevily velice vhodné systémy i nepoužitelné. Wordpress a Asana postrádaly některé základní funkční požadavky a tím pádem by encyklopedie s jejich použitím byla takřka nerealizovatelná. Taiga a Trac by se teoreticky daly použít, ale některé funkce by musely být vyřešeny poněkud složitě a systém by se stal nepřehledným a hůře použitelným. Do nejužšího výběru se dostaly dva - Redmine a OpenProject. Oba tyto systémy splňují funkční požadavky a nabízí funkční i vzhledovou rozšiřitelnost. Pro závěrečnou volbu nezbývá, než porovnat tyto dva. OpenProject nabízí modernější platformu a velice pěkný vzhled, Redmine obrovské množství pluginů a grafických témat, navíc se dá očekávat větší podpora, protože systém je starší a má mnohem větší komunitu uživatelů. OpenProject je proti tomu nováček, který se sám označuje za nástupce Redminu. Funkcionalitou jsou takřka totožné, protože OpenProject je vývojová odnož Redminu. Nakonec jsem zvolil Redmine, zejména kvůli velké podpoře a nabídce pluginů.

4.2 Podoba slovníků

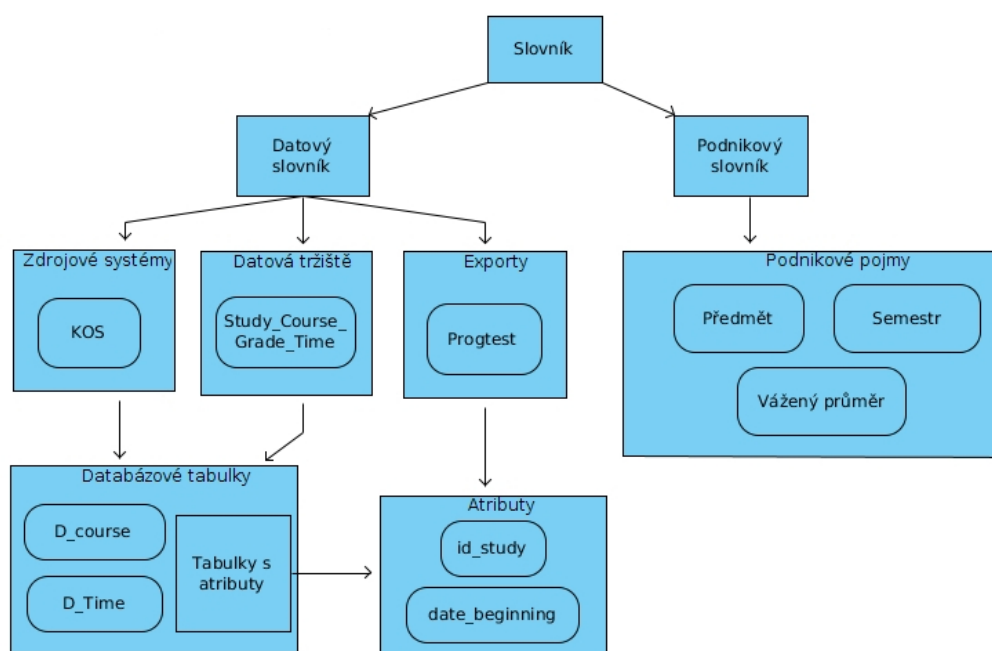
Jak jsem již uvedl, slovníky budu realizovat pomocí wiki modulu, který dovoluje základní formátování textu, seznamy, nadpisy, obrázky, tabulky a další. Díky hypertextovým odkazům a vyhledávání budou informace zachovávané přehledně a dostupně. Všechny stránky mohu rozdělit do několika skupin:

- **Podnikový pojem** - Popis podnikového výrazu (business metadata), jak byl navržen v teoretické části. Bude obsahovat informace pro vedení školy a zároveň odkazy do souvisejících položek technického slovníku.
- **Databázový atribut** - Bude obsahovat popis atributu v pár větách a všechny důležité technické informace. Může se jednat o atribut ze zdrojové i cílové databáze a v případě, že nějaký zdrojový atribut má svůj ekvivalent v datovém skladu, budou oba popsány v jedné stránce. Tato stránka navíc obsahuje i logickou datovou mapu, jejíž schéma se nachází v teoretické části 2.1, a informace o transformaci dat.
- **Databázová tabulka** - Tato stránka bude obsahovat pouze tabulku, ve které budou atributy z této databázové tabulky. V tabulce se bude nacházet datový typ atributu, velice krátký business popis a případně informaci, zda se jedná o primární nebo cizí klíč. Název atributu bude zároveň odkaz přímo na jeho stránku, kterou jsem popsal v předchozím bodě. Dále se zde bude nacházet popis databázové tabulky a odkazy do business slovníku.
- **Databáze** (KOS, Progtest, datová tržiště...)- Tato stránka bude obsahovat diagram datového modelu databáze a seznam všech databázových tabulek vložených pomocí makra *include*¹⁰.
- **Export** - Stránka o exportech bude obsahovat tabulku atributů v daném exportu podobnou tabulce atributů ve stránce o databázové tabulce. V tabulce se bude nacházet datový typ atributu, velice krátký business popis a případně informaci, zda se jedná o primární nebo cizí klíč. Atributy bude možné otevřít a přejít přímo na stránku o atributu.

Základní schéma slovníku jsem znázornil na diagramu 4.7. V praxi budou jednotlivé sekce daleko více propojeny díky odkazům na všech stránkách. Jeden z hlavních důvodů tvorby slovníku je sjednocení informací, a

¹⁰Makro *include()* vloží celý obsah jiné stránky na místo, kde použito.

tak jsem při návrhu slovníku dbal na to, aby každá informace byla uložena na co nejméně místech. Jinak by při neopatrným úpravám mohlo dojít k rozporu v informacích mezi různými stránkami.



Obrázek 4.7: Základní schéma návrhu slovníku

Realizace a testování

V Redminu se všechny moduly (wiki, dokumenty, issue tracker...) váží k určitému projektu, a tak jsem vytvořil primární projekt „Encyklopedie datového skladu“. Na hlavní stranu tohoto projektu jsem vložil rozcestník, který odkazuje do těch nejdůležitějších modulů. Dále jsem přejmenoval atributy ve zdrojových souborech, a tak portál ukazuje „Slovník“, „Zadat žádost“ a „Žádosti“ namísto původních nápisů „Wiki“, „Úkoly“ a „Nový úkol“. V zabudovaném issue trackeru jsem vytvořil typ issue „žádost o report“ a nastavil jsem ho jako defaultní. Nastavil jsem cyklus žádosti tak, jak je znázorněn v diagramu aktivit 3.1 Nastavitelné priority jsem zvolil jako nízká, střední a vysoká. Vytvořil jsem uživatelské role dle diagramu případu užití 3.2 a seznamu účastníků. Dále jsem vytvořil uživatelskou skupinu „Analytici“, které je možno přiřazovat žádosti o report. Ukázka současné obrazovky zadaných reportů se nachází na obrázku B.5.

5.1 Slovník

Ačkoli se v analýze vyskytovaly podnikový a datový slovník odděleně, rozhodl jsem se je tvořit nad jedním wiki modulem. Alternativní možnost by byla vytvořit druhý projekt, ve kterém by byla oddělená wiki, ale tato možnost by byla velice nepraktická, protože by se oddělily i issue tracker, dokumenty a ostatní moduly. Sloučení slovníků ničemu nevedí z hlediska funkčnosti ani z hlediska bezpečnosti.

Vytvořil jsem dvě základní stránky ve wiki modulu - podnikový a datový slovník. Podnikový slovník bude abecední seznam položek s odkazy na jejich stránky. Tyto položky se mohou případně seskupit do kategorií pro lepší přehlednost, bude-li to v budoucnu třeba. Technický slovník má seskupené zdrojové databáze, databáze ve skladu a experty, jak je naznačeno

na diagramu 4.7. Samotná hlavní stránka slovníku obsahuje jak podnikový, tak i datový slovník - použil jsem opět makro *include*. Zcela dole na hlavní stránce se nachází návody, kam jsem zároveň zanesl standardy a šablony pro další tvorbu slovníků. Zároveň jsem vytvořil několik ukázkových stránek jako vzor. Vytvořil jsem několik ukázkových stránek atributů, tabulek, databází a exportů dle návrhu. Změnil jsem podobu pravého menu a vložil do něj odkazy na podnikový a datový slovník a rozbalovací seznam všech stránek zobrazených na hlavní stránce slovníku. Toto menu urychlí navigaci ve slovníku. Podobu stránek slovníku si lze prohlédnout v příloze B.

5.2 Uchovávání soukromých informací

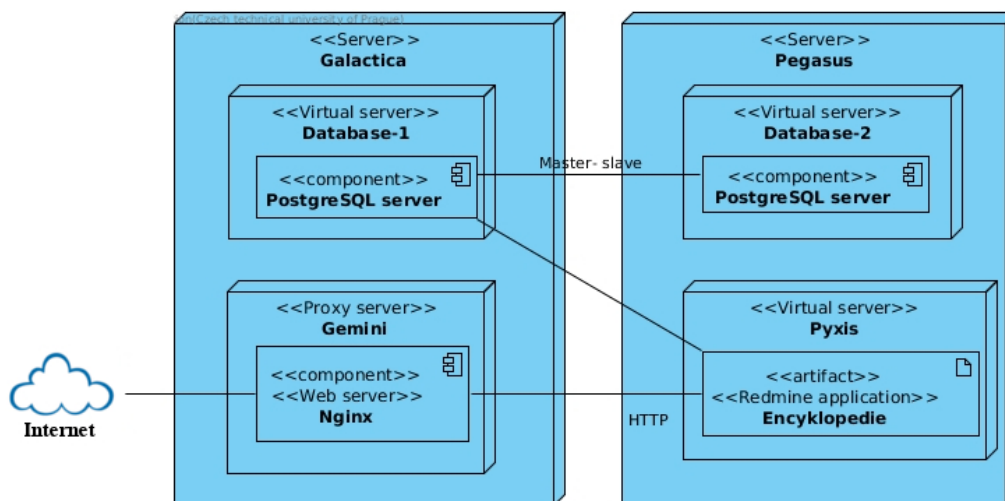
Ačkoli to ze současného zadání nevyplývá, v budoucnu by mohla vyvstat potřeba uchovávat v encyklopedii některé citlivé informace výhradně pro úzkou skupinu uživatelů, například jen pro ETL tým. Pokud by chtěli mít vlastní soukromé položky ve slovníku a chtěli použít wiki, za současné situace by to nebylo možné, protože řízení přístupu se vztahuje na celou wiki, ne na konkrétní stránky. Jako použitelné řešení se nabízí vytvořit druhý projekt, který bude mít svůj wiki modul, a přidat do něj pouze oprávněné uživatele. Velice užitečná je i ta vlastnost, že ze slovníku v hlavním projektu mohou vést odkazy a *include* makra do této soukromé wiki, ale samotný obsah se zobrazí pouze uživatelům ze soukromého projektu. Tento přístup zároveň umožňuje úplné skrytí požadovaných informací i je snadný k používání.

5.3 Nasazení

Jelikož nemám dostatečná oprávnění, nasazení prázdného systému Redmine na školní infrastrukturu provedl ing. Jakub Jirůtka, který administruje fakultní webové servery. Aktuální diagram nasazení jsem znázornil na obrázku 5.1. Protože se jedná o prototyp encyklopedie, bude zatím nasazena na školní VPN¹¹ a bude přístupná pouze ty, kdo mají přístup do VPN. Současná adresa je <http://dw-redmine.dev.vpn>. Plné nasazení se předpokládá v řádu několika týdnů až měsíců a je podmíněno několika fundamentálními činnostmi pro správný provoz encyklopedie (ustanovení týmu, vytvoření slovníku...). Aktuálně je přihlašování nastaveno na školní LDAP server, a lze se přihlásit pomocí údajů platných v celém ČVUT. Po prvním přihlášení

¹¹Virtual Private Network, soukromá pracovní síť

musí být osoba schválena administrátorem, který jí může přiřadit jednu z rolí.



Obrázek 5.1: Diagram nasazení encyklopedie na školní infrastrukturu.

5.4 Testování

Testování jsem provedl s ohledem na fakt, že jsem nakonfiguroval portál a navrhl slovník, ale neimplementoval jsem veškerou funkčnost portálu. Vzhledem k tomu, že se jedná o systém s velkou komunitou fungující již od roku 2006, se dá předpokládat, že systém je stabilní a bez větších funkčních nebo jiných chyb. Otestoval jsem všechny případy užití uvedené v analýze. Pro účely testování jsem vytvořil několik testovacích uživatelských účtů s rolí odpovídajícími diagramu případu užití 3.2. Ověřil jsem přiřazení žádosti o report konkrétními analytikovi i celé skupině a ověřil, že cyklus žádosti probíhá tak, jak jsem ho navrhl v diagramu aktivit 3.1. Lze vyhledávat napříč slovníkem, žádostmi o report i již vytvořenými reporty. Vytvořil jsem několik zkušebních položek ve slovníku dle mých šablon, abych ověřil jejich použitelnost. Všechny zmíněné testy jsem provedl na prohlížečích Chrome 38.0 [13] a Firefox 33.0 [12]. Nenalezl jsem žádnou funkční nebo jinou chybu na žádném z prohlížečů. Současnou podobu encyklopedie jsem zanesl do přílohy B.

Závěr

Náplní mé bakalářské práce bylo provést analýzu, návrh a realizaci encyklopedie pro efektivní využívání datového skladu fakulty. Mé řešení je aktuálně nasazeno a připraveno pro další fázi - vytváření obsahu podnikového a datového slovníku. Jakmile bude tato fáze hotová, encyklopedie se bude moci stát centrem znalostí pro vedení školy i analytiky a správce datového skladu. Podoba stránek slovníku, kterou jsem v této práci navrhl, zahrnuje všechny důležité informace a pokud bude dodržována, budou obsah stránek slovníku i celá struktura přehledné.

Návrhy, které jsem vypracoval, jsou obecně platné a mohou sloužit jako podklad pro tvorbu encyklopedie v menších a středních společnostech a organizacích. Z provedené rešerše vyplynulo, že i s použitím open source systémů se dá vytvořit vysoce funkční encyklopedie podobná komerčně nasazeným. Díky katalogu reportů budou všechny analýzy dostupné na jednom místě a snadno dohledatelné. Věřím, že mé řešení usnadní uživatelům získávání znalostí z univerzitních systémů a technikům zpřístupní informace o datových zdrojích. Zadání této práce považuji za splněné.

Použité zdroje

- [1] LARSON, B.: *Delivering business intelligence with Microsoft SQL Server 2008*. McGraw - Hill Companies, 2009, ISBN 9780071549455.
- [2] HEINZE, J.: History of Business Intelligence [online] poslední úprava 26.9.2014 [cit. 7.4.2015]. Dostupné z: <http://www.bisoftwareinsight.com/history-of-business-intelligence/>
- [3] MOSS, L.; ATRE, S.: *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Boston, MA, USA: Addison Wesley, 2003, ISBN 0201784203.
- [4] KIMBALL, R.; REEVES, L.; THORNTHWAITE, W.; aj.: *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses with CD Rom*. New York, NY, USA: John Wiley & Sons, Inc., první vydání, 1998, ISBN 0471255475.
- [5] KUZNETSOV, S.: *Datový sklad fakulty*. Diplomová práce, České vysoké učení technické v Praze, Fakulta informačních technologií, Praha, 2013.
- [6] BERÁNEK, L.: *Vizualizace technických a business metadat*. Diplomová práce, VŠE-FIS, Praha, 2012.
- [7] Organization, N. I. S.: *Understanding metadata*. NISO Press, 2004, ISBN 1880124629.
- [8] INMON, W.: *Business metadata: capturing enterprise knowledge*. Denise E.M. Penrose, 2008, ISBN 9780123737267.

- [9] KREJČÍ, J.: *Metadata pro datový sklad fakulty*. Bakalářská práce, České vysoké učení technické v Praze, Fakulta informačních technologií, Praha, 2015.
- [10] DE IOANNA, L.: Business metadata management [online], poslední úprava 11.6.2014 [cit. 7.4.2015]. Neveřejný školní server Edux, předmět BI-EDW, přednáška 9.
- [11] DE IOANNA, L.: Soukromé e-maily [online], únor - duben 2015 [cit. 7.4.2015]. E-mailová konverzace.
- [12] MOZILLA FOUNDATION and MOZILLA CORPORATION. Firefox 30.0 [software]. [přístup 18.4.2015]. Dostupné z: <https://www.mozilla.org/en-US/firefox/products/>
- [13] GOOGLE INC. Chrome 38.0 [software]. [přístup 18.4.2015]. Dostupné z: <https://www.google.com/chrome/browser/desktop/index.html>
- [14] ROUSE, M.: Issue tracking system (ITS) [online], poslední úprava 22.9.2010 [cit. 12.4.2015]. Dostupné z: <http://searchcrm.techtarget.com/definition/issue-tracking-system>
- [15] LEBAR, Z.: What are Wikis, and Why Should You Use Them? [online], poslední úprava 28.2.2014 [cit. 12.4.2015]. Dostupné z: <http://computers.tutsplus.com/tutorials/what-are-wikis-and-why-should-you-use-them--cms-19540>
- [16] LANG, J. a BITROCK INC. Bitnami Redmine stack 3.0.2 [software]. [přístup 20.3.2015]. Dostupné z: <https://bitnami.com/stack/redmine>
- [17] WORDPRESS FOUNDATION a BITROCK INC. Bitnami Wordpress stack 4.1.1 [software]. [přístup 13.3.2015]. Dostupné z: <https://bitnami.com/stack/wordpress>
- [18] EDGEWALL SOFTWARE a BITROCK INC. Bitnami Trac stack 1.0.4 [software]. [přístup 15.3.2015]. Dostupné z: <https://bitnami.com/stack/trac>
- [19] OPENPROJECT FOUNDATION a BITROCK INC. Bitnami OpenProject stack 4.0.9 [software]. [přístup 10.3.2015]. Dostupné z: <https://bitnami.com/stack/openproject>

Seznam použitých zkratk

- BI** Business Intelligence
- CMS** Content Management System
- DBMS** Database Management System
- ERD** Entity Relation Diagram
- ETL** Extract, Transform, Load
- SQL** Structured Query Language
- UML** Unified Modeling Language
- VPN** Virtual Private Network

Ukázka encyklopedie

Poznámka: Informace uvedené v obrázcích mohou být smyšlené.

The screenshot shows the main interface of the 'Encyklopedie Datového skladu ČVUT' website. The top navigation bar includes links for 'Úvodní', 'Moje stránka', 'Projekty', 'Administrace', and 'Nápověda'. The user is logged in as 'cejkama3'. The main header features a search bar and the site title. The 'Slovník' (Dictionary) section is active, displaying a list of terms under three categories: 'Podnikový slovník' (Company Dictionary), 'Datový slovník' (Data Dictionary), and 'Zdrojové databáze' (Source Databases). A right sidebar contains a 'Rozcestník' (Index) section with links to 'Podnikový slovník', 'Datový slovník', and 'Slovník'.

Encyklopedie Datového skladu ČVUT

Hledat: Encyklopedie Datového skladu ČVUT

Přehled Aktivita Žádosti Zadat Žádost Dokumenty **Slovník** Nastavení

Slovník [Upravit](#) [Sledovat](#) [Zamknout](#) [Přejmenovat](#) [Odstranit](#) [Historie](#)

Podnikový slovník

Všechny podnikové položky v abecedním pořadí:

- Bakalářská práce
- Kredit
- Semestr
- Student
- Vážený průměr
- Studium
- Předmět
- ID studenta

Datový slovník

Datový sklad

- Datamart Test
- Datamart study_grade_course_time
- Datamart progtestTask_study_time

Zdrojové databáze

- Zdroj Test
- KOS
- Edux
- Progtest

Exporty

- export KOS
- export Edux
- export Progtest

Návody pro uživatele:

- [Jak vytvořit žádost o report](#)

Návody pro správce:

- [Jak by měl vypadat uložený report](#)
- [Jak by měla vypadat položka podnikového slovníku](#)
- [Jak by měl vypadat atribut v datovém slovníku](#)
- [Jak by měla vypadat databáze a tabulka v datovém slovníku](#)

Rozcestník

- Podnikový slovník
- Datový slovník
- Rozbalit obsah hlavní stránky

Slovník

- Výchozí stránka
- Index dle názvu
- Index dle data

Obrázek B.1: Vzhled celé encyklopedie, hlavní stránka slovníku

B. UKÁZKA ENCYKLOPEDIE

The screenshot shows a navigation bar with tabs: Přehled, Aktivita, Žadosti, Zadat žádost, Dokumenty, **Slovník**, and Nastavení. Below the navigation bar, the breadcrumb trail is 'Slovník » Podnikový slovník »'. The main heading is 'Předmět'. The text describes the course as a basic unit of study, mentioning attributes like 'garant předmětu', 'rozsah kontaktní výuky', and 'způsob zakončení'. It also mentions ECTS credits and the organization of the course. At the end, it lists 'Obchodní pravidla:' and 'Související tabulky nebo atributy:' with a reference to a database table 'D_student_(KOS)_Test'.

Obrázek B.2: Položka v podnikovém slovníku

The screenshot shows a navigation bar with tabs: Přehled, Aktivita, Žadosti, Zadat žádost, Dokumenty, **Slovník**, and Nastavení. Below the navigation bar, the breadcrumb trail is 'Slovník » Datový slovník » Zdroj Test »'. The main heading is 'D_course (KOS) Test'. The text describes the table as representing a course. Below the text is a table with the following structure:

Český originální název	Popis	Databázový atribut	Typ atributu	PK/CK
KOD	Kód předmětu	K_Course_Code	VARCHAR	PK
ROLE	Role předmětu ve studiu	K_Role	VARCHAR	
ZPUSZAK	Způsob zakončení předmětu	K_Completion	VARCHAR	
ECTS_KREDITY	Počet kreditů podle ECTS	K_ECTS_Credits	INTEGER	

Below the table, there is a link 'Nový soubor' and a footer 'Také k dispozici: PDF | HTML | TXT'.

Obrázek B.3: Popis databázové tabulky v datovém slovníku

Přehled Aktivita Žádosti Zadat žádost Dokumenty **Slovník** Nastavení

Slovník » Datový slovník » Export KOS » [Upravit](#) [Sledovat](#) [Zamknout](#) [Přejmenovat](#) [Odstranit](#) [Historie](#)

k_semester code

Popis:
Kód semestru je čtyřznakový řetězec jednoznačně definující semestr.
Začíná písmenem B, pokračuje posledním dvoučíslím daného roku (14 pro 2014 atd).
Na jeho konci je 1 jako zimní nebo 2 pro letní semestr. (Zdroj)

Problémy s kvalitou dat:
žádné

Možné hodnoty a příklady:
od B951 výš, momentálně B142

Obchodní pravidla:
Viz popis

Specifikace entity v rámci zdroje:
Kód semestru se používá k časovému určení v mnoha zdrojových databázích . Například [předmět](#) je v rámci [studia](#) studován v určitém semestru.

Transformace:

Zdroje:

#	Databáze	Tabulka	Sloupec	Datový typ
1	Zdroj test	d_time	semestr_code	VARCHAR (4)
2	Edux	d_semester	semestr_code	VARCHAR (15)
3	Progtest	d_sem	seme_code	VARCHAR (4)

Cíle:

Databáze	Tabulka	Sloupec	Datový typ	Typ tabulky	Typ historizace
Datamart test	d_time	k_semester_code	VARCHAR (4)	Faktová	1

Obrázek B.4: Ukázka atributu v datovém slovníku

Přehled Aktivita **Žádosti** Zadat žádost Dokumenty Slovník Nastavení

Žádosti

▼ Filtry
 Stav vše Přidat filtr
 ▶ Nastavení

[Použít](#) [Smazat](#) [Uložit](#)

#	Předmět	Fronta	Stav	Priorita	Přiřazeno	Aktualizováno	Autor	Uzavřít do
4	průchodnost BI-ZMA	Žádost o report	Hotovo	Střední	analytik test	2015-05-10 16:23	vedeni test	2015-05-04
3	Progtest PA2 úloha 1	Žádost o report	Nový	Střední	Analytici	2015-05-10 15:55	vedeni test	2015-05-13
2	průchodnost BI-LIN	Žádost o report	Rozpracované	Vysoká	Martin Čejka	2015-05-10 13:19	vedeni test	2015-05-13
1	přijímací řízení	Žádost o report	Upravené	Nizká	Analytici	2015-05-10 15:18	vedeni test	

(1-4/4)

Obrázek B.5: Vytvořené žádosti o report

Přehled Aktivita Žádosti Zadat žádost **Dokumenty** Slovník Nastavení

Přijímací řízení 2014



Report
2015-05-10

Tento report ukazuje množství přijatých studentů dle jejich středních škol

související pojmy:
[Student](#)
[Přijímací řízení](#)

Vypracoval: Jan Souček
Zadavatel: Pavel Tvrdik, žádost #3
Report je přiložen

Soubory

 [Prijimaci rizeni 2014 report.pdf \(1.14 MB\)](#)  Martin Čejka, 2015-05-10 16:35

Obrázek B.6: Uložený report v katalogu

Obsah přiloženého CD

	readme.txt	stručný popis obsahu CD
	src		
		impl Zdrojové kódy implementace
		thesis zdrojová forma práce ve formátu L ^A T _E X
		text text práce
		thesis.pdf text práce ve formátu PDF