

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

AUTOMATICKÉ GENEROVÁNÍ WIKI STRÁNEK
Z E-MAILŮ A IS

DIPLOMOVÁ PRÁCE

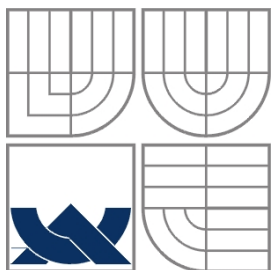
MASTER'S THESIS

AUTOR PRÁCE

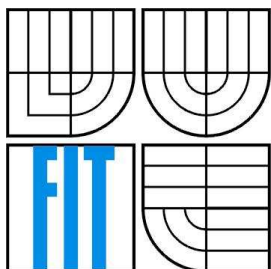
AUTHOR

Bc. ALEŠ VAVŘÍNEK

BRNO 2011



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

AUTOMATICKÉ GENEROVÁNÍ WIKI STRÁNEK Z E-MAILŮ A IS

GENERATION OF WIKI PAGES FROM E-MAILS AND IS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. ALEŠ VAVŘÍNEK

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. JAROSLAV DYTRYCH

BRNO 2011

Abstrakt

Tato diplomová práce se zabývá vytvořením rozšíření pro stávající informační systém NLPIS. Součástí řešení je vytváření a údržba automaticky generovaných stránek pro projekty do aplikace MediaWiki. Data budou do MediaWiki generována z e-mailové komunikace a databáze informačního systému NLPIS.

Abstract

This master's thesis deals with the creation of an extension to the existing information system NLPIS. A part of solution is the creation and maintenance of automatically generated pages for projects in the MediaWiki application. The data will be generated into the MediaWiki from database NLPIS information system and e-mail communications.

Klíčová slova

Automatické generování, web, stránka, informační systém, NLPIS, MediaWiki, HTML, CSS, PHP.

Keywords

Automatic generation, web, page, information system, NLPIS, MediaWiki, HTML, CSS, PHP.

Citace

Vavřínek Aleš: Automatické generování wiki stránek z e-mailů a IS, semestrální projekt, Brno, FIT VUT v Brně, 2011.

Automatické generování wiki stránek z e-mailů a IS

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením Ing. Jaroslava Dytrycha.

Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Aleš Vavřínek
18. 5. 2011

Poděkování

Za poskytnutou pomoc, odborné vedení a čas věnovaný při konzultacích k tématu mé diplomové práce děkuji svému vedoucímu Ing. Jaroslavu Dytrychovi.

© Aleš Vavřínek, 2011.

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

Obsah	1
Úvod	4
1 Současný stav	6
1.1 MediaWiki	6
1.1.1 Nástroje pro import dat	7
1.2 NLPIS	7
1.2.1 Pracovní výkazy	8
1.2.2 Plán práce	8
1.2.3 Projekty	8
1.2.4 E-mailové zprávy	9
2 Využité technologie	10
2.1 Jazyk HTML	10
2.1.1 Historie HTML	10
2.2 CSS - Kaskádové styly	11
2.3 Jazyk JavaScript	11
2.3.1 Historie JavaScriptu	11
2.4 Jazyk PHP	12
2.4.1 Historie PHP	12
2.5 Databáze MySQL	12
2.5.1 Historie MySQL	13
2.6 Syntaxe wiki (wiki markup)	13
2.6.1 Normalizace	14
2.7 PHP – CURL (Client URL Library)	14
2.7.1 HTTP	14
2.7.2 HTTPS	15
2.7.3 FTP	15
2.8 URL (Uniform Resource Locator)	16
2.9 Hashovací funkce	16
2.10 IMAP (Internet Message Access Protocol)	17
3 Návrh implementace	18
3.1 Návrh automaticky generovaných stránek	18
3.1.1 Aktualizace stránek skupiny projektů a projektu	18
3.1.2 Prostředky skupiny projektů a projektu	18
3.1.3 Stránka skupiny projektů	19

3.1.4	Stránka projektu.....	19
3.1.5	Pracovní výkazy.....	21
3.2	Stránka s pracovními výkazy.....	22
3.2.1	Stránka s odkazy na týdenní výkazy.....	22
3.2.2	Stránka s týdenním výkazem.....	23
3.2.3	Hashovací funkce.....	24
3.2.4	Hierarchie webových stránek.....	25
3.3	Stránky projektů v MediaWiki.....	26
3.3.1	Vkládání a aktualizace stránek v MediaWiki.....	26
3.4	Přístup k e-mailové schránce.....	27
3.4.1	Kontrola e-mailů a jejich zpracování.....	27
4	Implementace.....	29
4.1	Připojení k databázi NLPIS a přihlášení do MediaWiki.....	29
4.2	Generování a aktualizace wiki stránek.....	29
4.2.1	Aktualizace tabulky prostředků skupiny projektů a projektu.....	30
4.2.2	Titulek wiki stránky.....	31
4.2.3	Automaticky generovaná tabulka obsahu stránky.....	31
4.2.4	Wiki stránka skupiny projektů.....	32
4.2.5	Wiki stránka projektu.....	33
4.2.6	Wiki stránka s odkazy na výkazy.....	37
4.2.7	Wiki stránka s týdenním výkazem.....	39
4.2.8	Zpracování e-mailů.....	41
4.2.9	Odeslání e-mailu s chybou.....	44
4.2.10	Nastavení.....	44
4.2.11	Zabezpečení proti zneužití zdrojových souborů.....	44
4.2.12	Centrální ovládání aplikace.....	44
5	Testování a ladění.....	45
5.1	Založení stránek v MediaWiki.....	45
5.2	Aktualizace nově vytvořených stránek.....	46
5.3	Testování aktualizací jednotlivých částí stránek.....	46
5.3.1	Testy aktualizace stránky skupiny projektů.....	47
5.3.2	Testy aktualizace stránky projektu.....	47
5.3.3	Testy aktualizace stránky s odkazy na výkazy.....	49
5.3.4	Testy aktualizace stránky s týdenním výkazem.....	49
5.4	Výsledky testování.....	50
	Závěr.....	51
	Literatura.....	52

Přílohy	54
Seznam příloh:.....	55

Úvod

Ve své diplomové práci se budu zabývat vytvořením rozšíření pro stávající informační systém NLPIS (Informační systém výzkumné skupiny zpracování přirozeného jazyka), kde jsou uloženy rozpracované i hotové projekty řešitelů, kteří by měli ukládat informace o svých projektech do aplikace MediaWiki. Informačnímu systému NLPIS i aplikaci MediaWiki se budu v textu diplomové práce dále věnovat.

Cílem práce bylo generování a aktualizace wiki stránek, mezi které patří stránky skupin projektů, projektů, odkazy na výkazy a týdenní výkazy. V MediaWiki jsou již některé stránky skupin projektů a projektů vloženy a tak u nich proběhne pouze aktualizace, respektive vložení hlavičky ve správném formátu. Ostatní chybějící stránky budou vygenerovány. Dalším zadaným úkolem bylo zpracování e-mailů, respektive zjištění odpovědí (reakcí) a jejich vložení do týdenního výkazu.

Druhá kapitola je věnována problematice MediaWiki, která bude v diplomové práci využita jako datové úložiště. V informačním systému NLPIS jsou uloženy rozpracované i hotové projekty. Informace a aktuální stav řešení těchto projektů se ukládá do aplikace MediaWiki. Jedním z úkolů diplomové práce je kontrola obsahu webových stránek projektů v MediaWiki na výskyt základních údajů o projektu a jejich případné doplnění z databáze informačního systému NLPIS. Dále je třeba na stránku projektu přidat odkazy na další automaticky generované stránky, které budou obsahovat pracovní výkazy a plány uživatele.

Třetí kapitola se věnuje použitým technologiím, což jsou programovací jazyky HTML, JavaScript, CSS a PHP, a databáze MySQL. U jednotlivých jazyků jsem stručně zaznamenal oblasti použití a stručnou historii jejich vzniku. Dále jsem se v této kapitole zabýval syntaxí wiki, užívané v aplikaci MediaWiki včetně možností normalizace, a rozšířením PHP CURL, které jsem použil ke komunikaci s aplikací MediaWiki.

Čtvrtá kapitola práce je věnována návrhu implementace tohoto projektu. Její náplní je popis využití informačního systému NLPIS a aplikace MediaWiki. Podrobněji je zde rozebráno automatické generování pracovních výkazů z e-mailů a z informačního systému NLPIS. Nejprve jsem se zaměřil na tento problém jako celek a řešil některé problémy, které jsou s touto problematikou spojeny. Kromě návrhu jednotlivých wiki stránek jsem se zde také zabýval problematikou připojení k e-mailové schránce a získání jejího obsahu pomocí protokolu IMAP.

Pátá kapitola méj diplomové práce je věnována implementaci celé aplikace. Nejprve jsem se zde věnoval připojení k databázi informačního systému a přihlášení do aplikace MediaWiki. Dalším tématem byl podrobnější popis generování a aktualizace jednotlivých stránek a speciálně i jednotlivých funkcí MediaWiki, jako je vypnutí generování automatického titulku či obsahu na jednotlivých stránkách. Následně jsem se zabýval komponentami celé aplikace, jako je zpracování

e-mailů, odeslání e-mailu s chybou, nastavením, řešením zabezpečení proti nechtěnému čtení zdrojových kódů a poslední částí je ovládání skriptů z hlavní stránky programu.

Poslední část tohoto textu je věnována testování a ladění. Zde jsem umístil vyhodnocení některých základních testů jako je vygenerování stránek a uložení do MediaWiki, nebo aktualizace respektive kontrola vygenerovaných stránek. Následující část jsem věnoval provádění testů a aktualizování wiki stránek, do kterých jsem úmyslně vkládal chybné informace.

V závěrečné kapitole práce jsem zhodnotil dosažené výsledky a navrhl rozšíření do budoucna.

1 Současný stav

Vyvíjený program bude pracovat za podpory webového serveru, na kterém běží aplikace MediaWiki, kam zpracovatelé projektů postupně zapisují informace o řešených projektech.

1.1 MediaWiki

MediaWiki je webová aplikace, která slouží pro vytváření uživatelských webových stránek v jednoduchém textovém editoru, který podporuje HTML a speciální syntaxi wiki. Vytváří úložiště dat, která jsou vhodně strukturována a snadno editovatelná. MediaWiki je volně šiřitelný software, který byl původně vytvořen pouze pro internetovou encyklopedii Wikipedii. Její výhodou je snadné vkládání a úpravy těchto stránek a pravděpodobně i proto byla zvolena k ukládání informací o řešených projektech, uložených v databázi informačního systému NLPIS.

Jedná se o profesionální systém, který byl navržen pro práci s velkým množstvím aktivních uživatelů, kteří se systémem pracují, a webových stránek. Tato aplikace je právě díky své dobré výkonnosti a propracované struktuře stránek velmi oblíbená.

MediaWiki je napsána v jazyku PHP s využitím databáze MySQL nebo PostgreSQL. Původním autorem je Magnus Manske, když vytvořil první verzi tohoto systému 25. ledna 2002. Poslední dostupná stabilní verze je MediaWiki 1.16.0 a vyžaduje PHP minimálně ve verzi 5.1 nebo vyšší a databázi MySQL ve verzi 4.0 nebo vyšší nebo databázi PostgreSQL ve verzi 8.1 a vyšší. V projektu se bude pracovat s MediaWiki 1.15.5, která běží na serveru merlin.fit.vutbr.cz.

Systém nabízí velké množství funkcí:

- volitelné zobrazení obsahu pro dlouhé články
- rozšíření pro multimédia (nahrávání obrázků a zvuků, vkládání matematických vzorců pomocí jazyka LaTeX¹)
- sledování změn konkrétních článků (watchlist = seznam sledovaných položek (je možné sledovat vybrané články, které se do tohoto seznamu vloží, a podle něj bude systém uživatele upozorňovat na změny), zobrazení posledních článků, na kterých pracoval konkrétní uživatel, propracovaný systém změn)
- pro vkládání je možné využít syntaxi HTML v kombinaci se speciální syntaxí wiki
- lze editovat i malé části stránek
- systém je přeložen do mnoha jazyků včetně češtiny
- různý stupeň oprávnění pro anonymní uživatele, registrované uživatele, administrátory a vývojáře

¹ Program, který umožňuje autorům textů sázet a tisknout svá díla ve vysoké topografické kvalitě. [19]

- speciální stránky pro diskusi a široké možnosti přizpůsobení uživateli
- ...

[21], [22]

1.1.1 Nástroje pro import dat

Běžný způsob vkládání obsahu do MediaWiki je vkládat a zapisovat informace prostřednictvím uživatelského rozhraní, které kombinuje syntaxi wiki a HTML [21]. Problém ale nastává při potřebě automatického vkládání informací, což budu potřebovat pro implementaci rozšíření informačního systému NLPIS. Zde je možné využít množství nástrojů, které umožňují převádět obsah z vkládaného HTML a jiných formátů na značení wiki (přeloženo z anglického termínu „Wiki markup“), použité v aplikaci MediaWiki.

Aplikace MediaWiki obsahuje API pro správu dat. Pro import dat je možné využít funkce API:Edit, které slouží jak pro vkládání nových stránek, tak pro aktualizaci stávajících. Podrobnější popis použití API bude popsán v kapitole 3.

1.2 NLPIS

Informační systém NLPIS slouží ke správě projektů (jsou zde uloženy nezadané, rozpracované i hotové projekty), řešitelé zde zapisují svoje pracovní výkazy a upravují plán práce.

V informačním systému jsou uloženy informace o řešených projektech:

- řešitel projektu
- typ projektu,
- název,
- datum zadání (datum vložení od informačního systému),
- stav,
- vedoucí projektu,
- URL na stránku projektu v MediaWiki, uloženou na serveru merlin.fit.vutbr.cz.

Dále jsou v databázi uloženy informace o řešitelích projektu:

- datum zahájení řešení a datum ukončení (často nastaveno na nekonečno),
- aktivita,
- úkoly.

Tyto informace bude možné využít pro automatické generování wiki stránky v aplikaci MediaWiki.

1.2.1 Pracovní výkazy

Pracovní výkazy lze upravovat na samostatné stránce pojmenované pracovní výkazy. Pro jednotlivé projekty jsou pracovní výkazy oddělené a dělí se na denní a týdenní, přičemž denní jsou nedílnou součástí týdenních.

1.2.1.1 Týdenní pracovní výkaz

Tento výkaz obsahuje souhrn, plán, problémy a nejasnosti a omluvy a výmluvy. Souhrn je informace o práci provedené na projektu v průběhu celého týdne. Plán je stručná informace o činnostech na další týden. Problémy a nejasnosti slouží k upozornění vedoucího projektu na skutečnosti, které je třeba konzultovat a řešit. Omluvy a výmluvy slouží k uvedení omluv a výmluv.

Pokud jsou problémy zásadní a brání v dalším pokračování práce na projektu, je možné je ihned odeslat vedoucímu a případně dalším zainteresovaným osobám na projektu.

Pracovní výkaz musí řešitel odeslat vždy do pondělí následujícího týdne. Případně může odeslat výkaz dříve. Pokud není výkaz odeslán v určeném termínu, odešle systém řešiteli upozornění na neodeslaný výkaz v úterý. Pokud řešitel nereaguje, odešle se výkaz ve středu ráno automaticky.

1.2.1.2 Denní výkaz

Obsahuje datum, časový rozsah práce, počet odpracovaných hodin a popis činnosti. Řešitel tak může každý den uvést rozsah a náplň svojí práce na projektu. Denní výkaz slouží pro popis jednotlivých činností řešitele, a je proto možné vkládat více denních výkazů denně. Každý týden, kdy má řešitel naplánovanou nějakou činnost, musí zadat alespoň jeden denní pracovní výkaz.

1.2.2 Plán práce

Každý řešitel, který má odpracovat nějaké hodiny, má povinnost vytvořit si plán práce. Pokud nemá osoba naplánované všechny hodiny, které má odpracovat, naplánuje je při týdenní kontrole systém automaticky a informuje o naplánování řešitele e-mailem. Automaticky vytvořený plán je vhodné upravit a zohlednit tak například zkouškové období nebo Vánoce. Plán lze kdykoliv změnit a při neodpracování zadaného počtu hodin se plán práce automaticky upraví.

1.2.3 Projekty

Každý projekt v informačním systému má přiřazen kromě základních údajů také adresář projektu. Je možné vytvořit skupinu projektů, která umožňuje seskupit související projekty a následně přiřazovat příjemce zpráv celé této skupině. Dále je možné u skupiny projektů přiřazování prostředků z ostatních projektů ve skupině. Zprávy a prostředky budou popsány dále.

1.2.3.1 Řešené projekty

Ke každému projektu je dispozici seznam řešitelů. Každý řešitel má stanoven časový rozsah, ve kterém projekt řeší, a seznam úkolů, které má v rámci projektu splnit.

1.2.3.2 Prostředky projektu

Fyzické prostředky (adresáře, ke kterým potřebuje řešitel přístup) jsou obvykle umístěny na serveru minerva1. Každý prostředek má svoje jednoznačné identifikační číslo, název a vlastníka. Prostředky jsou přiřazovány k projektům s příslušnými oprávněními (tři možnosti: žádné, čtení nebo zápis). Každá osoba má k prostředku nejvyšší oprávnění, které potřebuje v rámci všech svých projektů. Vlastník prostředku má vždy oprávnění zápisu a vedoucí projektu má stejná oprávnění jako řešitelé. Příjemci zpráv mají oprávnění pro čtení.

1.2.4 E-mailové zprávy

Pro hromadné odesílání zpráv více osobám je nejprve nutné na stránce „výpis osob“ zvolit příjemce. Po výběru příjemců je nutné vyplnit předmět a text zprávy a případně seznam adresátů pro příjem kopie zprávy. Systém umožňuje zaslání zpráv jednotlivě (skrytí ostatních příjemců) nebo hromadně.

Systém automaticky rozesílá zprávu po vložení nové osoby do systému, dále při změně cesty k nějakému prostředku odešle zprávy všem osobám, které mají oprávnění k danému prostředku. Při změně úkolů je zaslána zpráva řešiteli a v kopii i všem zainteresovaným osobám na projektu. Všem zainteresovaným osobám je také odeslána zpráva při změně adresáře projektu, či při jeho přiřazení i k jiným projektům. Zprávy s výkazy a s tím související upozornění, která byla popsána výše. Zprávy o změnách plánu práce jsou zasílány všem osobám, kterým byl plán práce automaticky upraven. Zpráva je také zaslána řešitelům, kteří měli naplánovanou práci na předchozí týden, ale nevložili žádný pracovní výkaz, nebo pracovní výkaz neodeslali. Jednou týdně se odešle souhrnná zpráva vedoucímu projektu s přehledovými informacemi o hříšnicích (řešitelích, kteří do informačního systému nezasílali pracovní výkazy). Dále informační systém odešle zprávu vedoucímu, pokud některému řešiteli zbývá odpracovat malý počet hodin, pokud se u nějaké osoby blíží datum ukončení aktivity, nebo ukončení ředitelství posledního řešitele na projektu. Krajní případ tvoří řešitelé, kteří zadali poslední výkaz před více než 18 dny a měli na minulé dva týdny naplánovanou nějakou práci. Tito řešitelé obdrží varovný e-mail, a pokud na něj nereagují, odešle systém e-mail vedoucímu, který je povinen danou situaci řešit (návrh na zrušení účtu řešitele).

[4]

2 Vyžité technologie

Tato kapitola se věnuje použitým programovacím jazykům, které jsem si v mojí diplomové práci pro implementaci zvolil, a databázi MySQL. Jsou to jazyky obvykle používané pro implementaci dynamických webových aplikací s využitím databázového přístupu. Tyto technologie byly použity při implementaci informačního systému NLPIS a aplikace MediaWiki, a proto bude vhodné je pro implementaci rozšíření těchto aplikací využít. Dále jsem se v této kapitole věnoval syntaxi wiki i s možnostmi normalizace a rozšíření PHP CURL, včetně některých protokolů, které podporuje (HTTP, HTTPS, FTP). Závěr kapitoly jsem věnoval URL, hashovací funkci a protokolu IMAP, který jsem využil pro práci s e-mailovou schránkou.

2.1 Jazyk HTML

Hypertext Markup Language (dále jen HTML) je značkovací jazyk pro vytváření webových stránek. Dokumenty HTML mohou obsahovat hypertextové odkazy a pokročilejší formátování. Aktuálně používaná stabilní verze, přijatá konsorciem W3C (The World Wide Web Consortium), je verze 4.01.

2.1.1 Historie HTML

Značkovací jazyk HTML vznikl v roce 1990 jakožto verze HTML 1.0, ale v roce 1993 byl rozšířen verzí HTML + (obsahoval 78 elementů). Nová verze HTML 2.0, která byla podstatně zjednodušena (obsahovala 49 elementů) vznikla v roce 1994. Ta se jako první stala i uznávaným standardem. V roce 1995 se objevila nová verze HTML 3.0, což byl jakýsi kompromis mezi HTML + a HTML 2.0. Objevily se tu tabulky, matematické elementy, možnost formátování, formuláře. Přes verzi 3.2 v roce 1996 se plynule přešlo k nové verzi HTML 4.0, která se používá dnes. Tato verze umožňuje internacionalizace (používání zvláštních znaků různých abeced), úpravy tabulek a formulářů. Velký přínos byla podpora kaskádových stylů, na které se zaměříme níže. V prosinci 1999 byla vydána verze 4.01, která opravuje některé chyby předešlé verze. Předpokládalo se, že se jedná o poslední verzi před přechodem na XHTML. V březnu 2007 začaly práce na nové verzi HTML 5, které vede pracovní skupina WHATWG (Web Hypertext Application Technology Working Group). Základem specifikace nové verze jsou webové aplikace 1.0 a webové formuláře 2.0 ze specifikace pracovní skupiny WHATWG. Specifikace by měla být hotova v letech 2011-2012.

[7], [10], [14]

2.2 CSS - Kaskádové styly

Cascading Style Sheets (CSS) slouží k popisu prezentace dokumentů, aniž by jakkoli ovlivňovaly jejich obsah a strukturu. Kaskádové styly je možné vložit do externího souboru, který určuje vzhled elementů v HTML dokumentu. Další možností je vložení do hlavičky dokumentu nebo je možné psát formátovací značky přímo do těla dokumentu HTML. Vkládání přímo do dokumentu HTML ale může snižovat čitelnost a přehlednost kódu a navíc není tak snadné měnit definované styly jako při použití externího souboru.

CSS byl navržen organizací W3C. Autor prvotního návrhu byl Håkon Wium Lie. Publikoval jejich koncept v roce 1994. Prozatím byly vydány dvě verze – CSS1 a CSS2. Nyní se dokončuje verze CSS 2.1 a pracuje se na nové verzi CSS3.

Nevýhoda kaskádových stylů je různá reprezentace stejného CSS kódů v různých webových prohlížečích. Nástupem nových verzí prohlížečů se ale situace zlepšuje.

[1], [8], [10], [18]

2.3 Jazyk JavaScript

JavaScript je multiplatformní objektově orientovaný skriptovací jazyk používaný v prostředí WWW (World Wide Web), dále je možné jej využít například pro tvorbu aplikací. Skript v tomto jazyce je možné vkládat přímo do kódu HTML stránky nebo použít externí soubor. Obvykle jsou jím ovládány interaktivní prvky jako např. tlačítka nebo textová pole.

Syntaxe jazyka je podobná C/C++/Java. I když název napovídá, že jde o verzi Javy, není tomu tak a společně mají vedle podobného názvu jen podobnou syntaxi.

Program v JavaScriptu se obvykle spouští po stažení WWW stránky z Internetu = na straně klienta. Po stažení stránky z webu obvykle běží JavaScriptová aplikace v pískovišti (anglicky „sandbox“), aby nemohla provádět škodlivé operace. Je ale možné využívat doplňky prohlížeče (např. Firefox), které jsou napsány v tomto jazyce a umožňují i práci se soubory (ve webové aplikaci práce se soubory není možná z důvodu bezpečnosti).

[2], [3], [10], [17]

2.3.1 Historie JavaScriptu

Byl vyvinut v roce 1995 firmou Netscape autorem Brendanem Eichem nejdříve pod názvem Mocha, potom LiveScript a nakonec JavaScript. Firma Microsoft si jazyk upravila a používá název JScript.

V roce 1997 byl standardizován asociací ECMA (European Computer Manufacturers Association) a v srpnu 1998 i organizací ISO. Poslední verzí je 1.8.5 z července 2010.

[2], [10], [17]

2.4 Jazyk PHP

PHP je skriptovací jazyk určený především pro programování dynamických webových stránek. Používá se jako proveditelný kód, vložený do HTML stránek, nebo je možné použít opačný přístup – použití jazyka PHP jako generátor HTML stránky. Skripty jsou prováděny na straně serveru. Programátor má možnost s pomocí PHP se připojit k databázi (podpora mnoha typů databází, včetně MySQL, kterou ve své práci používám).

PHP se stalo velmi oblíbeným především díky jednoduchosti použití a tomu, že kombinuje vlastnosti více programovacích jazyků a nechává tak vývojáři částečnou svobodu v syntaxi. V kombinaci s operačním systémem Linux, databázovým systémem (obvykle MySQL nebo PostgreSQL) a webovým serverem Apache je často využíván k tvorbě webových aplikací. Pro tuto kombinaci se vžila zkratka LAMP – tedy spojení operační systém Linux, webový server Apache, databáze MySQL a skriptovací jazyk PHP nebo Perl. LAMP je tedy použitelné vývojové prostředí. Jazyk PHP je platformě nezávislý a je možné jej použít například i v operačním systému Windows nebo dokonce vytvářet multiplatformní aplikace.

[5], [6], [10], [27]

2.4.1 Historie PHP

PHP vyvinuli Rasmus Lerdorf (tvůrce), Zeev Suraski a Andi Gutmas v roce 1994 jako Personal Home Page. V polovině téhož roku se systém PHP spojil s programem Form Interpreter a tak vzniklo PHP/FI 2.0. V roce 1997 přepsali parser a zformovali tak základ PHP3. Tehdy byl také změněn název na PHP Hypertext Preprocessor, který je platný i dnes. Verze 3 vyšla v roce 1998. Byla rychlejší a obsahovala více funkcí. Navíc bylo možné tuto verzi spustit i na operačním systému Windows (předešlé verze fungovaly pouze na operačním systému Linux). Verze 4 vyšla v roce 2000.

Od roku 2004 se používá verze 5 s vylepšenou podporou objektově orientovaného programování, které se inspirovalo jazykem Java. Aktuálně nejnovější verze je PHP 5.3.3 z 22. července 2010.

[9], [10], [27], [28]

2.5 Databáze MySQL

MySQL je databázový systém, vytvořený švédskou firmou MySQL AB, kterou nyní vlastní společnost Oracle. Hlavní autoři jsou Michael Widenius a David Axmark. MySQL je multiplatformní databáze (lze provozovat na různých operačních systémech (Linux, Windows,...)). Spojení klient-server je optimalizováno především na rychlost i přes některá zjednodušení. Dalším kladem je možnost volného šíření databáze při použití v interním informačním systému nebo při dodání zdrojového kódu k dané distribuci aplikace (dle podmínek licence GPL, neboli General Public

Licence (česky všeobecná veřejná licence)). Díky těmto vlastnostem je MySQL jedna z nejvíce používaných databází.

Komunikace s databází probíhá pomocí jazyka SQL (Structured Query Language).

[6], [10], [24], [25]

2.5.1 Historie MySQL

Verze 3.23 poprvé podporovala cizí klíče a transakce v tabulkách typu InnoDB (formát úložiště dat v databázovém systému MySQL). Další verze 4.0 podporovala sjednocování dotazů pomocí příkazu UNION. Následující verze 4.01 podporovala různé znakové sady a porovnávání na úrovni databáze, tabulek nebo sloupců, časová pásma v datech nebo poddotazy. Verze 5.0 podporuje uložené procedury, trigger, pohledy, práci s metadaty, kurzory a INFORMATION.SCHEMA. Verze MySQL 5.1 přináší novou funkčnost jako časování událostí, replikaci na úrovni řádků či logování na straně serveru. Nejnovější verzí je nyní MySQL 5.5. V projektu bude využita verze 5.0.51, která je použita v aplikaci MediaWiki na serveru merlin.fit.vutbr.cz a verze 5.0.77, která je využita v informačním systému NLPIS na serveru nlp.fit.vutbr.cz. V současné době je plánovaná verze 6.x, která přinese možnost cizích klíčů i pro jiné než InnoDB tabulky a datové úložiště Falcon.

[24], [25]

2.6 Syntaxe wiki (wiki markup)

Wiki markup je značkovací jazyk, který se používá k psaní wiki webových stránek, jako je například Wikipedie². Jedná se o zjednodušenou alternativu značkovacího jazyka HTML. Výsledný kód napsaný ve wiki značkovacím jazyce je převeden do HTML (pomocí wiki software), který mohou zobrazit webové prohlížeče.

Neexistuje žádný běžně přijímaný standard tohoto jazyka, který by se ve všech wiki aplikacích používal. Gramatika, struktura, zarovnání nebo i klíčová slova závisí na konkrétní wiki aplikaci, která je použita na určité webové stránce. Příkladem může být vytváření odkazů na jiné stránky v rámci webu, pro které existuje mnoho různých konvencí. Wiki značkovací jazyky, přesněji dřívější verze, používaly „*CamelCase*“ k označení slov, která mají být automaticky odkazována. V aplikaci MediaWiki byla tato konvence nahrazena pomocí notace [[...]], kterou Wikipedie označuje jako „*free links*“ (volné odkazy).

Rozdílné wiki aplikace mohou podporovat používání různých množin elementů HTML v značkovacím jazyce wiki. V některých případech mohou být elementy HTML konfigurovány na wiki stránkách individuálně. V aplikaci MediaWiki je podporováno mnoho běžných značek HTML.

² URL české verze Wikipedie: http://cs.wikipedia.org/wiki/Hlavn%C3%AD_strana

2.6.1 Normalizace

Je snaha o vytvoření společného wiki značkovacího jazyka, který by měl být použit v různých variantách wiki aplikací. Existuje také několik wiki systémů, které toto značkování zavedly. Verze 1.0 této specifikace byla vydána v červenci 2007.

[32]

2.7 PHP – CURL (Client URL Library)

Knihovna libcurl vytvořená Danielem Stenbergem umožňuje spojení a komunikaci s mnoha různými typy serverů řadou protokolů, mezi které patří HTTP, HTTPS, FTP a další.

Knihovna libcurl bude v navrhovaném rozšíření sloužit k zasílání HTTP požadavků (HTTP Request) do aplikace MediaWiki a získávání odpovědí na ně. Požadavky budou zasílány na danou adresu s proměnnými metodou POST.

[27]

2.7.1 HTTP

HTTP (Hypertext Transfer Protocol) je internetový protokol určený pro výměnu hypertextových dokumentů ve formátu HTML. Používá obvykle port TCP/80, verze 1.1 protokolu je definována v RFC 2616. Tento protokol je spolu s elektronickou poštou tím nejvíce používaným a zasloužil se o obrovský rozmach internetu v posledních letech.

K protokolu HTTP existuje také jeho bezpečnější verze HTTPS, která umožňuje přenášet data šifrovat a tím chránit před odposlechem či jiným narušením.

Protokol funguje způsobem dotaz-odpověď. Uživatel (pomocí programu, obvykle webového prohlížeče) pošle serveru dotaz ve formě čistého textu, obsahujícího označení požadovaného dokumentu, informace o schopnostech prohlížeče apod. Server poté odpoví pomocí několika řádků textu popisujících výsledek dotazu (zda se dokument podařilo najít, jakého typu dokument je atd.), za kterými následují data samotného požadovaného dokumentu.

Pokud uživatel bude mít po chvíli další dotaz na stejný server (např. proto, že uživatel v dokumentu kliknul na hypertextový odkaz), bude se jednat o další, nezávislý dotaz a odpověď. Z hlediska serveru nelze poznat, jestli tento druhý dotaz jakkoli souvisí s předchozím. Kvůli této vlastnosti se protokolu HTTP říká bezstavový protokol – protokol neumí uchovávat stav komunikace, dotazy spolu nemají souvislost. Tato vlastnost je nepříjemná pro implementaci složitějších procesů přes HTTP (např. internetový obchod potřebuje uchovávat informaci o identitě zákazníka, o obsahu jeho „nákupního košíku“ apod.). K tomuto účelu byl protokol HTTP rozšířen o tzv. HTTP cookies, které umožňují serveru uchovávat si informace o stavu spojení na počítači uživatele.

[15]

2.7.2 HTTPS

HTTPS (Hypertext Transfer Protocol Secure) je nadstavba síťového protokolu HTTP, která umožňuje zabezpečit spojení mezi webovým prohlížečem a webovým serverem před odposloucháním, podvržením dat a umožňuje též ověřit identitu protistrany. HTTPS používá protokol HTTP, přičemž přenášená data jsou šifrována pomocí SSL³ nebo TLS⁴ a standardní port na straně serveru je 443.

Základy pro dnešní podobu HTTPS sahají do devadesátých let minulého století. Tehdy společnost Netscape Communications přišla s první verzí protokolu SSL, který vytvořila pro svůj webový prohlížeč. Tento protokol pak umožnil aplikačním protokolům možnost šifrovaného přenosu informací a ověření identity.

Protokol HTTPS využívá asymetrické šifrování. Obě strany si před zahájením komunikace vygenerují pár klíčů (privátní a veřejný). Při zahájení komunikace si vymění veřejné klíče, které by obě strany měly ověřit pomocí jiného komunikačního kanálu. Ověření může proběhnout kontrolou výtahu (otisk, miniatura, hash) veřejného klíče u protistrany například pomocí telefonu nebo lze použít princip přenosu důvěry, kdy nám protistrana předá veřejný klíč, který je digitálně podepsaný (nejlépe certifikační autoritou, které důvěřujeme a jejíž veřejný klíč máme v důvěryhodném úložišti). Digitální certifikáty jsou základním kamenem zabezpečení poskytovaného protokoly SSL/TLS.

[13]

2.7.3 FTP

FTP (anglicky File Transfer Protocol) je protokol pro přenos souborů mezi počítači pomocí počítačové sítě. Využívá protokol TCP z rodiny TCP/IP a může být požíván nezávisle na použitém operačním systému (je platformě nezávislý). Definován byl v roce 1985 v RFC 959 a rozšířen byl v roce 1997 v RFC 2228. Jeho podpora je součástí webových prohlížečů nebo specializovaných programů (tzv. FTP klientů).

FTP je jeden z nejstarších protokolů, využívá porty TCP/21 a TCP/20. Port 21 slouží k řízení a jsou jím také přenášeny příkazy FTP. Port 20 slouží k vlastnímu přenosu dat, který je 8bitový. Přenos může být *binární* nebo *ascii* (textový). Při textovém přenosu dochází ke konverzi konců

³ Secure Sockets Layer, SSL (doslova vrstva bezpečných socketů) je protokol, resp. vrstva vložená mezi vrstvu transportní (např. TCP/IP) a aplikační (např. HTTP), která poskytuje zabezpečení komunikace šifrováním a autentizací komunikujících stran. Více informací na [29].

⁴ Protokol Transport Layer Security (TLS) a jeho předchůdce, Secure Sockets Layer (SSL), jsou kryptografické protokoly, poskytující možnost zabezpečené komunikace na Internetu pro služby jako WWW, elektronická pošta, internetový fax a další datové přenosy. Více informací na [30].

řádků-CR/LF (DOS, Microsoft Windows) nebo jen LF (unixové systémy), pokud jsou koncové systémy rozdílné. Při binárním přenosu není do dat nijak zasahováno.

Protokol je interaktivní a umožňuje řízení přístupu (přihlašování login/heslo), specifikaci formátu přenášeného souboru (znakově - binárně), výpis vzdáleného adresáře atd. V současné době už není považován za bezpečný a z tohoto důvodu pro něj byla definována některá rozšíření (RFC 2228).

V protokolu je použit model klient-server. FTP server poskytuje data pro ostatní počítače. Klient se k serveru připojí a může provádět různé operace (výpis adresáře, změna adresáře, přenos dat atd.). Operace jsou řízeny sadou příkazů, které jsou definovány v rámci FTP protokolu, proto kdokoli může vytvořit klienta pro jakékoliv prostředí nebo operační systém. Existuje mnoho programů pro FTP servery i klienty a mnoho je jich volně dostupných.

[12]

2.8 URL (Uniform Resource Locator)

URL, celým názvem Uniform Resource Locator („jednotný lokátor zdrojů“) je řetězec znaků s definovanou strukturou, který slouží k přesné specifikaci umístění zdrojů informací (ve smyslu dokument nebo služba) na Internetu.

URL definuje doménovou adresu serveru, umístění zdroje na serveru a protokol, kterým je možné zdroj zpřístupnit. Jednotlivá pole v URL jsou: schéma, doménové jméno, port, specifikace souboru, parametry.

Některá pole jsou nepovinná – buď nemají význam, nebo se předpokládá předdefinovaná hodnota, závislá např. na schématu (např. pro protokol HTTP je implicitní port 80), nebo na aplikaci (pro webový prohlížeč se předpokládá protokol HTTP).

[31]

2.9 Hashovací funkce

Message-Digest algorithm je rozšířená rodina hašovacích funkcí, která vytváří ze vstupních dat výstup (otisk) fixní délky. Otisk je též označován jako miniatura, kontrolní součet (v zásadě nesprávné označení), fingerprint, hash (česky někdy psán i jako haš). Jeho hlavní vlastností je, že malá změna na vstupu vede k velké změně na výstupu, tj. k vytvoření zásadně odlišného otisku.

Algoritmus MD5 se prosadil do mnoha aplikací (např. pro kontrolu integrity souborů nebo ukládání hesel). MD5 je popsán v internetovém standardu RFC 1321 a vytváří otisk o velikosti 128 bitů. Byl vytvořen v roce 1991 Ronaldem Rivestem, aby nahradil dřívější hašovací funkci MD4.

MD5 se často používá pro ukládání hesel. Přidáním soli (několik pseudonáhodných znaků) k heslu se ztěžují útoky na získání hesla a účinnost slovníkových útoků s využitím předpočítaných tabulek (anglicky rainbow table attack). Tento postup lze použít i pro jakékoli kryptografické

hašovací funkce, které lze dopředu "předpočítat", jejich výsledný haš uložit a pak rychleji vyhledat v databázi srovnáním s hašem, na který se útočí.

[23]

V mojí diplomové práci jsem funkci MD5 nepoužil pro ukládání hesel, ale jako kontrolní výpočet pro zjištění změny ve vkládaném textu.

2.10 IMAP (Internet Message Access Protocol)

IMAP (*Internet Message Access Protocol*) byl vyvinut Markem Crispinem v roce 1986 jako vzdálený poštovní protokol pro prohlížení obsahu e-mailové schránky. Je to internetový protokol pro vzdálený přístup k e-mailové schránce. Na rozdíl od protokolu POP3 umí IMAP pracovat v tzv. on-line i off-line režimu a nabízí pokročilé možnosti vzdálené správy (práce se složkami, přesouvání zpráv, prohledávání na straně serveru a podobně). V současné době se používá protokol IMAP4 (IMAP version 4 revision 1 - IMAP4rev1), který je definován v RFC 3501.

Protokol IMAP umožňuje trvalé (tzv. on-line) připojení k e-mailové schránce. Díky tomu je možné s celou poštovní schránkou plně pracovat z libovolného místa. Všechny zprávy a složky jsou uloženy na poštovním serveru a na počítač se stahují jen nezbytné informace, takže při zobrazení složky se stáhnou jen záhlaví zpráv a jejich obsah až v případě, že zprávu chce uživatel přečíst. U jednotlivých zpráv se uchovává jejich stav (nepřečtená, odpovězena, důležitá), uživatel může zprávy přesouvat mezi složkami, složky vytvářet, mazat, prohledávat na straně serveru apod. Protokol umožňuje současné připojení více klientů zároveň.

[16]

Ve svém programu jsem tento protokol využil ke čtení zpráv z e-mailové schránky, do které budou posílány kopie zpráv z informačního systému NLPIS a reakce na ně.

3 Návrh implementace

V této kapitole se budu zabývat návrhem implementace celé aplikace, včetně vzhledu automaticky generovaných webových stránek pro aplikaci MediaWiki. Nejprve jsem se zaměřil na obecné problémy u některých stránek a následně jsem se postupně blíže podíval na jednotlivé generované stránky, které budou pro skupiny projektů, projekty, odkazy na pracovní výkazy, týdenní výkazy a v neposlední řadě stránky s e-maily. Pro zpracování e-mailů jsem navrhl použití protokolu IMAP a popsal jsem připojení ke schránce a postupné získání a procházení jejího obsahu.

3.1 Návrh automaticky generovaných stránek

Tato část je věnována návrhu automaticky generovaných stránek. Jsou zde popsány stránky s pracovními výkazy a stránka projektu pro jednoho i více řešitelů.

Vyvíjená aplikace bude cyklicky (jednou denně) zjišťovat stav wiki stránek a porovnávat je s informačním systémem. Pokud stránka v aplikaci MediaWiki dosud neexistuje, vygeneruje se automaticky nová stránka s využitím dat z informačního systému NLPIS. Pokud stránka v aplikaci MediaWiki již je, provede se kontrola na aktuálnost dat a případná aktualizace.

3.1.1 Aktualizace stránek skupiny projektů a projektu

První kontrolou bude, zda se na stránce vyskytuje ohraničení (začátek a konec) generované hlavičky. Pokud se zde ohraničení nebude vyskytovat, vygeneruje se na stránku nová hlavička. Speciální případ tvoří projekty, které mají stejná URL. Tyto stránky projektu budou tvořeny jednou wiki stránkou a jednotlivé hlavičky budou vypsány pod sebou.

Aktualizace stránek skupin projektů a projektů bude spočívat ve vyhledávání klíčových slov a jejich obsahu, který se bude porovnávat s požadovaným. Pokud se nebude obsah stránky s daty v databázi shodovat, vytvoří se poznámka k danému klíčovému slovu, do které se aktuální obsah přesune a za klíčovým slovem se nahradí správnými daty z databáze.

3.1.2 Prostředky skupiny projektů a projektu

Prostředky skupiny projektů a projektu budou na stránku umístěny ve formě tabulky. V této tabulce budou jednotlivé řádky s názvem prostředku, cestou, oprávněním a informací o tom, ke kterému projektu prostředek patří.

3.1.2.1 Aktualizace prostředků skupiny projektů a projektu

Kontrola aktuálnosti tabulky s prostředky bude probíhat tak, že se budou postupně procházet jednotlivé řádky, které se budou porovnávat s požadovaným obsahem. V tabulce se řádky nebudou

mazat, ale bude možné řádky přesunovat podle obsahu databáze. Dále pokud bude prostředek skupině nebo projektu odebrán, změní se oprávnění na „žádné“ a prostředek se přesune na konec tabulky.

Pokud měl prostředek dříve práva na „zápis“ a nyní přechází na „čtení“ nebo „žádné“, připíše se za tento typ oprávnění ještě znak „*“. Tímto bude také uchováována historie přidělených prostředků, která se může hodit při zjišťování, kdo je mohl měnit.

3.1.3 Stránka skupiny projektů

Skupiny projektů v informačním systému NLPIS prozatím nejsou zcela dořešeny, ale ve vyvíjené aplikaci s nimi bude již počítáno a budou se vytvářet a aktualizovat jejich stránky.

Na této stránce budou údaje jako název skupiny, vedoucí, prostředky skupiny a projekty, které do této skupiny patří (jak je vidět na obrázku Obrázek 1).

Název skupiny				
Vedoucí skupiny: titul, jméno, příjmení vedoucího skupiny				
Prostředky skupiny:				
<table border="1"><tr><td>název</td><td>cesta</td><td>oprávnění</td><td>patří k projektu</td></tr></table>	název	cesta	oprávnění	patří k projektu
název	cesta	oprávnění	patří k projektu	
...v tabulce budou pod sebou umístěny jednotlivé prostředky skupiny				
Projekty skupiny:				
...zde budou jednotlivé projekty skupiny s odkazem na stránku příslušného projektu				

Obrázek 1 - Stránka skupiny projektů

3.1.3.1 Aktualizace stránky skupiny projektů

Klíčová slova jako název skupiny a vedoucí se budou aktualizovat výše popsaným způsobem. Projekty skupiny budou procházeny po řádcích a budou se porovnávat s obsahem databáze. Pokud bude nalezena nesrovnalost, tak se aplikace pokusí nalézt projekt na dalších řádcích. Projekty budou pod sebou umístěny dle jejich stavu v databázi. Nejprve budou vypsány řešené projekty, potom k rozhodnutí, ukončené a na konec nezadané. Jestliže bude název projektu na jiném řádku, tak se automaticky přesune na správné místo. Pokud bude v seznamu odebraný projekt, vloží se do poznámky.

3.1.4 Stránka projektu

Na stránce projektu budou klíčová slova název, řešitel, vedoucí, zahájení projektu, ukončení řešení, typ projektu, skupina, prostředky projektu a úkoly řešitele. Úkoly budou na stránku projektu vkládány, pouze pokud bude mít projekt právě jednoho řešitele. Pokud bude projekt řešit více řešitelů, bude pro jejich úkoly vytvořena samostatná stránka. Datum zahájení projektu bude na stránce umístěno vždy, ale datum ukončení řešení se na stránku umístí, pouze pokud projekt již není řešen

žádným řešitelem (jinak se za toto klíčové slovo vloží znak „-“). Projekt nemusí patřit do žádné skupiny. V takovém případě bude za klíčovým slovem „Skupina“ místo odkazu na konkrétní stránku pouze pomlčka.

[editovat]

Automatické generování wiki stránek z e-mailů a IS

Řešitel: Bc. Ales Vavrinek
 Vedoucí: Ing. Jaroslav Dytrych
 Zahájení projektu: 10. 06. 2010
 Ukončení řešení: 01. 07. 2011
 Typ projektu: Diplomová práce
 Skupina: [misc](#)
 Prostředky:

název	cesta	oprávnění	patří k projektu
wiki_z_nlpis	/mnt/minerva1/nlp/projects/wiki_z_nlpis	zápis	Automatické generování wiki stránek z e-mailů a IS

Poznámka k oprávnění s *: dříve bylo oprávnění zápis

[Pracovní výkazy](#)

Úkoly:

1. Seznamte se s jazyky a prostředky pro tvorbu webových informačních systémů
2. Prostudujte MediaWiki a informační systém skupiny zpracování přirozeného jazyka (NLPIS)
3. Navrhněte rozšíření NLPIS a další potřebné skripty pro automatické generování stránek projektů a skupin projektů z NLPIS
4. Navrhněte skripty pro generování wiki stránek z přijatých e-mailů (vyhrazená schránka, jejíž obsah se bude automaticky zpracovávat) tak, aby vytvářely stránky s různými informacemi o projektech se vzájemnými odkazy (řešitelé projektu a jejich úkoly, řešené problémy, historie změn prostředků apod.)
5. Implementujte navržená řešení
6. Zhodnotte dosažené výsledky, srovnajte zvolené přístupy s alternativními metodami a navrhněte další možná rozšíření a vylepšení do budoucna

Obrázek 2 - Automaticky generovaná stránka projektu pro jednoho řešitele

3.1.4.1 Stránka projektu pro více řešitelů

Pokud bude mít projekt více řešitelů, bude za jménem řešitele generován odkaz na samostatnou stránku s úkoly pro daného řešitele. Dále bude u každého řešitele zapsáno, od kdy projekt řeší, případně kdy řešení ukončil. Datum zahájení projektu bude spočteno jako minimum z dat zahájení řešení všech řešitelů. Podobně datum ukončení řešení bude spočteno jako maximum z dat ukončení řešení (v případě, že je projekt stále řešen, bude místo data zapsán znak „-“).

3.1.4.2 Stránka projektu bez řešitele

Tato situace může nastat při zakládání nových projektů, kdy řešitel ještě není přiřazen, ale protože je již založen projekt, musí pro něj existovat stránka v aplikaci MediaWiki. Jinou variantou může být ukončení řešení přiděleným řešitelem, který na projektu vůbec nezačal pracovat.

3.1.4.3 Změna počtu řešitelů

Bude-li v databázi informačního systému zadán nový projekt, bez přiděleného řešitele, bude rozvržení a aktualizace řešena jako projekt s jedním řešitelem. Taktéž jestliže měl projekt jednoho řešitele a přechází na stav bez řešitele.

V případě, že měl projekt někdy přiděleno 2 a více řešitelů, bude aktualizace řešena stejně jako projekt s více řešiteli a to i v případě, že projekt přechází do stavu, kdy je řešen pouze jedním či žádným řešitelem.

Převod slovníků do formátu LMF [editovat]

Řešitelé:

- Filip Kocina ([Úkoly řešitele](#)), řešil: 29. 10. 2009 - 19. 05. 2010
- Ondrej Simicek ([Úkoly řešitele](#)), řešil: 19. 05. 2010 - 28. 02. 2011

Vedoucí: doc. RNDr., Ph.D. Pavel Smrz

Zahájení projektu: 29. 10. 2009

Ukončení řešení: 28. 02. 2011

Typ projektu: Obecný projekt

Skupina: [dict](#)

Prostředky:

název	cesta	oprávnění	patří k projektu
dicts2lmf	/mnt/minerva1/nlp/projects/dicts2lmf	čtení	Převod slovníků do formátu LMF
dicts2lmf3	/mnt/minerva1/nlp/projects/dicts2lmf3	zápis	Převod slovníků do formátu LMF

*Poznámka k oprávnění s *: dříve bylo oprávnění zápis*

[Pracovní výkazy](#)

Obrázek 3 - Stránka projektu pro více řešitelů

3.1.5 Pracovní výkazy

Řešitelé projektů pravidelně vytváří pracovní výkazy v informačním systému NLPIS. Informační systém sám hlídá zápis těchto výkazů do databáze a upozorňuje řešitele, pokud týdenní výkaz nezadal. Respektive zasílá e-mail s požadavkem o doplnění informací nejdříve řešiteli, a pokud ten nereaguje, pošle zprávu vedoucímu projektu.

Denní výkazy slouží k rozeepsání jednotlivých činností za odpracovaný den. Denní výkazy jsou nedílnou součástí týdenního výkazu. Týdenní výkaz obsahuje souhrn, plán, problémy a nejasnosti a omluvy a výmluvy. Problémy a nejasnosti může řešitel odeslat kdykoliv v týdnu a přijdou e-mailem všem zainteresovaným osobám a vedoucímu projektu. Při odeslání problému se odešle celý výkaz (včetně denních výkazů), aby mohli ostatní na problém lépe reagovat a viděli, co řešitel na projektu již udělal. Posledním typem jsou omluvy a výmluvy, které může uživatel v týdenním výkazu vyplnit, pokud například neměl čas řešit přiřazený projekt z různých důvodů (například z důvodu nemoci).

Nejprve se použije vyvíjená aplikace pro přenos některých dat z databáze informačního systému NLPIS do aplikace MediaWiki. V MediaWiki se budou data ukládat ke každému projektu na zvláštní webovou stránku. Na stránce projektu bude automaticky vygenerován odkaz na stránku s pracovními výkazy, které bude možné prohlížet. Odkaz ze stránky projektu povede na stránku

s odkazy na výkazy, kde budou pro každý týden v případě činnosti řešitele uvedeny týdenní výkazy, respektive hlavička týdne, jméno řešitele, odpracované hodiny a část souhrnu. Jméno řešitele bude zároveň tvořit odkaz na příslušný kompletní týdenní výkaz, kde budou umístěny informace jako souhrn, plán nebo jednotlivé denní výkazy. Pokud se v daném týdnu řešil přes informační systém nějaký problém (problémy a nejasnosti), budou k tomuto výkazu též přidány i s odpověďmi na daný problém. Pokud se bude uživatel omlouvat z práce (omluvy a výmluvy), nemá smysl tyto interní informace na webovou stránku vkládat.

Další použití aplikace bude tvořit druhý hlavní úkol zadání – generování wiki stránek pomocí e-mailů. Pro tuto část bude potřeba založit e-mailovou schránku, do které budou přicházet kopie veškeré odchozí komunikace z NLPIS a dále kopie komunikace mezi řešiteli a mezi řešiteli a vedoucím projektu či dalšími zainteresovanými osobami (včetně reakcí). Informační systém odesílá e-mailem týdenní výkazy (během týdne případně i řešené problémy) řešitelům (i sobě - volitelné), vedoucímu projektu a dalším zainteresovaným osobám. Zde bude potřeba pouze nastavit jako jednoho z příjemců novou speciální e-mailovou schránku. Vyvíjená aplikace bude procházet všechny přijaté e-maily a systematicky generovat do MediaWiki dané informace.

3.2 Stránka s pracovními výkazy

V aplikaci MediaWiki bude k webové stránce projektu automaticky generován odkaz na stránku s pracovními výkazy. První stránka bude obsahovat pouze rozpis týdnů s odpracovanými hodinami. Po výběru některého týdne se po kliknutí přejde na další stránku, kde bude vybraný týdenní výkaz s rozepsanými denními výkazy. Nahoře budou odkazy, které budou odkazovat na jednotlivé denní výkazy na této stránce.

3.2.1 Stránka s odkazy na týdenní výkazy

Navrhl jsem tuto stránku rozdělenou na jednotlivé týdny. U každého týdne budou pro lepší orientaci uvedeny data od (pondělí) – do (neděle). Na následujícím řádku bude jméno řešitele následované počtem odpracovaných hodin v daném týdnu. Pod jménem bude zobrazeno několik počátečních slov (bude nastavitelné) souhrnu.

Po kliknutí na jméno řešitele se otevře stránka s vybraným týdenním výkazem. Texty omluv a výmluv by nebylo vhodné uvádět, protože tuto stránku mohou prohlížet všichni registrovaní uživatelé.

Pokud bude mít projekt více řešitelů, budou jednotliví řešitelé vypsáni pod sebou v každém týdnu a každý tak bude mít svoje stránky s týdenními výkazy. Všichni řešitelé budou mít za jménem počet odpracovaných hodin a na dalším řádku bude nastavitelný počet slov ze souhrnu týdenního výkazu.

Jak je možné vidět na obrázku (Obrázek 4), v záhlaví stránky je umístěn nadpis tvořený zpětným odkazem na stránku projektu a dále nadpisem stránky.

Pro možnosti aktualizace bez vytváření zbytečných poznámek (historie odpracovaných hodin či souhrnu, jenž se během týdne může měnit), které zde nejsou nijak důležité, jsem zvolil využití hashovací funkce MD5 (viz kapitola 2.9), již bude spočten hash z odpracovaných hodin a souhrnu řešitele. Poznámky se tak budou vytvářet pouze pro texty upravené přímo ve wiki stránce.

Automatické generování wiki stránek z e-mailů a IS - odkazy na pracovní výkazy	[editovat]
Rok 2010, 48. týden (29. 11. 2010 - 05. 12. 2010)	[editovat]
Řešitel Bc. Ales Vavrinek, odpracováno 5 hodin, 0 minut Souhrn: Několik počátečních slov z výkazu řešitele ...	
Rok 2010, 47. týden (22. 11. 2010 - 28. 11. 2010)	[editovat]
Řešitel Bc. Ales Vavrinek, odpracováno 0 hodin, 0 minut Souhrn: Několik počátečních slov z výkazu řešitele ...	
Rok 2010, 46. týden (15. 11. 2010 - 21. 11. 2010)	[editovat]
Řešitel Bc. Ales Vavrinek, odpracováno 3 hodin, 0 minut Souhrn: Několik počátečních slov z výkazu řešitele ...	
Rok 2010, 45. týden (08. 11. 2010 - 14. 11. 2010)	[editovat]
Řešitel Bc. Ales Vavrinek, odpracováno 3 hodin, 0 minut Souhrn: Několik počátečních slov z výkazu řešitele ...	

Obrázek 4 - Stránka s odkazy na týdenní výkazy pro jednoho řešitele

3.2.2 Stránka s týdenním výkazem

Na této stránce bude zobrazen vybraný týdenní výkaz - souhrn aktivit řešitele, plán na příští týden, problémy a nejasnosti, denní výkazy. Pokud řešitel do výkazu zadá problémy a nejasnosti, budou zde umístěny spolu s nadpisem „Reakce na problémy a nejasnosti“, pod který se bude automaticky vkládat odpověď z e-mailu. U reakce na daný problém bude automaticky z e-mailu vybrána část s odpovědí. Aby bylo možné zjistit další souvislosti, nebo v případě chybného automatického výběru, bude za textem odpovědi odkaz na stránku s celým textem e-mailu. Informace o odpovědi lze od dotazu rozpoznat pomocí speciální značky (nejčastěji „>“) na začátku každého řádku. Před denními výkazy bude na stránce umístěn ještě nadpis „Reakce na výkaz“. Pod tento nadpis se budou vkládat e-maily, které budou na daný výkaz reagovat.

Vkládané e-maily budou ve tvaru: datum, odesílatel, reakce na výkaz (nebo problém), odkaz na celý text e-mailu.

Informace budou generovány pravděpodobně každý večer. Po upravení výkazu řešitelem nebo po změně dat v databázi, budou informace automaticky pozměněny.

Další část této stránky bude věnována denním výkazům, které budou umístěny ve spodní části. Odkazy na denní výkazy v horní části stránky budou generovány pouze na ty dny v daném týdnu, kdy řešitel zapíše svoji činnost do informačního systému NLPIS.

V horní části se bude nacházet nadpis tvořený odkazem na stránku projektu, odkazem na stránku s odkazy na výkazy a samotným nadpisem vybraného výkazu.

Pro možnosti aktualizace jednotlivých částí stránky se bude využívat výpočtů hashovací funkce MD5 (viz kapitola 2.9). Tato funkce bude využita k porovnání dříve vloženého textu z databáze informačního systému NLPIS a aktuálního obsahu stránky. Bude vhodné počítat hash pro souhrn, plán, problémy i jednotlivé denní výkazy. V případě, že bude stávající text odlišný od požadovaného, ale výpočet MD5 se od uloženého nebude lišit, může dojít k nahrazení neplatného obsahu. Jestliže se ale bude hash lišit, vloží se text do poznámky a vygeneruje se nový.

Automatické generování wiki stránek z e-mailů a IS - Pracovní výkazy - Rok 2010, 48. týden, výkaz řešitele Bc. Ales Vavrinek
Týdenní výkaz řešitele Bc. Ales Vavrinek za 48. týden (29. 11. 2010 - 05. 12. 2010), celkem odpracováno 6 hodin, 0 minut
Denní výkazy: Pondělí Úterý Středa Čtvrtek Pátek Sobota Neděle
Souhrn: ... zde bude umístěn text souhrnu daného řešitele
Plán na příští týden: ... zde bude umístěn text plánu daného řešitele
Problémy a nejasnosti: ... zde bude text problémů řešitele
Reakce na problémy a nejasnosti: ... zde budou umístěny reakce na dané problémy
Reakce na výkaz: ... zde budou umístěny reakce na daný výkaz
Pondělí 29. 11. 2010, odpracováno 1 hodin, 0 minut) ... zde bude umístěn text odpracované činnosti
Úterý 30. 11. 2010, odpracováno 2 hodin, 0 minut) ... zde bude umístěn text odpracované činnosti

Obrázek 5 - Týdenní výkaz

3.2.3 Hashovací funkce

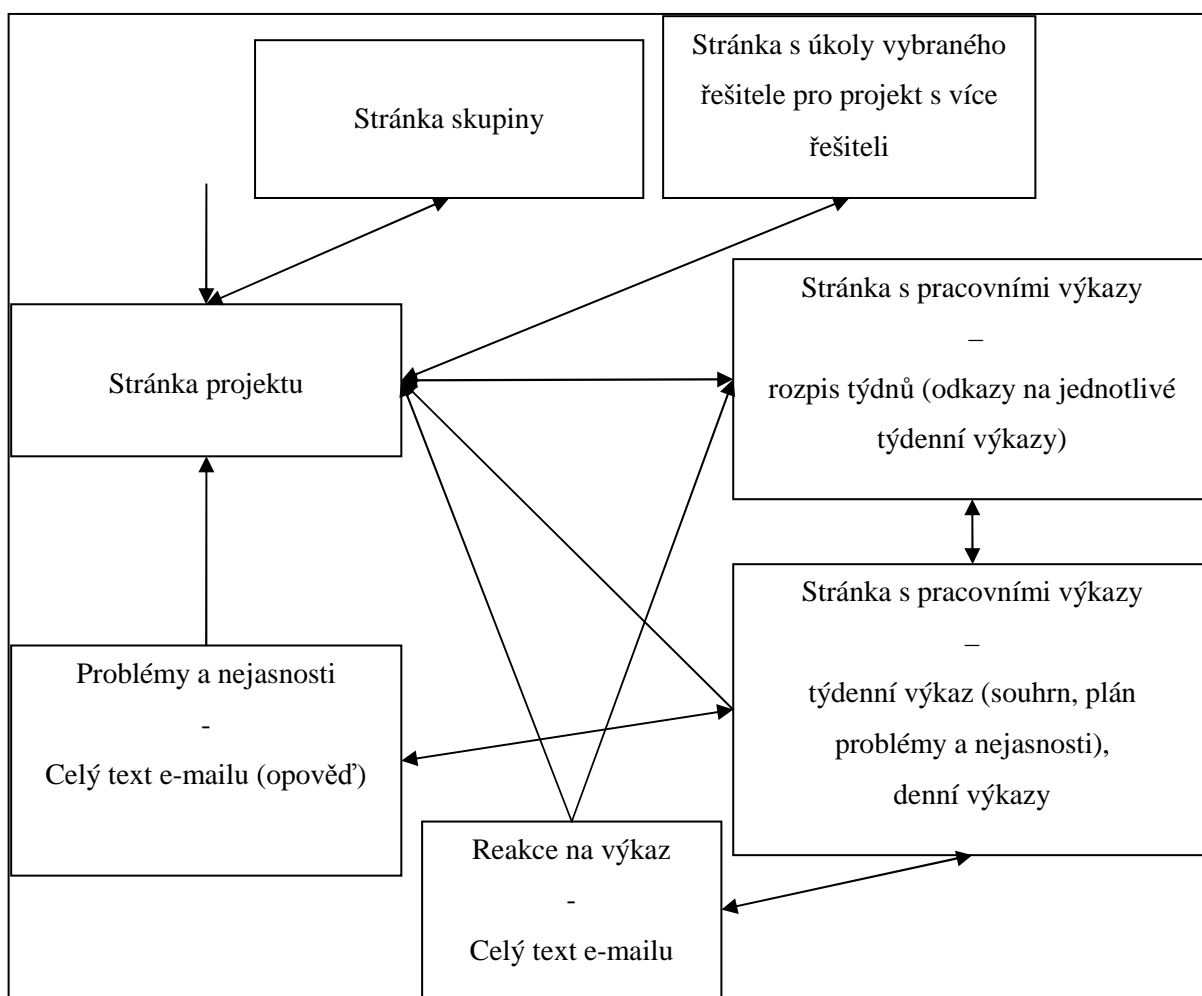
Pro stránky jako odkazy na výkazy a týdenní výkazy, kdy není příliš důležitá historie zadaných údajů, ale zajímá nás spíše aktuální stav z databáze informačního systému, jsem tento výpočet využil ke zjištění, zda se jedná o dříve vložený text, nebo zda je jeho součástí i nějaký komentář uživatele. Je-li text na stránce od požadovaného odlišný, kontroluje se s uloženým

výsledkem hashe a při shodě je možné ho nahradit. Naopak při neshodě uloženého hashe a stávajícího textu se vloží starý text do poznámky, aby nedošlo ke ztrátě informací, dopsaných uživatelem přímo do wiki stránky, a následně je vložen nový obsah s odpovídajícím výpočtem MD5.

3.2.4 Hierarchie webových stránek

Ze stránky projektu v aplikaci MediaWiki bude možné přejít na stránku skupiny projektů, ke které daný projekt patří (pokud k nějaké skupině patří). Pokud bude mít projekt více řešitelů, bude možné přejít na stránku s úkoly každého řešitele. Dále bude možné přejít na stránku s odkazy na jednotlivé týdenní výkazy. Zde se po kliknutí na jméno řešitele přejde na stránku s vybraným týdenním výkazem daného řešitele.

Stránky budou hierarchicky propojeny a bude možné se jednoduše a rychle „proklikat“ k požadovaným informacím. Stránky budou provázány odkazy tak, aby se dalo snadno přecházet mezi jednotlivými úrovněmi zanoření, jak je vidět na následujícím obrázku (Obrázek 6).



Obrázek 6 – Hierarchie stránek v aplikaci MediaWiki

3.3 Stránky projektů v MediaWiki

V aplikaci MediaWiki jsou uloženy webové stránky všech řešených a hotových projektů, které jsou uloženy v informačním systému NLPIS. Všichni řešitelé projektů mají povinnost udržovat tyto stránky aktuální a doplňovat do nich aktuální informace o řešeném projektu a současném postupu.

Do této aplikace lze vkládat data přes nabízené webové rozhraní, což u automatického generování ale nebude možné. Pro automatické generování a aktualizaci wiki stránek bude třeba využít dostupné API.

3.3.1 Vkládání a aktualizace stránek v MediaWiki

Pokud budeme chtít pracovat s MediaWiki a vkládat nebo editovat wiki stránky, je třeba se nejprve přihlásit. K tomuto účelu slouží rozhraní API:Login. Jak jsem již zmínil v kapitole 1.1.1, v aplikaci MediaWiki existuje rozhraní pro správu dat. K účelu vkládání a aktualizace wiki stránek bude dobré využít API:Edit. Dále bude ale ještě potřeba získávat aktuální obsah wiki stránek, který se dá získat přes rozhraní API:Query.

3.3.1.1 MediaWiki API:Login

Toto rozhraní slouží k přihlášení do aplikace MediaWiki. Použití je poměrně jednoduché. Nejprve se provede dotaz do MediaWiki, kam jsou zaslány přihlašovací údaje, a je získán přihlašovací token. Pokud ale budou přihlašovací údaje chybné, přihlašovací token nezískáme.

Ve druhém průchodu se opět odešlou přihlašovací údaje, ale k nim je ještě přidán získaný token. Nyní zjistíme, zda bylo přihlášení úspěšné.

Zasílání dotazu do MediaWiki a získání odpovědi bude realizováno pomocí modulu PHP CURL (viz kapitola 2.7). Tento modul umožňuje komunikaci přes HTTP, případně HTTPS, která je pro přihlášení a editovací funkce požadována.

3.3.1.2 MediaWiki API:Query

API:Query slouží k zasílání požadavků do MediaWiki, přes které je možné získat například obsah wiki stránky nebo editovací token (viz dále). V aplikaci MediaWiki jsou stránky identifikovány titulkem.

K získání obsahu hledané stránky se tedy použije v dotazu titulek hledané stránky. Aplikace buď odpoví identifikačním číslem stránky, nebo chybovým hlášením, že stránka neexistuje.

Dalším využitím API:Query bude získání editovacího tokenu, který slouží k vytváření a editaci wiki stránek. Dotaz bude obsahovat opět titulek požadované wiki stránky. MediaWiki odpoví vždy editovacím tokenem a to pro existující i neexistující wiki stránky. Editovací token slouží jako identifikátor pro editaci a vytváření wiki stránek.

3.3.1.3 MediaWiki API:Edit

Pro práci s tímto API je třeba vždy nejprve získat tzv. „edit token“ (editovací token). K získání tohoto tokenu bude třeba využít rozhraní API:Query a modul PHP CURL (viz kapitola 2.7).

Nejprve se tedy zašle požadavek do MediaWiki na získání editovacího tokenu pro vybranou wiki stránku. Druhým krokem bude aktualizace nebo vytvoření požadované wiki stránky, jejíž obsah se bude předávat jako parametr editovací funkce (dalšími parametry jsou samozřejmě titulek a editovací token). Po provedení vložení nebo aktualizace wiki stránky odpoví MediaWiki buď „Success“ (operace proběhla v pořádku), nebo chybovým hlášením.

Další možností použití této funkce je k vkládání nových sekcí do stránky, což bude možné využít například u stránky s nezpracovanými e-maily. Tuto funkci je možné využít přidáním parametru „section=new“ do dotazu do MediaWiki.

3.4 Přístup k e-mailové schránce

Přístup k e-mailové schránce na poštovním serveru lze poměrně snadno naprogramovat pomocí jazyka PHP pomocí protokolu IMAP (Internet Message Access Protocol). Každý adresář na e-mailovém serveru má zprávy číslované pomocí UID – jednoznačného identifikátoru, který se v průběhu času nemění a každá další zpráva má o 1 vyšší identifikátor (funkce *auto_increment* = automatické inkrementování identifikátoru UID).

K e-mailovému serveru je možné se připojit pomocí funkce *imap_open*. Dále je možné získat seznam schránek, z kterých lze následně zjistit počet zpráv a jejich hodnoty UID. Po přečtení zprávy je dobré kontrolovat status a při jeho změně se musí aktualizovat informace ze schránky pomocí funkce *imap_reopen*. Protože jsou přístupy na e-mailový server pomalé, je lepší udržovat současný stav schránky lokálně a aktualizovat pouze změny (nově přijaté zprávy). Zjistí se tedy stav schránky pomocí *imap_status* a pokud došlo ke změně, musíme z e-mailového serveru stáhnout nová data. Hlavičky všech nových zpráv lze získat pomocí funkce jazyka PHP, které se předávají parametry od-do. Takto zjištěné zprávy lze následně procházet stažením těla vybrané zprávy.

[11]

3.4.1 Kontrola e-mailů a jejich zpracování

Jak již bylo napsáno výše, bude vytvořena databázová tabulka, kde bude uložen název schránky, následující UID a počet zpráv ve schránce. Kontrola a zpracování e-mailů ve schránce bude prováděna každý den.

Do speciálně vytvořené e-mailové schránky budou přeposílány kopie všech zpráv, odeslaných informačním systémem NLPIS i samotnými uživateli. Některé zprávy jsou ale z hlediska zpracování

nezajímavé, protože v nich obsažené informace jsou také v databázi informačního systému. E-maily, které se budou zpracovávat, jsou pouze reakce a komentáře k zaslaným týdenním výkazům, problémům a dalším.

Rozpoznání e-mailu, který je reakcí na předchozí zprávu, bude prováděno několika způsoby. První nejjednoznačnější varianta je identifikace „Re:“ v předmětu. Protože může reagující osoba toto slovo z předmětu smazat, provede se ještě kontrola těla zprávy. Pokud se v těle nalezne část textu, kde řádek začíná odsazením speciální značkou (nejčastěji „>“), bude se tento e-mail brát také jako odpověď.

Pro zpracování e-mailu je nutné rozpoznat hlavičku, která je při odeslání z informačního systému automaticky vkládána. Klíčovými údaji jsou datum pro zjištění týdne výkazu, login řešitele a identifikace projektu.

Jednotlivé části e-mailu jsou kódovány různými znakovými sadami a je nutné je pro správnou identifikaci vždy rozpoznat. Taktéž tělo zprávy je kódováno nějakou znakovou sadou, která je ale zapsána v parametrech zprávy v případě jednoduchého i „multipart“ e-mailu (e-mail, který se skládá z více částí). Tato funkce bude realizována nejprve načtením správného kódování a překódováním do používané znakové sady (zvolil jsem UTF-8) pomocí funkce *iconv*.

4 Implementace

V této kapitole se budu zabývat implementací navrhovaného rozšíření informačního systému NLPIS. Nejprve funkcemi pro připojení k databázi informačního systému a funkcemi pro přihlášení do aplikace MediaWiki. Dále postupně zmíním implementované moduly a některé stěžejní funkce popíšu podrobněji.

4.1 Připojení k databázi NLPIS a přihlášení do MediaWiki

K připojení do databáze informačního systému NLPIS jsem využil standardní funkce jazyka PHP.

Zajímavější částí bylo přihlášení do aplikace MediaWiki, jelikož není možné se přihlásit vyplněním přihlašovacího formuláře a odesláním pomocí JavaScriptu, protože k přihlášení je potřeba přihlašovací token, který se při každém přihlášení mění. Jako nejjednodušší funkční alternativu jsem zvolil využití modulu PHP CURL.

Nejprve bylo nutné získat přihlašovací token zasláním HTTP požadavku s přihlašovacími údaji pomocí CURL. Potom se odešle další HTTP požadavek s přihlašovacími údaji a se získaným přihlašovacím tokenem. Přijaté odpovědi jsou ve formátu XML, a proto je možné z nich potřebné údaje snadno zjistit pomocí funkcí PHP. Po zpracování druhé odpovědi zjistíme, zda bylo přihlášení úspěšné nebo nikoliv.

4.2 Generování a aktualizace wiki stránek

Pro generování i aktualizaci stránek v aplikaci MediaWiki je možné použít stejné funkce rozhraní API:Query a API:Edit.

Nejprve je třeba zjistit editovací token zasláním HTTP požadavku s využitím PHP CURL a funkce API:Query. Tato operace se provádí stejným způsobem pro existující i nově vytvářené wiki stránky. Jako parametr, který se do MediaWiki v požadavku zasílá, je použit titulek vytvářené resp. editované stránky.

Po obdržení editovacího tokenu je možné do MediaWiki zaslat obsah požadované stránky. Tato operace se ale provádí zasláním HTTP požadavku, a tak je nutné nový obsah wiki stránky nejprve zakódovat do kódování URL pomocí funkce *urlencode*, protože jedním z parametrů je právě obsah wiki stránky. Pokud by se tato operace neprovedla, mohla by se v požadavku chybně dekodovat některá část požadavku jako jiný požadavek nebo parametr. Vzhledem k tomu, že funkce

modulu CURL zasílá požadavek metodou POST, není velikost stránky příliš omezena (maximální velikost dat, zasílaných metodou POST, je nastavitelná).

Posledním krokem je tedy zaslání HTTP požadavku na rozhraní API:Edit, pomocí kterého je wiki stránka nahrána do aplikace MediaWiki.

Využitím této posloupnosti úkonů není wiki stránka nijak znehodnocena, naopak fungují všechny funkce MediaWiki jako je například historie.

Pomocí HTTP požadavku s využitím API:Query lze zjistit, zda je stránka v MediaWiki již vložena a bude se pouze aktualizovat, nebo jde o vytvoření. Vzhledem k tomu, že je nutné vědět, zda stránku vložit, nebo aktualizovat, vytvořil jsem si funkci, která jednoduše zjistí text stránky, nebo vrátí hodnotu *false* pro neexistující stránku.

Pokud tedy funkce na zjištění obsahu stránky vrátí hodnotu *false*, je možné generovat wiki stránku s požadovanými údaji, načtenými z databáze informačního systému. Vzhledem k tomu, že uživatelé aplikace MediaWiki mají přístup i pro zápis, mohou stránky také aktualizovat a měnit tak jejich obsah. Zde může vzniknout konflikt mezi databází a aktuální verzí wiki stránky. Stránky je třeba udržovat v konzistentním stavu vůči databázi, ale zároveň se nesmí zahodit dříve zapsané údaje z důvodu historie, ani údaje zapsané uživatelem z důvodu nechtěné ztráty důležitých dat. Pokud je tedy na stránce nalezena nějaká nesrovnalost, je toto řešeno několika způsoby dle typu stránky. Konkrétní řešení aktualizace bude u jednotlivých typů stránek podrobněji popsáno dále.

4.2.1 Aktualizace tabulky prostředků skupiny projektů a projektu

Aktualizace tabulky prostředků byla implementována dle návrhu. Tabulka je procházena po řádcích a vždy se kontroluje obsah s požadovaným. Pokud se narazí na nějakou nesrovnalost, navrhl jsem několik způsobů řešení.

První možností nekonzistence je přesunutí prostředku v tabulce na jiný řádek. Toto může vzniknout přidáním prostředku k projektu či skupině, aktualizací tabulky s prostředky v databázi informačního systému, nebo úpravou tabulky s prostředky uživatelem na stránce skupiny projektů či projektu.

Jiný případ může nastat úpravou oprávnění. V tomto případě se u daného prostředku pouze oprávnění nahradí, ale i zde je několik variant (viz Tabulka 1). Pokud není stávající oprávnění rozpoznáno, nahradí se správným dle dat v databázi.

Poslední možností je odebrání prostředku v informačním systému respektive v jeho databázi. V tomto případě dojde k přesunutí prostředku na konec tabulky a nahrazení oprávnění na *žádné* dle výše popsaných pravidel.

Výchozí oprávnění	Požadované oprávnění	Výsledné oprávnění
žádné	zápis	zápis
žádné*	zápis	zápis
žádné	čtení	čtení
žádné*	čtení	čtení*
čtení	zápis	zápis
čtení*	zápis	zápis
čtení	žádné	žádné
čtení*	žádné	žádné*
zápis	čtení	čtení*
zápis	žádné	žádné*

Tabulka 1 - Upravení oprávnění

4.2.2 Titulek wiki stránky

Titulek je v aplikaci MediaWiki tvořen automaticky z odkazu na danou stránku. Takto tvořené titulky ale nemusí vždy plně odpovídat požadovanému titulku, a proto jsem využil rozšíření MediaWiki „No Title“ (žádný titulek). Po vložení klíčového slova „`__NOTITLE__`“ na začátek wiki stránky je automatické vytvoření titulku vypnuto.

Titulek stránky je vždy tvořen názvem skupiny nebo projektu, další nadpisy se skládají z více částí. Aby bylo možné rychle přecházet strukturou stránek, je například na stránce s reakcí na pracovní výkaz nadpis tvořen odkazem na:

1. stránku projektu,
2. stránku s odkazy na pracovní výkazy,
3. stránku s týdenním výkazem
4. a konečně vlastní nadpis s datem a odesilatelem e-mailu.

Podobně je tvořena struktura nadpisů i na ostatních stránkách, které budou popsány dále.

4.2.3 Automaticky generovaná tabulka obsahu stránky

Tabulka s obsahem se v MediaWiki automaticky zobrazuje, pokud je na stránce 4 a více nadpisů. Na stránce skupiny projektů či projektu je toto rozšíření vhodné využít, ale na stránce s odkazy na výkazy nebo s týdenním výkazem by šlo o příliš dlouhý výčet, který by stránky spíše zneprůhlednil a důležitý obsah odsunul dále.

Pro vypnutí této funkce je možné využít klíčové slovo „`__NOTOC__`“ (z angličtiny NO Table Of Contents), které se vloží na začátek stránky.

4.2.4 Wiki stránka skupiny projektů

V době návrhu nebyly skupiny projektů v informačním systému NLPIS ještě úplně dořešeny, ale přesto jsem se je pokusil zpracovat dle požadovaných kritérií, jak je vidět na obrázku (Obrázek 7). Nyní jsou skupiny projektů navrženy, ale tabulky nejsou ještě zcela naplněny daty.

relatedness

Vedoucí skupiny: doc. RNDr., Ph.D. Pavel Smrz

Prostředky skupiny:

název	cesta	oprávnění	patří k projektu
-------	-------	-----------	------------------

*Poznámka k oprávnění s *:* dříve bylo oprávnění **zápis**

Projekty skupiny:

- Dolování znalostí ze slovních profilů (semantic_relatedness - Řešený) [↗](#)
- Hledání souvislostí mezi slovy (souvislosti2 - Řešený) [↗](#)
- Testovací data pro měření sémantické blízkosti (testovaci_sada_blizkosti_clanku - Řešený) [↗](#)
- Automatický thesaurus (automaticky_thesaurus - K rozhodnutí) [↗](#)
- Metody sémantické blízkosti ve folksonomiích (similarity_in_folksonomies - K rozhodnutí) [↗](#)
- Sémantická blízkost dokumentů (related_docs - Ukončený) [↗](#)
- Automatická tvorba rejstříku publikace (rejstrik2 - Ukončený) [↗](#)
- Automatické navrhování klíčových slov (keywords - Ukončený) [↗](#)
- Hledání souvislostí mezi slovy (souvislosti - Ukončený) [↗](#)
- Automatické generování tezauru (term_similarity - Ukončený) [↗](#)
- Hledání souvislostí mezi slovy (hledani_souvislosti - Ukončený) [↗](#)
- Sémantická blízkost termů (blizkost_termu - Ukončený) [↗](#)
- Převod odpovědí z internetových diskusí pro určení relevance (relevantni_dokumenty - Ukončený) [↗](#)

Obrázek 7 - Stránka skupiny

4.2.4.1 Aktualizace stránek skupiny projektů

Tato stránka obsahuje jako všechny ostatní wiki stránky titulek, který je zde tvořen názvem skupiny. U titulku není v případě nekonzistence vytvářena poznámka, ale titulek se nahradí. Nahrazení se provádí několika způsoby a to podle nalezeného obsahu na wiki stránce.

Dalším klíčovým slovem je vedoucí skupiny. Zde se již provádí klasický výše popsaný algoritmus aktualizace, kdy se neaktuální obsah vkládá do poznámky a nahrazuje se aktuálními daty z databáze informačního systému.

Třetí část wiki stránky skupiny projektů je tvořena prostředky této skupiny. Tabulka s prostředky je aktualizována dle kapitoly 4.2.1.

Předposlední, ale neméně důležitou částí je výpis projektů. Jednotlivé řádky obsahují vždy odkaz na stránku daného projektu skupiny a zobrazen je jeho název, zkratka a stav projektu

(Nezadaný, Řešený, Ukončený nebo K rozhodnutí). Aktualizaci jsem implementoval tak, že se provede například přeuspořádání řádků a jejich oprava. Chybějící řádky se doplní dle dat v databázi, naopak přebývajících projekty se vloží do poznámky. Jednotlivé řádky s projekty se kontrolují pouze po zkratku, takže stav projektu se může aktualizovat, aniž by se řádek vkládal zbytečně do poznámky.

Posledním klíčovým slovem je ukončení stránky, pomocí kterého se nachází konec obsahu předchozího klíče. Za ukončením hlavičky stránky je tedy možné psát další libovolný text, který již ale není předmět kontroly a aktualizace.

4.2.5 Wiki stránka projektu

Stránka projektu je z hlediska generování a aktualizace asi nejkomplicovanější částí celého projektu. Zde je obzvlášť důležité, aby nedošlo ke znehodnocení nebo ztrátě dat, protože se jedná o nejdůležitější, respektive výchozí stránku celé hierarchie.

Složitost této stránky spočívá především v možnostech variace počtu řešitelů a jejich změnách. Dalším problémem jsou projekty se stejným URL. Projekty se stejnou adresou jsou totiž generovány a aktualizovány na jedné společné wiki stránce. Celou problematiku generování a aktualizací této stránky budu postupně rozebírat v následujících kapitolách.

Klíčová slova na stránce projektu s jedním řešitelem jsou titulky, řešitel, vedoucí, zahájení projektu, ukončení řešení, typ projektu, skupina, prostředky projektu, odkaz na stránku s pracovními výkazy, úkoly daného řešitele a ukončení. Stránka s více řešiteli má jinak řešenou část s řešiteli a jejich úkoly jsou přesunuty na samostatnou stránku. Obsah za těmito klíčovými slovy je poměrně snadno odvoditelný.

Rozvedl bych ale nadpis, který se skládá z klíčového slova `__NOTITLE__`, poznámky o začátku automaticky generované sekce a samotného nadpisu stránky (název projektu a jeho zkratka). Dále typ projektu, který může nabývat hodnot disertační, diplomová nebo bakalářská práce a obecný projekt. Ukončení je složeno z oddělovací čáry a poznámky o konci automaticky generované sekce.

Ukončení generované hlavičky je důležité nejen pro zjištění obsahu klíčového slova před ním, ale také slouží jako zarážka, aby se klíčová slova nehledala v celém textu stránky. Pokud není ukončení nalezeno, automaticky se generuje nová hlavička. Za tímto účelem jsem implementoval funkci, která vytváří a vkládá chybějící sekci neboli hlavičku projektu na správnou pozici na wiki stránce. Pozice hlavičky je ve většině případů zcela na začátku, kromě případů, kdy je na jedné stránce více generovaných hlaviček (pro projekty se stejným URL).

4.2.5.1 Aktualizace stránky projektu

Kontrola údajů je rozdělena na několik částí. První částí je zjištění, zda stránka vůbec obsahuje hlavičku, neboli pokus o nalezení titulku a ukončení. Pokud není nalezeno ukončení, generuje se nová

hlavička. Nyní se postupně prochází jednotlivá klíčová slova a hledají se mezi titulkem a ukončením. Když slovo není nalezeno, použije se funkce pro vložení chybějícího klíčového slova s obsahem za obsah předchozího klíčového slova. Naopak když je nalezeno, zavolá se funkce na kontrolu jeho obsahu. Aktualizace chybných údajů u řešitele či řešitelů, vedoucího, data zahájení projektů, ukončení řešení a úkolů řešitele se provádí přesunutím do poznámky k danému klíči a vložním validních dat z databáze informačního systému NLPIS. Titulek se nahrazuje a prostředky se přesouvají a aktualizují v příslušné tabulce.

Pokud se provádí kontrola projektů se stejným URL, provedou se výše popsané úkoly pro každou hlavičku projektu.

4.2.5.2 Projekt bez řešitele

Projekt, který nemá přiděleného žádného řešitele, může být buď nový, nebo je již řešený. U již řešeného projektu se může stát, že řešitel byl z projektu odvolán. Tato situace většinou nastane, když přidělený řešitel na projektu neudělá vůbec nic a svou činnost ukončí. Takový projekt může být například přenechán jinému řešiteli.

Co se týče obsahu, jsou zde zastoupena všechna klíčová slova jako u ostatních projektů, jen za některými není žádný obsah.

V prvním případě jde o jednoznačné vygenerování stránky projektu, kterou jsem řešil stejným stylem, jako stránku pro jednoho řešitele. Pokud byl projekt již řešen, aktualizuje se tato stránka dle dřívějšího typu, tedy řešitel, zahájení projektu, ukončení řešení a případné úkoly se dají do poznámky.

Projekt bez řešitele (bez_resitele)

Řešitel:
Poznámky k Řešitel: Bc. Ales Vavrinek

Vedoucí: doc. RNDr., Ph.D. Pavel Smrz

Zahájení projektu:
Poznámky k Zahájení projektu: 08. 05. 2011

Ukončení řešení:
Poznámky k Ukončení řešení: 09. 05. 2011

Typ projektu: Obecný projekt

Skupina: -

Prostředky:

název	cesta	oprávnění	patří k projektu
dve_zkratky_1	/mnt/minerva1/nlp/projects/dve_zkratky_1	žádné	Pokusný projekt se stejnými zkratkami 1

*Poznámka k oprávnění s *:* dříve bylo oprávnění zápis

[Pracovní výkazy](#)

Úkoly:
Poznámky k Úkoly:
 úkoly řešitele projektu

Obrázek 8 - Upravený projekt bez řešitele (dříve 1 řešitel)

4.2.5.3 Projekt s jedním řešitelem

Tento typ projektu je asi nejčastější variantou. Naprostá většina bakalářských či diplomových prací je řešena právě jedním řešitelem. Ale i zde se může stát, že student práci nedokončí, nebo na jeho část bude navazovat jiný řešitel.

Vzhledem k tomu, že projekty mohou být za jistých okolností přidělovány opakovaně, existuje ještě možnost založení nového projektu, který ale bude mít stejnou zkratku nebo jednodušší varianta přidání dalšího řešitele. Tyto případy budu dále rozvíjet v následujících kapitolách.

4.2.5.4 Projekt s více řešiteli

Rozdíly v generované stránce oproti projektu s jedním řešitelem jsou pouze u klíče řešitelé, kde jsou jednotliví řešitelé vypsáni pod sebou. U každého z nich je odkaz na stránku s jeho úkoly a od kdy, do kdy projekt řešil, nebo ještě řeší. Z toho již vyplývá, že úkoly jsou přesunuty na samostatné stránky pro každého řešitele.

Při aktualizaci se provádí kontrola nejen stránky projektu, ale také stránek s úkoly řešitelů daného projektu. Aktualizace stránky s úkoly se provádí tak, aby nedošlo ke ztrátě dat. Dojde-li k nalezení nesrovnalostí, přesune se aktuální text do poznámky s časovým razítkem (datum aktualizace) a požadovaný obsah se vloží za nadpis (před vytvořenou poznámku).



Obrázek 9 - Stránka s úkoly řešitele projektu s více řešiteli, úprava textu

4.2.5.5 Více projektů se stejným URL

V databázi respektive v informačním systému NLPIS je možné vytvořit více projektů se stejným URL. V MediaWiki jsou pro tuto variantu vkládány hlavičky projektů na jednu stránku pod sebe. Hlavičky se řeší stejným způsobem jak pro jednoho, tak pro více řešitelů jednotlivých projektů. Pokud je v informačním systému u projektu nastavena proměnná url_dle_zkratky, vytváří se URL z odkazu do MediaWiki následovaného zkratkou projektu.

Rozpoznání projektů se stejným URL se provádí několika způsoby. První možností je, že daný projekt nemá nastavenou proměnnou url_dle_zkratky. Potom je ale nutné provést kontrolu, zda začátek URL skutečně odpovídá požadovanému, tedy odkazu do MediaWiki. Pokud ano, vyhledají se v databázi další projekty se stejným URL a také projekty, které nemají nastavenou proměnnou url_dle_zkratky, ale začátek adresy stránky neodpovídá adrese MediaWiki a zároveň je zkratka toho projektu shodná se zbytkem odkazu (za adresou MediaWiki) tohoto projektu.

Další možností je, že proměnná `url_dle_zkratky` není nastavena a při tom zadaná URL neodpovídá adrese do MediaWiki. V tomto případě se bude adresa stránky vytvářet stejným způsobem, jako kdyby proměnná byla nastavena. Vyhledání projektů se stejným URL bude probíhat pro oba tyto případy (nastavená proměnná `url_dle_zkratky` nebo chybně zadaná URL) stejně. Proveďte se vyhledání projektů se stejným URL, jako má daný projekt a opět také s projekty, které nemají nastavenou proměnnou `url_dle_zkratky`, vložená URL je chybná a zkratka takového projektu odpovídá zbytku odkazu tohoto projektu do aplikace MediaWiki.

Aktualizace se provádí vždy kontrolou údajů mezi hlavičkou a ukončením, takže se úpravy realizují postupně pro všechny hlavičky projektů na dané wiki stránce (viz Obrázek 10).

Pokusný projekt se stejnými zkratkami 1 (dve_zkratky)

Řešitel: Ing. Jaroslav Dytrych
 Vedoucí: Ing. Jaroslav Dytrych
 Zahájení projektu: 06. 04. 2011
 Ukončení řešení: 07. 04. 2011
 Typ projektu: Diplomová práce
 Skupina: [misc](#)
 Prostředky:

název	cesta	oprávnění	patří k projektu
dve_zkratky_1	/mnt/minerva1/nlp/projects/dve_zkratky_1	zápis	Pokusný projekt se stejnými zkratkami 1

*Poznámka k oprávnění s *: dříve bylo oprávnění zápis*
[Pracovní výkazy](#)

Úkoly:
 ukoly resitele prvnioho projektu se stejnou zkratkou

Pokusný projekt se stejnými zkratkami 2 (dve_zkratky)

Řešitelé:
 ■ Bc. Ales Vavrinek ([Úkoly řešitele](#), řešil: 01. 04. 2011 - 07. 04. 2011)
 ■ doc. RNDr., Ph.D. Pavel Smrz ([Úkoly řešitele](#), řešil: 01. 04. 2011 - 30. 04. 2011)
 Vedoucí: doc. RNDr., Ph.D. Pavel Smrz
 Zahájení projektu: 01. 04. 2011
 Ukončení řešení: 30. 04. 2011
 Typ projektu: Obecný projekt
 Skupina: [misc](#)
 Prostředky:

název	cesta	oprávnění	patří k projektu
dve_zkratky_2	/mnt/minerva1/nlp/projects/dve_zkratky_2	zápis	Pokusný projekt se stejnými zkratkami 2

*Poznámka k oprávnění s *: dříve bylo oprávnění zápis*
[Pracovní výkazy](#)

Obrázek 10 - Stránka projektů se stejným URL

4.2.6 Wiki stránka s odkazy na výkazy

Přístup na tuto stránku je možný ze stránky projektu, kde je odkaz „Pracovní výkazy“. Obsahem této stránky je výpis jednotlivých týdnů, kdy měl řešitel nějakou vykázanou činnost. Každý takový výpis se skládá z:

- hlavičky,
- jména řešitele, které zároveň tvoří odkaz na stránku s daným týdenním výkazem,
- odpracovanými hodinami řešitele za daný týden a
- několika úvodních slov souhrnu, což je možné nastavit.

Pokud jde o projekt s více řešiteli, jsou jejich údaje, pokud byly zadány, vypsány pod jednou hlavičkou (viz Obrázek 11).

Rok 2010, 38. týden (20. 09. 2010 - 26. 09. 2010) [editovat]	
Řešitel Jan Svec , odpracováno 10 hodin, 31 minut	
Souhrn: Po konzultaci, som skontroloval pocitanie vysledkov korelacie spearmanom a objavil som chybu v skripte ktory prevadza vypocet Spustil som vyfiltravanie zaujmych slov s novych dat.	
Řešitel Alena Pavelkova , odpracováno 1 hodin, 35 minut	
<hr/>	
Rok 2010, 37. týden (13. 09. 2010 - 19. 09. 2010) [editovat]	
Řešitel Alena Pavelkova , odpracováno 3 hodin, 54 minut	
Souhrn: Úprava cest python skriptů, víceznačná lemmata (nedokončeno)	
<hr/>	
Rok 2010, 36. týden (06. 09. 2010 - 12. 09. 2010) [editovat]	
Řešitel Alena Pavelkova , odpracováno 6 hodin, 44 minut	
Souhrn: Úprava wiki, nezobrazování vlastních jmen u anglických souvislostí, oprava u českých	
Řešitel Jan Svec , odpracováno 23 hodin, 19 minut	
Souhrn: Tento tyzden som dokoncoval data pre porovnanie anglickej casti. Prepocital som hodnoty dat percentualne, aby sa dali porovnat. Mam zostavene porovnanie vsetkych 3 suborov(cz, eng, wordsim) su v <code>./wat/(czech,english,wordsim353)/compared/cmp_(eng,cz,wordsim353)</code> . Toto porovnanie som robil na zaciatku....	

Obrázek 11 - Odkazy na výkazy dvou řešitelů projektu

4.2.6.1 Generování stránky s odkazy na výkazy

Titulek stránky je tvořen odkazem na stránku projektu a samotným nadpisem stránky. Protože je zde velké množství nadpisů, bylo vhodné vypnout funkci zobrazování tabulky s obsahem a samozřejmě bylo také důležité vypnutí automatického titulku pomocí vložení kombinace klíčových slov `__NOTITLE__` a `__NOTOC__` na začátek stránky. Při generování této stránky se prochází postupně všechny týdenní výkazy daných řešitelů projektu a pro každý týden, kdy alespoň jeden řešitel vykázal nějakou činnost, se generuje hlavička týdne následovaná proměnným počtem výkazů řešitelů, které se skládají z výše popsaných částí. Počet hodin je spočítán jako součet odpracovaného času

za jednotlivé denní výkazy v daném týdnu. Souhrn je uváděn, pouze pokud byl zadán do informačního systému NLPIS. Pro případ, že by souhrn byl příliš dlouhý, se provádí jeho zkrácení na nastavený maximální počet znaků, kdy se od tohoto čísla dojde do konce slova a ukončí se „...“. Jednotlivé výkazy jsou na stránku vkládány sestupně, tedy nejnovější výkaz bude nahoře.

Protože tyto údaje mohou být v informačním systému často doplňovány a měněny, použil jsem pro odpracované hodiny a souhrn kontrolní součet – hash, který je umístěn za odkazem na stránku s týdenním výkazem. Je tedy spočítán z odpracovaných hodin a souhrnu. Pro výpočet jsem využil hashovací funkci MD5.

Vzhledem k tomu, že jednotlivé týdenní výkazy musí být nějak odděleny, použil jsem vodorovnou čáru, která se v MediaWiki vytváří pomocí vložení „----“ na začátek řádku. Aby nedošlo k chybné reprezentaci oddělovače, provádí se automaticky za pomlčkou vložení mezery.

Pokud se jedná o více projektů se stejnými URL, jsou pracovní výkazy generovány pod sebou, ale jsou odděleny ukončením a titulkem projektu. Aby byl snazší přístup k jednotlivým výkazům projektů, je na stránce projektu v odkazu na výkazy přidána část, která odkazuje přímo na nadpis daného projektu na stránce s odkazy na výkazy.

4.2.6.2 Aktualizace stránky s odkazy na výkazy

Nejdříve se provede kontrola titulku, zda se skládá z klíčových slov `__NOTITLE__` a `__NOTOTC__` následovaných vytvořeným nadpisem. Aktualizace této stránky se provádí postupným procházením jednotlivých týdenních výkazů. Vždy se nejprve hledá hlavička výkazu a jeho ukončení, tedy „----“. Tímto způsobem je možné najít obsah daného výkazu.

Pokud není nalezena hlavička, generuje se nový výkaz, který se vloží na správné místo, které je za novějším a před starším výkazem. Nejčastěji se bude vkládat nový výkaz pod titulek stránky.

Naopak při nalezení hlavičky výkazu se může provést kontrola jeho obsahu. Nejprve se zjistí, zda je výkaz shodný s požadovaným, pokud ano, přejde se na další, pokud ne, provede se kontrola a případná aktualizace. Kontrola probíhá nejdříve zjištěním, zda výkaz neobsahuje poznámku, pokud ano, načte se výkaz pouze po ni. V dalším kroku se zjistí generovaný výsledek funkce MD5. Jestliže není kontrolní výpočet MD5 nalezen, vkládá se obsah výkazu do poznámky a generuje se nový. V opačném případě je načten, získá se obsah výkazu a provede se jeho porovnání s načteným hashem. Jestliže je výsledek porovnání shodný, je možné odpracované hodiny a začátek souhrnu nahradit. V opačném případě se generuje nový obsah (včetně hashe) a starý se vkládá do poznámky.

Pokud má projekt více řešitelů, provádí se kontrola stejným způsobem, jen se pod jednou hlavičkou týdne hledá více odkazů na týdenní výkaz. Další postup je ale stejný. Jestliže je výkaz již u předchozího řešitele nahrazen, není třeba provádět kontrolu u dalších řešitelů daného projektu.

4.2.7 Wiki stránka s týdenním výkazem

Do aplikace jsou vkládány týdenní výkazy pro každého řešitele zvlášť i v případě, že se jedná o projekt s více řešiteli. Adresa stránky má tvar: adresa projektu/výkazy/login řešitele/rok_číslo týdne. Jak je možné vidět na obrázku (Obrázek 12), týdenní výkaz řešitele se skládá z nadpisu, podnadpisu, odkazů na jednotlivé dny, souhrnu, plánu, problémů a nejasností s reakcemi, reakcí na výkaz a jednotlivých denních výkazů.

Hlavní nadpis stránky je složen opět z klíčových slov `__NOTITLE__` a `__NOTOC__`, dále z odkazu na stránku projektu, odkazu na stránku s odkazy na výkazy a vlastního nadpisu s identifikací týdne a řešitele. Pro snadnější procházení mezi výkazy se po kliknutí na odkaz na stránku s odkazy na výkazy vrátíme na tuto stránku na týden s právě prohlíženým výkazem.

Podnadpis stránky je tvořen detailnějšími informacemi o zobrazeném výkazu doplněnými o odpracované hodiny.

Odkazy na jednotlivé denní výkazy odkazují na první denní výkaz z daného dne. Tyto odkazy jsou generovány pouze na dny, kdy byl výkaz zadán.

Dále je zde zobrazen celý souhrn, u kterého je spočítán hash pro snadnou aktualizaci, která bude popsána dále. Následuje plán na příští týden, u kterého je taktéž spočítán hash.

Problémy a nejasnosti s reakcemi jsou na stránku vkládány, pouze pokud je řešitel do výkazu zadal. V tom případě je u problémů opět spočítána kontrolní hodnota pomocí funkce MD5. U reakce jsou vkládány přímo z e-mailu, který reaguje na daný problém. Není třeba kontrolní funkci počítat, protože zde není možné provádět aktualizaci, jelikož e-maily se prochází pouze jednou.

Reakce na výkaz jsou na stránce zobrazeny vždy a právě sem se vkládají všechny e-maily s reakcemi, pokud nejsou zadány problémy a nejasnosti.

Z e-mailu je vybrána pouze odpověď na problém či výkaz, ale pro získání dalšího kontextu, je vložen celý jeho text do aplikace MediaWiki na samostatnou stránku, na kterou vede odkaz umístěný za odpovědí z tohoto e-mailu.

Následující sekce je věnována jednotlivým denním výkazům, které jsou vždy uvozeny hlavičkou. Hlavička se skládá z id výkazu, dne v týdnu, data, času od – do a odpracovaných hodin s minutami. Čas řešení od – do je vkládán do hlavičky, pouze pokud byl zadán. Za hlavičkou je kontrolní výpočet MD5 spočtený z dané činnosti, která jej hned následuje. Jednotlivé denní výkazy jsou odděleny opět vodorovnou čarou. Obdobně jako na stránce s odkazy na výkazy je i u denních výkazů za pomlčku vkládána mezera, aby nedošlo k chybné reprezentaci oddělovače.

Automatické generování wiki stránek z e-mailů a IS (wiki_z_nlpis) - Pracovní výkazy - Rok 2010, 43. týden, výkaz řešitele Bc. Ales Vavrinek
<p>Týdenní výkaz řešitele Bc. Ales Vavrinek za 43. týden (25. 10. 2010 - 31. 10. 2010), celkem odpracováno 1 hodin, 0 minut</p> <p>Denní výkazy: Neděle</p> <p>Souhrn: pokus o připojení na minerva1</p> <p>Plán na příští týden: projit výtah z databáze NLPIS a mailů</p> <p>Problémy a nejasnosti: Dobrý den, mohl bych se prosím zeptat, jak se dá připojit na počítač minerva1? Pokoušel jsem se přes WinSCP3, ale pouze vyprší timeout a nic se nestane (Server unexpectedly closed network connection). Děkuji za odpověď.</p> <p>Reakce na problémy a nejasnosti:</p> <p>Dne 28. 04. 2011 od ales.vavrinek@seznam.cz () Dobrý den, na minerva1 se lze připojit pouze z domény fit.vutbr.cz přes IPv4. Pokud se připojujete z domu, je potřeba se připojit přes jiný server. Pokud chcete pouze přistupovat k adresáři projektu, můžete se připojit pouze na merlina (třeba přes to WinSCP3) a využít toho, že jako člen skupiny NLP máte minervu1 připojenou na merlinovi (/mnt/minerva1/...). Pokud chcete na minervě dělat něco dalšího, je třeba se nejprve připojit přes ssh (pod Windows lze využít Putty) na merlina a z toho potom přes ssh na minervu.</p> <p>Přeji pěkný den</p> <p>Ing. Jaroslav Dytrych idytrych@fit.vutbr.cz</p> <p>Celý text e-mailu</p> <p>Reakce na výkaz:</p> <hr/> <p>Neděle 31. 10. 2010, odpracováno 1 hodin, 0 minut pokus o připojení na minerva1</p>

Obrázek 12 - Týdenní výkaz s problémy a reakcí

4.2.7.1 Aktualizace stránky s týdenním výkazem

První porovnání se provádí na rozdíl v celé stránce. V případě, že žádné nejsou, pokračuje se další stránkou, jinak se provede kontrola a případné aktualizace. Při aktualizaci jsou kontrolovány postupně jednotlivé položky týdenního výkazu. Pokud je chybný titulek nebo podtitulek, je nahrazen. Kotvy na jednotlivé dny se také nahrazují podle změn v denních výkazech. U všech položek se vždy nejprve zkusí porovnání celého pole a při neshodě se pokračuje v hledání změn a opravách.

Souhrn je na této stránce umístěn vždy a tak je možné jej kontrolovat dle spočtené hashovací funkce MD5 s aktuálním obsahem. Pokud je výsledek shodný, text se může nahradit aktuální verzí z databáze informačního systému. V opačném případě se přesune do poznámky a nahraje se nový, pro který se opět spočte kontrolní výpočet funkce MD5. Stejným způsobem se řeší také plán na příští týden a případné problémy. Reakce na problémy a reakce na výkaz se nekontrolují, pouze je možné vkládat nové reakce z e-mailů.

Je možné, že řešitel nejprve během týdne odešle výkaz s problémy, ale na konci týdne je už neuvede. V takovém případě se dají problémy a nejasnosti včetně reakcí do poznámky, aby nedošlo ke ztrátě informací.

U denních výkazů se postupuje obdobně. Nejprve se hledá hlavička, pokud se nenajde, vkládá se nový výkaz, jinak se provede načtení kontrolního součtu a porovnání s aktuálním obsahem.

Vzhledem k tomu, že jsou jednotlivé výkazy odděleny vodorovnou čarou (v MediaWiki vytvářenou pomocí „----“), je poměrně snadné jednotlivé texty činností mezi hlavičkou a koncem rozpoznat a porovnat je s hashem. V případě, že hash s textem nesouhlasí, vytvoří se k danému výkazu poznámka a vloží se aktuální text, jinak je obsah včetně výpočtu nové hodnoty MD5 aktualizován.

Následně se provádí kontrola, zda na stránce nejsou zbytečné výkazy, které nebyly rozpoznány kvůli změně v hlavičce. Pokud je takový výkaz nalezen, je přesunut na konec stránky a jeho hlavička upravena, aby nedošlo k chybnému rozpoznání při příští aktualizaci. Aby bylo jasné, kdy byl takový výkaz smazán, je před něj vložen text „Neplatný výkaz ze dne “ a aktuální datum.

Poslední kontrolou je aktualizace kotev na denní výkazy. V případě, že je nalezena nějaká nesrovnalost, jsou automaticky nahrazeny novými.

Kontrola a aktualizace týdenního výkazu se spouští při kontrolování stránky s odkazy na výkazy a to vždy při zpracování daného výkazu.

4.2.8 Zpracování e-mailů

Pro kontrolu zpráv jsem v databázi informačního systému vytvořil tabulku, ve které je úplný název čtené schránky, UID následující očekávané zprávy a celkový počet zpráv. Tyto údaje se vždy nejprve načtou a poté jsou porovnávány se získanými informacemi ze schránky.

Následně se provede funkce k připojení k definovanému serveru a výběr příslušné schránky. Nyní je možné zjistit, zda se v e-mailové schránce něco změnilo, tedy porovnájí se získané hodnoty s informacemi z databáze. V případě, že jsou stejné, může se operace ukončit, v opačném případě se provede kontrola přijatých zpráv. Provede se načtení rozsahu zpráv od UID z databáze po nově získané.

Vzhledem k tomu, že nás zajímají pouze reakce na problémy a nejasnosti nebo celé výkazy, je třeba nalézt pouze tyto odpovědi. Vyhledávání se provádí na dvou místech a to v předmětu a tělu zprávy. Rozpoznání odpovědi z předmětu je nalezením „Re:“, které se porovnává nezávisle na velikosti písmen. Druhou možností je, že reagující osoba tuto část předmětu smaže a tak je třeba se podívat do samotného těla zprávy, kde se na začátcích řádků vyhledávají příslušné symboly, značící odpověď. Nejčastěji to bývá znak „>“.

Aby bylo možné porovnávat text v předmětu či těle zprávy, je nutné nejprve zjistit, v jakém kódování byl vložen a překódovat jej na mnou zvolené „UTF-8“. K tomuto účelu se vyhledává parametr „charset“ (z angličtiny: znaková sada) ve struktuře zprávy načtené funkcí *imap_fetchstructure*. V případě, že bude zpracováván e-mail obsahuje HTML kód, provede se kontrola, zda se jedná o odesilatele z databáze informačního systému NLPIS nebo o adresu, která je v seznamu výjimek v konfiguračním souboru. V případě nenalezení adresy odesilatele se taková zpráva nebude procházet a vloží mezi nezpracované e-maily.

Nyní již víme, že příslušný e-mail je reakcí na problémy a nejasnosti nebo na výkaz. Aby bylo toto možné zjistit, musí se v těle zprávy najít hlavička, ve které jsou základní informace o projektu. Podle hlavičky jsou zprávy rozděleny, viz Tabulka 2.

Hlavička v těle zprávy	Odpověď se přiřadí do ...
Týdenní zpráva (login řešitele)	reakce na výkaz
Problémy a týdenní zpráva (login řešitele)	reakce na problémy a nejasnosti
Zpráva o problémech uživatele	reakce na problémy a nejasnosti

Tabulka 2 - Rozdělení přijatých e-mailů

Hlavička projektu v e-mailu je dále prohledávána na výskyt informací o uživateli, id projektu a identifikaci týdne. Pokud nejsou tato základní data nalezena, není možné zprávu zpracovat. V opačném případě je ze zprávy vybrána pouze odpověď, která je vložena do příslušného týdenního výkazu řešitele, nebo je stránka s výkazem teprve vytvořena. Vložená odpověď je ukončena odkazem na celý text zpracovávaného e-mailu, aby bylo možné zjistit další kontext z právě uložených výkazů a dalších informací, jak je možné vidět na obrázku (Obrázek 12), kde je zobrazena horní část stránky s celým textem zpracovaného e-mailu.

Celý text e-mailu je tedy vložen na samostatnou wiki stránku, která obsahuje nadpis, který se skládá z odkazu na stránku projektu, odkazu na stránku s odkazy na výkazy, odkazu na stránku s týdenním výkazem a samotného nadpisu stránky následovaného předmětem e-mailu a jeho textem (viz Obrázek 13).

Potom, co se projdou všechny nové e-maily, se v databázi informačního systému aktualizuje UID následující zprávy i počet zpráv ve schránce.

Automatické generování wiki stránek z e-mailů a IS (wiki_z_nlpis) - Pracovní výkazy - Rok 2010, 43. týden, výkaz řešitele Bc. Ales Vavrinek - Celý text e-mailu od ales.vavrinek@seznam.cz ze dne 02. 11. 2010

Předmět e-mailu: Re: NLPIS - problémy, týdenní zpráva (xvavri04)

Dobrý den,

na minerva1 se lze připojit pouze z domény fit.vutbr.cz přes IPv4. Pokud se připojujete z domu, je potřeba se připojit přes jiný server. Pokud chcete pouze přistupovat k adresáři projektu, můžete se připojit pouze na merlina (třeba přes to WinSCP3) a využít toho, že jako člen skupiny NLP máte minervu1 připojenou na merlinovi (/mnt/minerva1/...). Pokud chcete na minervě dělat něco dalšího, je třeba se nejprve připojit přes ssh (pod Windows lze využít Putty) na merlina a z toho potom přes ssh na minervu.

Přeji pěkný den

Ing. Jaroslav Dytrych
idytrych@fit.vutbr.cz

Dne 2.11.2010 09:23, Ales Vavrinek napsal(a):

- > Problémy a týdenní zpráva (xvavri04)
- > Týden: 43. (25.10.2010 - 31.10.2010)
- > Uživatel: Vavrinek Ales, Bc. (xvavri04)
- > Projekt: 237. DP: Automatické generování wiki stránek z e-mailů a IS (https://merlin.fit.vutbr.cz/nlp-wiki/index.php/wiki_z_nlpis)
- > Zkratka: wiki_z_nlpis (<http://nlp.fit.vutbr.cz/nlpis/index2.php?str=13&varStr=0&uprPol=237&uprU=203>)
- > Vedoucí: Dytrych Jaroslav, Ing. (idytrych@fit.vutbr.cz)
- > Poslední změna wiki:
- > Poslední změna adresáře projektu:
- > Za týden na projektu odpracováno: 1:00
- > Odpracovat zbývá: 147:00

Obrázek 13 - Celý text zpracovaného e-mailu (pouze část stránky)

4.2.8.1 Nezpracované e-maily

V případě, že je u e-mailu zjištěno, že se jedná o odpověď, tedy reakci na výkaz nebo problémy a nejasnosti, a není nalezena hlavička projektu, nebo některá z vyhledávaných položek, není možné tuto zprávu zpracovat a je umístěna na speciální wiki stránku s nezpracovanými e-maily (viz Obrázek 14).

Nezpracovane e-mail

Nezpracovaný e-mail pro xvavri04@stud.fit.vutbr.cz (Vavrinek Aleš) od idytrych@fit.vutbr.cz (Jaroslav Dytrych) s předmětem: Re: dp - pokračování, ze dne 02. 05. 2011: Text e-mailu

Obrázek 14 - Stránka s nezpracovanými e-maily

Tato stránka není nijak složitě formátována a slouží spíše správci systému k zjištění, zda nebyl zahozen důležitý e-mail. Aktualizace je prováděna pouze přidáním další položky do horní části. Pro tuto aktualizaci jsem vytvořil funkci, která využívá funkci API:Edit s parametrem „section=new“ (vytvoření nové sekce na začátku stránky). Kontrola stránky se neprovádí.

Nové záznamy jsou vkládány ve formátu: „Nezpracovaný e-mail pro“, adresa příjemce (jméno příjemce), „od“, adresa odesilatele (jméno odesilatele), „s předmětem:“, text předmětu, „ze dne“, dnešní datum následované odkazem na celý text e-mailu, který je umístěn na samostatné stránce.

V případě, že není jméno příjemce nebo odesilatele z hlavičky e-mailu načteno, na stránku se nekládá, aby zde nebyly prázdné závorky „()“.

4.2.9 Odeslání e-mailu s chybou

Tak jako v každé aplikaci se může stát, že dojde k selhání připojení k databázi nebo k jinému problému, který bude třeba řešit. Aby bylo možné problém opravit, musí se o něm správce systému dozvědět. K tomuto účelu jsem vytvořil modul, který v případě vyplnění e-mailové adresy správce a správného nastavení, odešle v případě chyby jednoduchou zprávu s jejím popisem.

4.2.10 Nastavení

Soubor s nastavením slouží pro konfiguraci celé aplikace. Jsou zde možnosti konfigurace webové adresy serveru s MediaWiki, klíčová slova na různých wiki stránkách a další nastavení různých podpůrných proměnných.

Jednou z proměnných je například *ladění*. Jedná se o logickou proměnnou, pomocí které se ovládá zapnutí či vypnutí ladících a kontrolních výpisů při generování a aktualizaci wiki stránek.

Důležitým je také vypnutí chybových hlášení nebo nastavení maximálního času provádění skriptů na neomezený.

4.2.11 Zabezpečení proti zneužití zdrojových souborů

Tato poměrně důležitá funkce je ovládána z nastavení aplikace. Je zde definována proměnná *zobrazeniVkladanych*, která musí být pro otevření modulu v prohlížeči nastavena na požadovanou hodnotu (jedná se o náhodně vygenerovaný řetězec s malými a velkými písmeny i čísly). Je-li tato proměnná v požadovaném modulu nastavena na správnou hodnotu, provede se jeho kód. V opačném případě dojde k přesměrování na stránku fakulty.

4.2.12 Centrální ovládání aplikace

Tímto modulem jsou spouštěny všechny implementované funkce. Provádí se tedy nejprve aktualizace skupin projektů, dále aktualizace stránek s projekty, stránek s odkazy na výkazy, týdenními výkazy a posledním úkonem je kontrola e-mailové schránky.

V případě, že dojde k chybě a je nastaven e-mail správce, je mu tato chyba odeslána (pokud nebude problém s připojením k Internetu).

5 Testování a ladění

V této kapitole se budu zabývat testováním jednotlivých modulů i celého systému. Nejprve se zaměřím na měření času, potřebného k vytvoření všech wiki stránek, které je možné z aktuálního stavu databáze získat. V další části se budu věnovat testování aktualizací jednotlivých částí stránek a úmyslným vkládáním chybných informací za některá klíčová slova.

Pro testování jsem používal notebook s 64bitovým operačním systémem Microsoft Windows 7 Professional, procesorem Intel Core i3 M370 @ 2.4 GHz a pamětí RAM 4 GB.

5.1 Založení stránek v MediaWiki

Před těmito testy jsem smazal všechny stránky z aplikace MediaWiki a měřil čas, za jak dlouho budou vytvořeny. Pro věrohodnost výsledků, bylo měření provedeno dvakrát.

typ stránek	počet	čas zpracování	
<i>skupiny projektů</i>	16	10 s	12 s
<i>projekty</i>	175	2 min, 30 s	2 min, 29 s
<i>odkazy na výkazy</i> +	+	30 min, 6 s	29 min, 41 s
<i>týdenní výkazy</i>	2529		
celkem	2895	32 min, 46 s	32 min, 22 s
průměrný čas na jednu stránku	1	68 ms	67 ms

Tabulka 3 - Test 1: Založení všech stránek

Z tabulky 3 je možné vyčíst, že bylo v tomto testu vygenerováno 16 stránek skupin projektů za 10 sekund, 175 stránek projektů za 2 minuty a 29 sekund. Stejně množství jako stránek projektů je samozřejmě i stránek s odkazy na výkazy, které se generují zároveň s týdenními výkazy. Potřebný čas k vytvoření těchto stránek byl 29 minut a 41 sekund. Celkově bylo vytvořeno 2895 wiki stránek s různě objemným obsahem. Průměrný čas vytvoření a vložení jedné stránky byl asi 67 ms.

Zakládání nových stránek se bude pravděpodobně provádět každý týden, protože alespoň někteří řešitelé budou postupně vkládat svoje výkazy. Po vložení nového týdenního výkazu, respektive prvního denního v daném týdnu, se vytvoří nová stránka s týdenním výkazem, která se bude během týdne pouze aktualizovat.

Asi nejdelší čas zabírá vkládání stránky do MediaWiki. Testování jsem prováděl lokálně, a tak zpoždění mezi přenosy z databáze do MediaWiki byla minimální. Při reálném používání běží ale každá část na jiném serveru, a tak může dojít ke zvýšení doby provádění skriptů.

5.2 Aktualizace nově vytvořených stránek

Pro obsah aplikace MediaWiki, vytvořený v předchozí podkapitole, jsem spustil aktualizací skripty. Výsledkem byla samozřejmě nula aktualizovaných stránek, ale šlo o zjištění, jak dlouho bude takováto aktualizace trvat. Časově nejnáročnější operací zde může být opět komunikace mezi servery, protože obsah kontrolované stránky je nutné nejprve stáhnout. V případě, že by se musely některé stránky aktualizovat, musí se nahrát jejich nový obsah zpět do MediaWiki.

typ stránek	počet	čas zpracování		
<i>skupiny projektů</i>	16	2 s	2 s	2 s
<i>projekty</i>	175	39 s	37 s	39 s
<i>odkazy na výkazy</i>	175	12 min, 24 s	12 min, 6 s	12 min, 22 s
+ <i>týdenní výkazy</i>	+ 2529			
celkem	2895	13 min, 5 s	12 min, 45 s	13 min, 3 s
průměrný čas na jednu stránku	1	27.1 ms	25.6 ms	26.2 ms

Tabulka 4 - Test 2: Kontrola všech stránek se správným obsahem

Celkem bylo zkontrolováno 2895 stránek a žádná nebyla aktualizována. Celkový nejkratší čas těchto operací, které jsou složeny ze získání aktuálního obsahu wiki stránky a jeho kontroly s obsahem databáze informačního systému NLPIS, byl 12 minut a 45 sekund. Kontrola jedné stránky se tedy provedla za necelých 26 milisekund.

5.3 Testování aktualizací jednotlivých částí stránek

V této podkapitole se zaměřím na konkrétní výsledky aktualizace nesprávných údajů na wiki stránkách. Testy budu provádět vkládáním chybného obsahu za některá klíčová slova na vybraných stránkách v aplikaci MediaWiki a následně budu spouštět aktualizací skripty pouze pro tuto vybranou stránku a porovnávat získané výsledky.

5.3.1 Testy aktualizace stránky skupiny projektů

Wiki stránka skupiny projektů se skládá z nadpisu, vedoucího skupiny, prostředků, projektů skupiny a ukončení. Testoval a ladil jsem všechny části, pro názornost zde uvedu některé z nich.

5.3.1.1 Aktualizace vedoucího skupiny

Nejprve jsem provedl změnu jména vedoucího na chybnou (viz Obrázek 15).

Po spuštění aktualizace bylo jméno vedoucího nahrazeno správnými daty z databáze informačního systému a stávající text přesunut do poznámky (viz Obrázek 16).

5.3.1.2 Aktualizace tabulky s prostředky skupiny

Vzhledem k tomu, že skupiny jsou nově vytvořeny, není databázová tabulka s prostředky skupiny prozatím naplněna žádnými daty. Na stránce skupiny jsem tedy provedl vložení řádku s neplatnými informacemi (Obrázek 17).

Záměrně jsem zvolil oprávnění pro zápis, aby bylo názorně vidět vložení znaku * za správné oprávnění. Z výsledku aktualizace je patrné, že nedošlo k odstranění neplatného řádku, nýbrž se pouze změnilo oprávnění a řádek byl přesunut na konec tabulky, což zde ale vidět není, protože je v tabulce pouze jeden řádek (Obrázek 18). V případě, že by v této tabulce bylo více řádků, byl by výsledek srozumitelnější.

5.3.1.3 Aktualizace projektů skupiny

Projekty skupiny jsou umístěny pod sebou a každý tvoří zároveň odkaz na vlastní stránku projektu. Provedl jsem zde tři typy kontrol. Přesunutí projektu „Automatické generování wiki stránek z e-mailů a IS“ z prvního místa na konec seznamu, vložení neexistujícího projektu na druhé místo v seznamu a stav projektu „weknowit“ přepsal na K rozhodnutí (viz Obrázek 19).

Aktualizace by tedy měla spočívat v přesunutí projektu „Automatické generování wiki stránek z e-mailů a IS“ na první místo, odsunutí neexistujícího projektu do poznámky a opravení stavu projektu „weknowit“ bez přesunu starého řádku do poznámky. Jak je vidět z obrázku (Obrázek 20), je opravená stránka upravena dle předpokladů.

5.3.2 Testy aktualizace stránky projektu

Na wiki stránce projektu je nadpis, řešitel, vedoucí, zahájení projektu, ukončení řešení, typ projektu, skupina, prostředky, odkaz na pracovní výkazy (stránku s odkazy na výkazy), úkoly (v případě jednoho řešitele) a ukončení. Klíčová slova řešitel a vedoucí se aktualizují stejným způsobem jako vedoucí na stránce skupiny projektů.

5.3.2.1 Aktualizace zahájení projektu

Obsahem klíče je datum přidělení prvnímu řešiteli. Aktualizace zahájení projektu se provádí dvěma způsoby. Jestliže je obsahem datum, provede se obnovení za aktuální. V případě, že obsahem není datum, ale text, vloží se do poznámky.

V testu jsem nastavil datum na 1. 1. 2000 (viz Obrázek 21). Při aktualizaci by se mělo rozpoznat, že jde o datum, a tedy provést pouze nahrazení bez vytvoření poznámky.

Výsledkem testu bylo skutečně pouze opravení data na správnou hodnotu (viz Obrázek 22).

5.3.2.2 Aktualizace ukončení řešení

Obsah klíče ukončení řešení je maximální datum ukončení všech řešitelů projektu. Pokud nemá některý řešitel stanoveno datum ukončení, není zde napsáno nic. Aktualizace se provádí stejně jako u zahájení projektu. Jako ilustrativní testovací případ jsem zde použil vložení textu místo správného data, což je vidět na obrázku (Obrázek 23).

Výsledek aktualizace, který je možné zkontrolovat na obrázku (Obrázek 24), by tedy měl být takový, že se text vloží do poznámky a nahradí se správným datem.

5.3.2.3 Typ projektu

Obsahem tohoto klíče je jeden ze čtyř typů projektu (Obecný projekt, Bakalářská práce, Diplomová práce, Disertační práce). Při aktualizaci se při nenalezení správného obsahu, vždy vytváří poznámka, aby bylo zřejmé, jaký byl typ dříve, ale také se tímto uchovávají poznámky uživatelů.

Při testu jsem místo správného obsahu vložil komentář (viz Obrázek 25). Aktualizovaná stránka bude obsahovat klíčové slovo typ projektu následované správným obsahem, pod který se vygeneruje poznámka, kam se přesune chybný obsah (Obrázek 26).

5.3.2.4 Skupina projektu

Skupina projektu tvoří odkaz na skupinu projektů, kam daný projekt patří, nebo nepatří do žádné skupiny a pak je zde znak pomlčka. Při aktualizaci se najde správný obsah, nebo je vytvořena poznámka, kam je stávající obsah odsunut a je nahrazen správným.

Test jsem provedl vložím pomlčky místo správné skupiny (Obrázek 27). Aktualizace proběhla správným způsobem (Obrázek 28).

5.3.2.5 Aktualizace prostředků projektu

Prostředky projektu jsou tvořeny i aktualizovány stejným způsobem jako na stránce skupiny projektů, kde ale nebylo zřetelné, jak výsledek dopadl. Aby byl vidět přesun v tabulce i změna oprávnění, vložil jsem na první místo neexistující prostředek projektu (Obrázek 29).

Dle předpokladů byl přidán prostředek odsunut na konec tabulky a jeho oprávnění bylo dle tabulky 1 změněno na žádné*.

5.3.2.6 Test změny počtu řešitelů

V tomto případě jsem testoval změnu z jednoho řešitele projektu na dva. Toto je klasická operace, která se bude provádět, pokud bude další řešitel navazovat na prvního. Jako výchozí wiki stránku jsem použil projekt Automatické generování wiki stránek z e-mailů a IS (Obrázek 32). V databázi informačního systému jsem k tomuto projektu přidal dalšího řešitele.

Aktualizace stránky projektu po přidání řešitele se provádí tak, že se převede klíčové slovo „řešitel“ na „řešitelé“ a případné poznámky se vloží do poznámky k novému klíči. Další změnou jsou úkoly řešitele, které jsou pro jednoho řešitele umístěny na stránce projektu. Aby bylo možné úkoly přemístit na správnou stránku s úkoly, je nutné zjistit, zda se úkoly shodují s obsahem databáze. V případě, že se úkoly ze stránky projektu naleznou u jednoho z řešitelů, přesunou se na jeho stránku s úkoly včetně poznámek. Výsledek tohoto testu je na obrázku (Obrázek 31).

5.3.3 Testy aktualizace stránky s odkazy na výkazy

Na stránce s odkazy na výkazy je pro každý týden, kdy alespoň jeden řešitel projektu vykáže nějakou činnost, generován zkrácený týdenní výkaz, který se skládá z hlavičky týdne, jména řešitele, což je zároveň odkaz na kompletní týdenní výkaz daného řešitele, odpracovaných hodin a několika počátečních slov souhrnu. Z odpracovaných hodin a souhrnu je spočten hash, který je umístěn za jménem řešitele.

Při aktualizaci výkazu se při neshodě aktuálního textu na stránce a požadovaného obsahu z databáze provede kontrola, zda text na stránce odpovídá kontrolní hodnotě spočtené funkcí MD5. Jestliže hash odpovídá, provede se nahrazení, jinak se vloží starý výkaz do poznámky a nahradí se novým.

Provedl jsem nejprve test, kdy jsem upravil výkaz na stránce (Obrázek 33). Výsledkem tedy bude vytvoření poznámky a vložení správného výkazu (Obrázek 34), protože bude rozdílný text a hash.

Druhý test, který jsem vyzkoušel, bylo upravení dat v databázi a tím pádem na stránce s odkazy na výkazy byl správný hash a výkaz se mohl nahradit (Obrázek 36). Původní obsah je vidět na obrázku (Obrázek 35).

5.3.4 Testy aktualizace stránky s týdenním výkazem

Na stránce s týdenním výkazem je možné aktualizovat všechny jeho části, kromě nadpisu, který se při nenalezení vkládá nový. Podnadpis a kotvy na denní výkazy se při nalezení klíčového slova nahrazují správným obsahem, bez vytváření poznámek. Kromě reakcí na problémy a nejasnosti a reakcí na výkaz mají všechna klíčová slova spočten hash, podle kterého je možné porovnávat změny v textu na stránce provedené uživatelem od pouhé změny v databázi informačního systému NLPIS.

Na této stránce jsem otestoval všechny položky, ale protože se u údajů se spočteným řetězcem MD5 provádí aktualizace stejným způsobem jako na stránce s odkazy na výkazy, uvedu zde pouze aktualizaci podnadpisu. Podnadpis stránky se skládá z klíčového slova „Týdenní výkaz řešitele“, následovaného jménem řešitele, identifikací týdne a celkovým počtem odpracovaných hodin za daný týden. Ve vybraném testu jsem provedl aktualizaci odpracovaných hodin (Obrázek 37) na správné údaje (Obrázek 38).

5.4 Výsledky testování

Díky provedeným testům jsem ve svojí implementaci mohl odhalit několik drobných nesrovnalostí, které jsem následně opravil. Většina kódu byla již ale před tímto testováním odladěna a byly tak řešeny jen maličkosti.

V návaznosti na prováděné generování a aktualizace 2985 wiki stránek, byla nalezena i jedna závažnější chyba u stránek se stejným URL. Jednalo se o problém, který spočíval v tom, že se projekty na stránce navzájem přerovnávaly. Proto bylo třeba provést úpravu dotazu do databáze informačního systému, kde se pouze muselo vložit řazení podle id projektu a tím byl problém vyřešen.

Ve výsledné implementaci nebyla dle výsledků testů zjištěna žádná závažná chyba.

Závěr

V rámci semestrálního projektu jsem prostudoval jazyky a prostředky pro tvorbu webových informačních systémů, zejména jazyky PHP, HTML, JavaScript, databázi MySQL, dále jsem prostudoval aplikaci MediaWiki a informační systém výzkumné skupiny zpracování přirozeného jazyka NLPIS.

Následně jsem navrhl rozšíření pro stávající systém NLPIS, které bude automaticky generovat a aktualizovat webové stránky v aplikaci MediaWiki. Například jsem navrhl tabulku s prostředky projektu či skupiny projektů, ze které bude možné vyčíst i historii přiřazených prostředků, což zatím nebylo možné nijak jednoduše zjistit. Na některých stránkách se bude uchovávat historie všech změn, naopak na stránkách s odkazy na výkazy a týdenními výkazy se bude používat při aktualizaci hashovací funkce, pomocí které se bude porovnávat, zda uživatel dopsal do stránky nějaký komentář, nebo je možné starý text nahradit aktuálním.

V letním semestru jsem pracoval na implementaci jednotlivých skriptů pro automatické generování wiki stránek z informačního systému NLPIS a z přijatých e-mailů. Nejdříve jsem se zabýval stránkou projektu, která je pro celý systém stránek stěžejní a ze které je možné přejít na stránku skupiny projektů, kam daný projekt patří, nebo na stránku s odkazy na pracovní výkazy, která obsahuje zestručněné týdenní výkazy a odkazy na týdenní výkazy jednotlivých řešitelů projektu.

Při nasazení implementovaného rozšíření budou vytvořeny automaticky generované stránky projektů, které wiki stránku prozatím nemají. Ostatní projekty v aplikaci MediaWiki budou automaticky zkontrolovány na výskyt klíčových slov a případně budou doplněny chybějící údaje. V nastavení aplikace se před prvním spuštěním nakonfigurují požadovaná data jako URL MediaWiki, login a heslo správce, či údaje potřebné k připojení k e-mailové schránce. Vytvořené rozšíření bude v průběhu letošního roku nasazeno ve výzkumné skupině zpracování přirozeného jazyka.

Měřil jsem nejdříve čas generování 2985 wiki stránek, který byl 32 minut a 22 sekund, což je 67 milisekund na jednu stránku. Následně jsem měřil i čas aktualizace vygenerovaných stránek, jenž trval 12 minut a 25 sekund, neboli necelých 26 milisekund na jednu stránku. V kapitole testování a ladění jsem se ale také zaměřil na aktualizaci jednotlivých klíčových slov na několika typech generovaných wiki stránek. Výchozí stav i výsledky jsem řádně popsal a pro lepší dokreslení skutečnosti jsem vytvořil obrázkovou přílohu, kam jsem vstupy a výstupy testování uložil.

Do budoucna je možné aplikaci rozšířit o implementaci kontroly poslední aktualizace wiki stránky projektu. Po nějakém čase bez aktualizace stránky řešitelem je možné řešitele nebo případně i vedoucího projektu informovat e-mailem o neaktuálnosti wiki stránky, jelikož stránka projektu by měla být řešitelem pravidelně aktualizována. Dále připadá v úvahu vytvoření přímého přecházení mezi jednotlivými týdenními výkazy (přechod na předchozí a následující výkaz) v rámci řešitele projektu.

Literatura

- [1] BOS, B.: *CSS*. [online]. W3C, ©2010, poslední úpravy 6. 11. 2010 [cit. 2010-11-25]. Dostupné na WWW: <<http://www.w3.org/Style/CSS/>>.
- [2] CLARY, B., EICH, B., PONOMAREV, N.: *Javascript*. [online]. Mozilla Developer Network ©2010, poslední úpravy 4. 11. 2010 [cit. 2010-11-23]. Dostupné na WWW: <https://developer.mozilla.org/en/JavaScript_Language_Resources>.
- [3] DELLWIG, E., DELLWIG, I.: *JavaScript příručka programátora*. Praha: Grada Publishing a.s., 2003, ISBN 80-247-0298-3.
- [4] DYTRYCH, J.: *Nápověda v NLPIS*. [online]. NLPIS, ©2010 [cit. 2010-12-19]. Dostupné na WWW: <<http://nlp.fit.vutbr.cz/nlpis/index2.php>>.
- [5] LEE, J., WARE, B.: *Open Source – vývoj webových aplikací*. Praha: Mobil Media a.s., 2003, ISBN 80-86593-43-6.
- [6] LINHART, P.: *Apache & PHP*. Praha: Neokortex spol. s.r.o., 1998, ISBN 80-902230-2-8.
- [7] PÍSEK, S.: *HTML a XHTML*. Praha: Grada Publishing a.s., 2003, ISBN 80-247-0571-0.
- [8] STANÍČEK, P.: *CSS Kaskádové styly: Kompletní průvodce*. Brno: Computer Press a.s., druhé vydání, 2003, ISBN 80-7226-872-4.
- [9] TACHOVSKÁNET: *Skriptovací jazyk PHP*. [online]. ©2010, poslední úpravy 15. 1. 2010 [cit. 2010-11-23]. Dostupné na WWW: <<http://www.tachovska.cz/server/php.php>>.
- [10] VAVŘÍNEK, A.: *Interaktivní výuka pro U3V*. Brno, 2009, bakalářská práce, FIT VUT v Brně.
- [11] VRÁNA, J.: *PHP Okénko: Zobrazení pošty na webu*. [online]. ROOT.CZ, ©2006, poslední úpravy 20.4.2006 [cit. 2010-12-12]. Dostupné na WWW: <<http://www.root.cz/clanky/php-imap/>>.
- [12] *File Transfer Protocol*. [online]. Wikipedie, ©2011, poslední úpravy 10. 4. 2011 [cit. 2011-05-07]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/File_Transfer_Protocol>.
- [13] *HTTPS*. [online]. Wikipedie, ©2011, poslední úpravy 3. 4. 2011 [cit. 2011-05-07]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/HTTPS>>.
- [14] *HyperText Markup Language*. [online]. Wikipedie, ©2010, poslední úpravy 11. 10. 2010 [cit. 2010-12-06]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/HyperText_Markup_Language>.
- [15] *HyperText Transfer Protocol*. [online]. Wikipedie, ©2011, poslední úpravy 29. 4. 2011 [cit. 2011-05-07]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/HTTP>>.
- [16] *Internet Message Access Protocol*. [online]. Wikipedie, ©2010, poslední úpravy 23. 11. 2010 [cit. 2010-11-25]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/Internet_Message_Access_Protocol>.

- [17] *JavaScript*. [online]. Wikipedie, ©2010, poslední úpravy 20. 11. 2010 [cit. 2010-11-23]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/JavaScript>>.
- [18] *Kaskádové styly*. [online]. Wikipedie, ©2010, poslední úpravy 15. 11. 2010 [cit. 2010-11-25]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/Cascading_Style_Sheets>.
- [19] *LaTeX*. [online]. Wikipedie, ©2010, poslední úpravy 25. 11. 2010 [cit. 2011-01-05]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/LaTeX>>.
- [20] *MediaWiki - Manual:FAQ*. [online]. MediaWiki, ©2010, poslední úpravy 3. 12. 2010 [cit. 2010-12-06]. Dostupné na WWW: <<http://www.mediawiki.org/wiki/Manual:FAQ>>.
- [21] *MediaWiki*. [online]. MediaWiki.cz, ©2010, [cit. 2010-12-06]. Dostupné na WWW: <<http://www.mediawiki.cz/>>.
- [22] *MediaWiki*. [online]. Wikipedie, ©2010, poslední úpravy 24. 10. 2010 [cit. 2010-12-06]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/MediaWiki>>.
- [23] *Message-Digest Algorithm*. [online]. Wikipedie, ©2010, poslední úpravy 6. 12. 2010 [cit. 2010-12-07]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/Message-Digest_algorithm>.
- [24] *MySQL: MySQL Documentation: MySQL Reference Manuals*. [online]. MySQL, ©2010, [cit. 2010-11-25]. Dostupné na WWW: <<http://dev.mysql.com/doc/>>.
- [25] *MySQL*. [online]. Wikipedie, ©2010, poslední úpravy 9. 11. 2010 [cit. 2010-11-25]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/MySQL>>.
- [26] *Perl*. [online]. Wikipedie, ©2010, poslední úpravy 5. 1. 2011 [cit. 2011-01-06]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/Perl>>.
- [27] *PHP Manual*. [online]. The PHP Group, ©2010, poslední úpravy 12. 11. 2010 [cit. 2010-11-23]. Dostupné na WWW: <<http://php.net/manual/en/index.php>>.
- [28] *PHP*. [online]. Wikipedie, ©2010, poslední úpravy 17. 11. 2010 [cit. 2010-11-23]. Dostupné na WWW: <<http://cs.wikipedia.org/wiki/Php>>.
- [29] *Secure Sockets Layer*. [online]. Wikipedie, ©2011, poslední úpravy 4. 4. 2011 [cit. 2011-05-10]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/Secure_Sockets_Layer>.
- [30] *Transport Layer Security*. [online]. Wikipedie, ©2011, poslední úpravy 22. 3. 2011 [cit. 2011-05-10]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/Transport_Layer_Security>.
- [31] *Uniform Resource Locator*. [online]. Wikipedie, ©2011, poslední úpravy 2. 5. 2010 [cit. 2011-05-08]. Dostupné na WWW: <http://cs.wikipedia.org/wiki/Uniform_Resource_Locator>.
- [32] *Wiki markup*. [online]. Wikipedie, ©2010, poslední úpravy 15. 12. 2010 [cit. 2011-01-03]. Dostupné na WWW: <http://en.wikipedia.org/wiki/Wiki_markup>.

Přílohy

Seznam příloh:

A. Testovací vstupy a výstupy

A. Testovací vstupy a výstupy

Vedoucí skupiny: zde bude vedoucí skupiny

Obrázek 15 - Chybné jméno vedoucího skupiny

Vedoucí skupiny: doc. RNDr., Ph.D. Pavel Smrz
Poznámky k Vedoucí skupiny: zde bude vedoucí skupiny

Obrázek 16 - Opravené jméno vedoucího skupiny

Prostředky skupiny:

název	cesta	oprávnění	patří k projektu
zkratka_prostredku_skupiny	/mnt/minerva1/nlp/projects/zkratka_prostredku_skupiny	zápis	Testovací prostředek skupiny

Poznámka k oprávnění s *: dříve bylo oprávnění zápis

Obrázek 17 - Testovací řádek v prostředcích skupiny

Prostředky skupiny:

název	cesta	oprávnění	patří k projektu
zkratka_prostredku_skupiny	/mnt/minerva1/nlp/projects/zkratka_prostredku_skupiny	žádné*	Testovací prostředek skupiny

Poznámka k oprávnění s *: dříve bylo oprávnění zápis

Obrázek 18 - Opravený řádek s prostředky skupiny

Projekty skupiny:

- [weknowit \(weknowit - K rozhodnutí\)](#)
- [Neexistující projekt](#)
- [Pokusný projekt se stejnými zkratkami 1 \(dve_zkratky - Řešený\)](#)
- [NLPIS - Informační systém Skupiny zpracování přirozeného jazyka \(nlpis - Řešený\)](#)
- [Syntaktická analýza přirozeného jazyka \(syntax - Řešený\)](#)
- [Editor anotací pro WYSIWYG textové editory napsané v jazyce JavaScript \(wysiwyg - Řešený\)](#)
- [Úprava standardních unixových nástrojů, aby dokázaly pracovat se národním nastavením \(cscoreutils - K rozhodnutí\)](#)
- [Offline RSS čtečka \(offline_rss - Ukončený\)](#)
- [Informační technologie v psychologii \(psychosemantika3 - Ukončený\)](#)
- [Pokusný projekt se stejnými zkratkami 2 \(dve_zkratky - Nezadaný\)](#)
- [Automatické generování wiki stránek z e-mailů a IS \(wiki_z_nlpis - Řešený\)](#)

Obrázek 19 - Chybný seznam projektů skupiny

Projekty skupiny:

Automatické generování wiki stránek z e-mailů a IS (wiki_z_nlpis - Řešený) [↗](#)

weknowit (weknowit - Řešený) [↗](#)

Pokusný projekt se stejnými zkratkami 1 (dve_zkratky - Řešený) [↗](#)

NLPIS - Informační systém Skupiny zpracování přirozeného jazyka (nlpis - Řešený) [↗](#)

Syntaktická analýza přirozeného jazyka (syntax - Řešený) [↗](#)

Editor anotací pro WYSIWYG textové editory napsané v jazyce JavaScript (wysiwyg - Řešený) [↗](#)

Úprava standardních unixových nástrojů, aby dokázaly pracovat se národním nastavením (cscoreutils - K rozhodnutí) [↗](#)

Offline RSS čtečka (offline_rss - Ukončený) [↗](#)

Informační technologie v psychologii (psychosemantika3 - Ukončený) [↗](#)

Pokusný projekt se stejnými zkratkami 2 (dve_zkratky - Nezadaný) [↗](#)

Poznámky k **Projekty skupiny**:

Neexistující projekt [↗](#)

Obrázek 20 - Opravený seznam projektů skupiny

Zahájení projektu: 1. 1. 2000

Obrázek 21 - Chybné zahájení projektu (datum)

Zahájení projektu: 10. 06. 2010

Obrázek 22 - Opravené zahájení projektu (datum)

Ukončení řešení: na řešení se stále pracuje

Obrázek 23 - Chybné ukončení řešení (text)

Ukončení řešení: 01. 07. 2011

Poznámky k Ukončení řešení: na řešení se stále pracuje

Obrázek 24 - Opravené ukončení řešení (text)

Typ projektu: Diplomová práce, na které se stále pracuje

Obrázek 25 - Chybný typ projektu

Typ projektu: Diplomová práce

Poznámky k Typ projektu: Diplomová práce, na které se stále pracuje

Obrázek 26 - Opravený typ projektu

Skupina: -

Obrázek 27 - Chybná skupina projektu

Skupina: misc

Poznámky k Skupina: -

Obrázek 28 - Opravená skupina projektu

Prostředky:

název	cesta	oprávnění	patří k projektu
neexistujici_prostredok	/mnt/minerva1/nlp/projects/neexistujici_prostredok	zápis	Neexistující projekt
automaticky_thesaurus	/mnt/minerva1/nlp/projects/automaticky_thesaurus	čtení	Automatický thesaurus
ma	/mnt/minerva1/nlp/projects/ma	čtení	Morfologický slovník a morfologický analyzátor pro češtinu
souvislosti2	/mnt/minerva1/nlp/projects/souvislosti2	zápis	Hledání souvislostí mezi slovy
valence	/mnt/minerva1/nlp/projects/valence	čtení	Určování slovních vazeb na základě statistiky
to_manatee	/mnt/minerva1/nlp/projects/to_manatee	čtení	Zpracování a indexování rozsáhlých textových dat

Poznámka k oprávnění s *: dříve bylo oprávnění zápis

Obrázek 29 - Chybné prostředky projektu

Prostředky:

název	cesta	oprávnění	patří k projektu
automaticky_thesaurus	/mnt/minerva1/nlp/projects/automaticky_thesaurus	čtení	Automatický thesaurus
ma	/mnt/minerva1/nlp/projects/ma	čtení	Morfologický slovník a morfologický analyzátor pro češtinu
souvislosti2	/mnt/minerva1/nlp/projects/souvislosti2	zápis	Hledání souvislostí mezi slovy
valence	/mnt/minerva1/nlp/projects/valence	čtení	Určování slovních vazeb na základě statistiky
to_manatee	/mnt/minerva1/nlp/projects/to_manatee	čtení	Zpracování a indexování rozsáhlých textových dat
neexistujici_prostredok	/mnt/minerva1/nlp/projects/neexistujici_prostredok	žádné*	Neexistující projekt

Poznámka k oprávnění s *: dříve bylo oprávnění zápis

Obrázek 30 - Opravené prostředky projektu

Automatické generování wiki stránek z e-mailů a IS (wiki_z_nlpis) [editovat]

Řešitelé:

- Bc. Ales Vavrinek (Úkoly řešitele, řešil: 10. 06. 2010 - 01. 07. 2011)
- Ing. Jaroslav Dytrych (Úkoly řešitele, řešil: 01. 08. 2011 - 31. 07. 2011)

Vedoucí: Ing. Jaroslav Dytrych

Zahájení projektu: 10. 06. 2010

Ukončení řešení: 31. 07. 2011

Typ projektu: Diplomová práce

Skupina: misc

Prostředky:

název	cesta	oprávnění	patří k projektu
wiki_z_nlpis	/mnt/minerva1/nlp/projects/wiki_z_nlpis	zápis	Automatické generování wiki stránek z e-mailů a IS

Poznámka k oprávnění s *: dříve bylo oprávnění zápis

Pracovní výkazy

Obrázek 31 - Převod projektu s jedním řešitelem na dva

Automatické generování wiki stránek z e-mailů a IS (wiki_z_nlpis) [editovat]

Řešitel: Bc. Ales Vavrinek
 Vedoucí: Ing. Jaroslav Dytrych
 Zahájení projektu: 10. 06. 2010
 Ukončení řešení: 01. 07. 2011
 Typ projektu: Diplomová práce
 Skupina: [misc](#)

Prostředky:

název	cesta	oprávnění	patří k projektu
wiki_z_nlpis	/mnt/minerva1/nlp/projects/wiki_z_nlpis	zápis	Automatické generování wiki stránek z e-mailů a IS

*Poznámka k oprávnění s *:* dříve bylo oprávnění zápis

[Pracovní výkazy](#)

Úkoly:

1. Seznamte se s jazyky a prostředky pro tvorbu webových informačních systémů
2. Prostudujte MediaWiki a informační systém skupiny zpracování přirozeného jazyka (NLPIS)
3. Navrhněte rozšíření NLPIS a další potřebné skripty pro automatické generování stránek projektů a skupin projektů z NLPIS
4. Navrhněte skripty pro generování wiki stránek z přijatých e-mailů (vyhrazená schránka, jejíž obsah se bude automaticky zpracovávat) tak, aby vytvářely stránky s různými informacemi o projektech se vzájemnými odkazy (řešitelé projektu a jejich úkoly, řešené problémy, historie změn prostředků apod.)
5. Implementujte navržená řešení
6. Zhodnoťte dosažené výsledky, srovnajte zvolené přístupy s alternativními metodami a navrhněte další možná rozšíření a vylepšení do budoucna

Obrázek 32 - Jeden řešitel projektu

Rok 2011, 11. týden (14. 03. 2011 - 20. 03. 2011)
 Řešitel [Bc. Ales Vavrinek](#), odpracováno 40 hodin, 0 minut
Souhrn: Tento týden jsem řešil spoustu věcí

Obrázek 33 - Chybný výkaz na stránce s odkazy na výkazy

Rok 2011, 11. týden (14. 03. 2011 - 20. 03. 2011)
 Řešitel [Bc. Ales Vavrinek](#), odpracováno 40 hodin, 0 minut
Souhrn: úpravy v části aktualizace stránky projektu

Poznámky k výkazu:

Rok 2011, 11. týden (14. 03. 2011 - 20. 03. 2011)
 Řešitel [Bc. Ales Vavrinek](#), odpracováno 40 hodin, 0 minut
Souhrn: Tento týden jsem řešil spoustu věcí

Obrázek 34 - Opravený výkaz na stránce s odkazy na výkazy

Rok 2011, 13. týden (28. 03. 2011 - 03. 04. 2011)

Řešitel Bc. Ales Vavrinek [✉](#), odpracováno 55 hodin, 0 minut

Souhrn: Řešení generování a aktualizace stránky projektu a stránky s úkoly více řešitelů pro projekty se stejnou zkratkou.

Obrázek 35 - Upravení výkazu v databázi IS

Rok 2011, 13. týden (28. 03. 2011 - 03. 04. 2011)

Řešitel Bc. Ales Vavrinek [✉](#), odpracováno 55 hodin, 0 minut

Souhrn: Řešení generování a aktualizace stránky projektu a stránky s úkoly více řešitelů pro projekty se stejnou zkratkou. Provedení testu s aktualizací souhrnu.

Obrázek 36 - Upravený souhrn bez vložení poznámky

Týdenní výkaz řešitele Bc. Ales Vavrinek za 13. týden (28. 03. 2011 - 03. 04. 2011), celkem odpracováno 5 hodin, 5 minut

Obrázek 37 - Chybné odpracované hodiny na stránce s týdenním výkazem

Týdenní výkaz řešitele Bc. Ales Vavrinek za 13. týden (28. 03. 2011 - 03. 04. 2011), celkem odpracováno 55 hodin, 0 minut

Obrázek 38 - Opravené odpracované hodiny na stránce s týdenním výkazem