Slovenská technická univerzita

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ

Študijny program: Informatika

Martin Kováčik

Komunitná katalogizácia informácií

Bakalársky projekt

Vedúci projektu : Ing. Michal Takács Máj, 2006

Akademický rok: 2005/2006

Anotácia

Slovenská technická univerzita
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ
Študijny odbor: **Informatika**

Autor: Martin Kováčik
Bakalársky projekt: Komunitná katalogizácia informácií
Vedúci projektu: Ing. Michal Takács
Dátum odovzdania: Máj 2006

Cieľom tohto projektu je poskytnúť analýzu prístupov ku katalogizácií informácií prostredníctvom komunity. Súčasťou je prehľad technológií a metód určených na katalogizáciu informácií. Katalogizácia informácií v komunitnom prostredí využíva otvorené kolaboratívne prostredia. Tento projekt je zameraný na wiki. V otvorenom prostredí je problematické spravovať obsah z pohľadu relevantnosti a kvality. Výstupom tohto projektu je preto prehľad metód a prístupov k ochrane informácií uložených v otvorenom prostredí wiki.

Anotation

Slovak University of Technology
FAKULTY OF INFORMATICS AND INFORMATION TECHNOLOGY
Degree Course: Informatics

Author:	Martin Kováčik
Bachelor project:	Community-based Catalogization of Information
Project Supervisor:	Ing. Michal Takács
Submission date:	2006, May

Aim of this project is to analyze approaches to information catalogisation in community environment. The project brings an overview of technologies and methods used for information catalogization. Community based catalogization uses open collaborative environments. In this document I focus on wiki based software. Open environment is problematic in terms of information content. Part of project is overview of methods and approaches to protect information stored in open collaborative environments.

Čestne prehlasujem, že som bakalársky projekt vypracoval sám na základe literatúry uvedenej v dokumente a na základe vlastných vedomostí a schopností.

V Bratislave, 17. mája 2006

Obsah

1 Úvod

Táto práca približuje prístupy ku komunitnej katalogizácií informácií, ktoré sú roztrúsené medzi jednotlivými nositeľmi. Analýza prístupu zahŕňa prehľad potrebných technológií, informácií, ktoré je možné katalogizovať, ako aj opis používateľovho prístupu a jeho správanie sa v prostredí, ktoré umožňuje katalogizáciu informácií. Či už prispieva informáciami, ktoré nesie, alebo v tomto prostredí získava potrebné informácie.

Druhá kapitola sa venuje všeobecnému úvodu to tematiky kolaborácie. Opisuje ciele a dôvody kolaborácie, ako aj emergenciu uskladnenia a prepájania informácií. Súčasťou je opis modelov kolaborácie a naznačenie vhodnosti pre jednotlivé aplikácie. Táto časť obsahuje základné informácie o interaktívnej kolaborácií a výhody jej aplikácie na webe.

Tretia kapitola sa venuje otvorenej webovej kolaborácie a vysvetľuje prístup ku kolaboratívnemu získavaniu a vývoju informácií v otvorenom prostredí, ktorý zaviedol typ systémov označovaných wiki. Súčasťou je opis wiki a jeho porovnanie s konvenčnými webovými stránkami. Následne je rozobraný používateľov prístup a problémy spojené s otvorenosťou tohto prostredia.

Štvrtá kapitola práce navrhuje riešenie problémov spojených s wiki a otvorenosťou kolaboratívnych systémov. Uvádza odporúčania a konkrétne postupy a algoritmy na zamedzenie útokov na wiki. Tieto postupy zahŕňajú analýzu vkladaných informácií, ako aj správanie sa používateľa vo wiki. V niektorých prípadoch je uvedená aj odporúčaná reakcia.

Piata kapitola opisuje architektúru implementovaného wiki systému, v ktorom je vytvorené prostredie na implementáciu postupov a algoritmov uvedených v štvrtej kapitole.

2 Kolaborácia v komunitnom prostredí

Kolaborácia, ako taká, je pomerne všeobecný pojem. V súčasnosti neexistuje komplexná definícia kolaborácie [1]. Preto sa pod pojmom kolaborácia budem zaoberať kolaboráciou v prostredí internetu s použitím moderných informačných technológií.

[2] hovorí o kolaborácií ako o cielenej zmene kolaboratívnej entity. Kolaboratívna entita má nestabilnú formu. Príklady zahŕňajú spoločnú tvorbu návrhu, vývoj nápadov, dosiahnutie spoločného cieľa.

Kolaborácia nachádza preto miesto všade tam, kde sa na výsledku podieľa viacero ľudí. Predpokladom pre kolaboráciu je motivácia všetkých zapojených ľudí aby sa podieľali na samotnej kolaborácii.

Motivácia môže mať rôzny charakter. Jedným z možných motivačných faktorov je odmena príspevku. Odmena je realizovaná rôznymi spôsobmi. Či už hmotnými alebo nehmotnými. Kolaborácia je znakom tvorby otvorených a voľných systémov. Tie sú tvorené dobrovoľníkmi a jediná odmena je často krát uznanie v rámci komunity. Člen má lepšie preferencie v rámci komunity používateľov kolaboratívnej entity.

Motiváciou je taktiež vytvorenie entity, ktorú môžu neskôr použiť všetci používatelia podieľajúci sa na kolaborácií. Tento typ kolaborácie je vhodný pri realizácií projektov zameraných na výskum, tvorbu softvéru a i. Takáto motivácia vedie ku kolaborácií aj v komerčnej sfére.

Okrem stimulačných faktorov ovplyvňujúcich kolaboráciu vystupujú do popredia aj problémy, ktoré bránia kolaborácií. [3] opisuje príčiny možného neúspechu wiki ako kolaboratívneho média. Nasledujúce problémy sú všeobecné pre väčšinu známych systémov na podporu kolaborácie.

- Používatelia sa nezhodnú na cieľoch kolaborácie.
- Je zložité rozlíšiť relevantné informácie od nerelevantných.
- Je zložité usmerňovať kolaboráciu (v originále sa autor odkazuje na refaktoring obsahu wiki, avšak tento problém má všeobecnú platnosť). Na usmernenie je potrebné odlíšiť relevantné informácie od nerelevantných.
- Nedostatok a nevhodnosť katalogizačnej štruktúry. Tento fakt odradzuje ľudí od toho aby využívali kolaboratívne médium ako zdroj informácií a radšej hľadajú alternatívne zdroje.

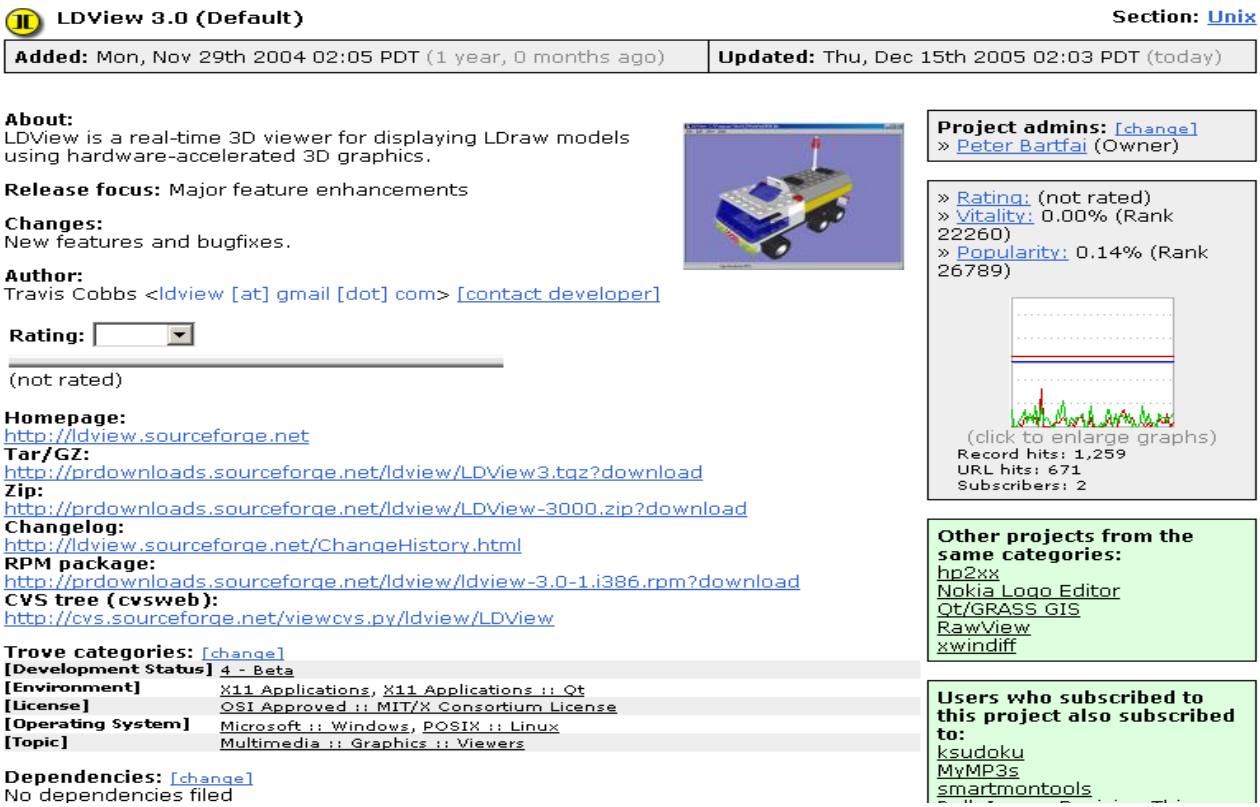
- Nedostatky v rozhraní pre prácu s kolaboratívnym médiom. Neprívetivé používateľské rozhranie pri modifikácie odrádza ľudí aby sa aktívne podieľali na kolaborácií. Tento problém je úzko prepojený s predchádzajúcim bodom.

2.1 Uskladnenie a prepájanie informácií

Uskladnenie informácií je predpoklad ich katalogizácie. Informácie, ale nemusia byť držané spolu s katalógom. Informácie, ktoré sú verejné dostupné, ale len nezaradené nie je potrebné duplicitne ukladať. V tomto prípade sa ukladajú len katalogizačné informácie spolu s prepojením s informáciou. Niektoré vyhľadávače na internete majú, ako jednu zo služieb, katalóg stránok (napr. Zoznam.sk, obr. 1). Katalógy internetových prehliadačov vytvárajú tematický prepojenú hierarchickú štruktúru.

Obrázok 1: Náhľad na titulnú stranu katalógu www.zoznam.sk

Položky v katalógu môžu obsahovať aj doplnkové údaje o danej informácií. Takýto prístup využíva portál freshmeat.net. Tento portál slúži na vyhľadávanie softvéru. Softvér je katalogizovaný podľa kategórií a každá položka nesie aj ďalšie informácie o softvéri a projekte samotnom, nie len odkaz naň (obrázok 2). Odkazy na softvér, ako aj informácie o ňom sú vkladané a aktualizované registrovanými používateľmi.



Obrázok 2: Náhľad na stránku freshmeat.net obsahujúcu katalogizačné informácie

Ďalším rozšírením klasického katalógu odkazov je uloženie samotnej informácie. Ukážkou tohto modelu je server Sourceforge.net. Ten bol primárne vytvorený na podporu otvorených projektov. Pri registrácii projektu je vytvorené aj celé projektové prostredie na manažment projektu a ukladanie súborov.

Systémy, ktoré vytvárajú katalóg definujú určité kategórie pričom ich zaradzujú do stromovej štruktúry. Avšak informácie môžu byť zaradené do viacerých kategórií. Takýto hybridný prístup umožňuje jednoduchšie nájdenie informácií, keďže hľadanú informáciu sprístupňuje z viacerých pohľadov.

Katalógové vyhľadávače vyžadujú zadanie informácie, ktorá má byť neskôr vyhľadávaná. Pre dynamické prostredia je vhodnejší fulltextový vyhľadávač. Fulltextové vyhľadávače periodicky prehľadávajú množinu stránok, hľadajú ich pripojenia a uschovávajú najvýznamnejšie výrazy z prehľadávanej stránky. Toto umožňuje vyhľadávať informácie podľa charakteristických slov.

Fulltextové vyhľadávače sledujú prepojenia medzi stránkami a vytvárajú hodnotenie relevantnosti prehľadávanej informácie. Týmto uprednostňujú relevantnejšie informácie a čiastočne tak riešia problém s nerelevantnými údajmi. Najznámejší reprezentant z kategórie

hodnotenia stránok je algoritmus používaný vo vyhľadávači Google PageRank [4].

Striktné využívanie stromovej štruktúry je smerované na špecifické informácie, medzi ktorými sú presne definované rozdiely. Stromová štruktúra je vhodná na reprezentáciu napr. organizačných informácií a i.

Stromová štruktúra definuje presné kategórie a podkategórie. Postupným zjemňovaním podmienok na informácie je možné pomerne rýchlo nájsť vhodné informácie.

Konkrétna informácia je zaradená v informačnom kontexte. Informačný kontext sprístupňuje informáciu, jej význam a prepojenia s inými informáciami. V prostredí Internetu sú informácie prepájané pomocou hypertextových odkazov.

2.2 Konverzácia ako predpoklad kolaborácie

Elementárny spôsob výmeny informácie je rozhovor. Osobný rozhovor často krát nie je vhodný na výmenu informácií, ktoré majú byť neskôr prístupné komunite. Vhodným rozšírením je zaznamenávať rozhovor alebo použiť prenosové kanály, ktoré záznam priamo podporujú. Záznam je v prípade týchto technológií realizovaný pomocou moderných informačných a telekomunikačných technológií.

V prípade komunikácie medzi fyzicky vzdialenými subjektami je možné využiť telefónne spojenie. Mobilné telekomunikačné siete priniesli možnosť komunikovať pomocou krátkych textových správ. Tento spôsob komunikácie je možné označiť za konverzáciu. Krátke textové správy, alebo aj SMS, je možné v súčasnosti možné integrovať do väčších informačných infraštruktúr. Takéto riešenie sa často krát využíva v podnikovej sfére. Existujúce riešenia podporujú záznam zachytenej konverzácie, ako aj je katalogizáciu podľa rôznych pravidiel.

S príchodom počítačových sieti a Internetu sa objavujú riešenia, ktoré využívajú tento prenosový kanál. V súčasnosti sa tento komunikačný kanál neobmedzuje len na text, ale je možné komunikovať aj pomocou hlasu a obrazu, ako aj celú konverzáciu zaznamenávať a sprístupniť širšiemu okruhu ľudí a systémov.

Revolúciu v konverzácií cez Internet spôsobili aplikácie ktoré priamo simulovali konverzáciu. Príkladom je IRC a podobné technológie umožňujúce textovú *konverzáciu* v takmer reálnom čase.

Ďalším míľnikom je Instant Messaging. Instant Messaging je pomerne populárny pri bežnej

komunikácií ako aj v podnikovej oblasti. V niektorých prípadoch je vhodnejší ako komunikácia poštou, či už elektronickou alebo klasickou. Výhodou je kratší reakčný čas konverzácie, ako aj možnosť záznamu konverzácie. Najznámejšie Instant Messaging systémy sú ICQ, AIM, MSN, AOL, Jabber a aj rôzne menej známe, alebo novšie systémy, napr. Google Talk, ktoré rozširujú funkcionalitu Instant Messagingu o nové vlastnosti a funkcie.

2.3 Modely kolaborácie

Iným predpokladom kolaborácie je aktuálnosť kolaboratívneho média. Spolu s výmenou informácií, na sprístupnenie kolaboratívneho média sa používajú *kolaboračné modely.*

Podľa [5] sú najčastejšie používané tri kolaboračné modely v prostredí Internetu.

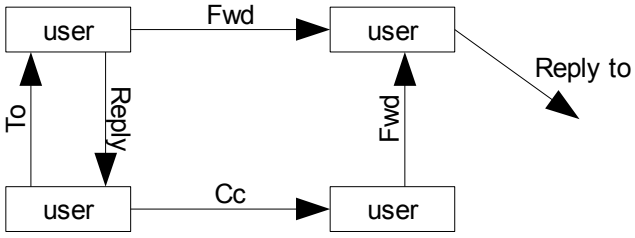
- Komunikácia prostredníctvom elektronickej pošty
- Zdieľané priečinky/zdieľaný prístup ku súborom
- Interaktívne aktualizácie obsahu/interaktívny prístup k informáciám

Iné kolaboračné modely sú oveľa zložitejšie z pohľadu zdieľania informácií a funkcionality v prostredí počítačových sietí a ešte nie sú dostatočne zastúpené, efektívne a prístupné pre bežných používateľov.

2.3.1 Elektronická pošta

Elektronická pošta poskytuje priamu výmenu informácií medzi členmi kolaboratívnej skupiny. Tento spôsob výmeny informácií vyžaduje len schopnosť poslať a prijať emailové správy.

Na výmenu informácií v rámci kolaboratívnej skupiny musí byť emailová správa poslaná každému členovi komunity. Každý má následne vlastnú kópiu informácie. Toto spôsobuje veľké zaťaženie siete, po ktorej sú správy posielané. Prepojenie členov kolaboratívnej skupiny je znázornené na obrázku 3.



Obrázok 3: Prepojenie členov kolaboratívnej skupiny pri komunikácií pomocou elektronickej pošty

Tieto neduhy je možné čiastočne redukovať špeciálnym serverovým softvérom (mailing list managers). Ten príjma správy s určitou emailovou adresou a preposiela ich na emailové adresy ostatných členov kolaboratívnej skupiny. Tieto typy softvéru umožňujú aj archiváciu správ a ich neskoršie prehľadávanie.

Správanie mailing listov je integrované v niektorých emailových klientoch, kde je možné manažovať zoznamy používateľov a adries lokálne.

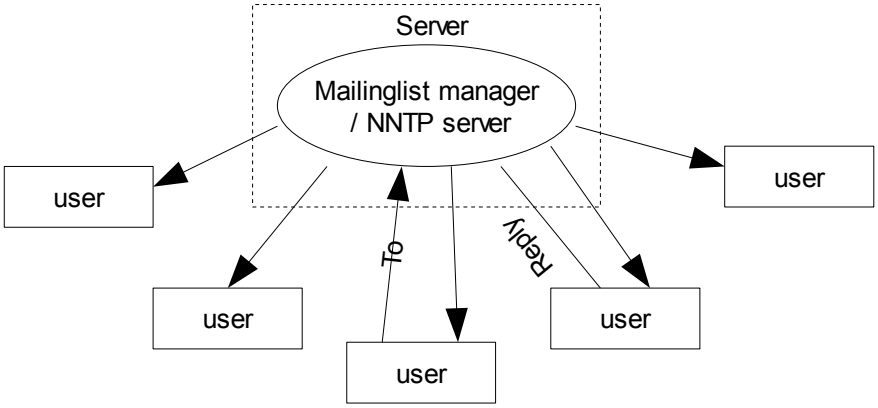
Priame využívanie emailov na kolaboratívnu výmenu informácií má rušivý charakter. Tento charakter je v niektorých prípadoch vhodný. Iné spôsoby zdieľania však využívajú aj iný spôsob notifikácie o aktualizácií.

Priama emailová komunikácia a mailing listy vyžadujú mať lokálnu kópiu informácií. Katalogizácia a práca s údajmi je realizovaná na lokálnej kópií. Aktualizácie na kolaboratívnom produkte musí robiť používateľ alebo program používaný na výmenu emailov sám na základe zaslaných aktualizácií. Aj zasielanie aktualizácií je nutné iniciovať explicitne.

Iný prístup k zdieľaniu elektronickej pošty je použitý pri protokole *News*. Práca s ním je takmer zhodná ako s elektronickou poštou. Informácie sú zdieľané pomocou jedného úložiska elektronickej pošty, ktoré je dostupné všetkým členom komunity.

News server pracuje podobným spôsobom ako mailinglist. Rozdiel je v tom, že maily sa nerozposielajú k jednotlivým členom kolaboratívnej skupiny. Prístup ku kolaboratívnemu médiu je možný priamo prístupom ku serveru. Aktuálne informácie sú ukladané a spravované centrálne. Toto čiastočne zaradzuje news servery k modelu reprezentujúcemu zdieľaný prístup.

Hlavným rozdielom news serverov oproti mailinglistom je, že správy po určitej dobe expirujú a informácie o aktualizácií (príchod novej správy) nie sú rušivo podané používateľovi. Obrázok 4 znázorňuje prepojenie kolaborujúcich členov prostredníctvom mailinglistu/news servera.



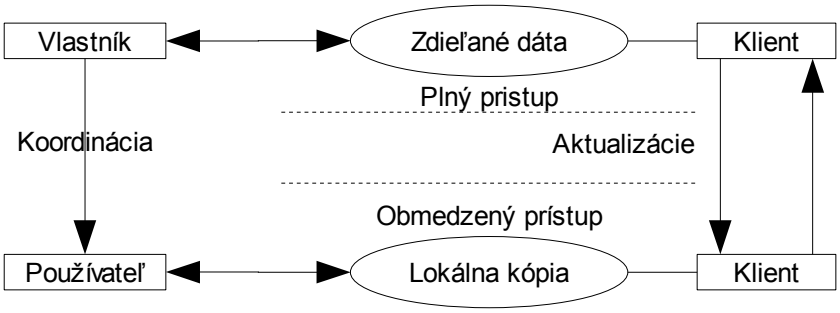
Obrázok 4: Prepojenie používateľov prostredníctvom mailinglistu/news servera. V prípade mailinglistu sú nové správy rozposielané. V prípade news servera sú správy žiadané zo servera.

Nevýhodou tohto modelu je, že jednotlivé správy už nemôžu byť neskôr editované a prepájané medzi sebou. Ak je potrebná posteditácia a prepájanie, je potrebné použiť niektorý z nasledujúcich modelov.

2.3.2 Zdieľaný priečinky/zdieľaný prístup k súborom

Tento model reprezentuje priamy prístup k súborom v spoločnom úložisku dát. Tento model je najvšeobecnejším modelom používaným na kolaboráciu v podnikovej sfére. Problémom tohto modelu je koordinácia modifikácií a aktualizácií. Spolu s týmto modelom sa preto využíva alternatívny komunikačný kanál, napr. Využívajúci prvý model (emailová komunikácia).

Jeden zo spôsobov kolaborácie je znázornený na obrázku 5. Tento spôsob predpokladá priebežnú implicitnú synchronizáciu informácií.



Obrázok 5: Práca so zdieľanými informáciami. Klient zabezpečuje synchronizáciu lokálnej kópie so zdieľanými dátami. Iný informačný kanál slúži na koordináciu kolaborácie.

Niekedy priebežné aktualizácie nie sú žiadúce. Modelovým príkladom je vývoj softvéru. Ak

programátor pristupuje k zdieľaným prostriedkom a iný programátor modifikuje ten istý program, vzniká nekonzistentný projekt, ktorý nebude fungovať. Problémom je možné zamedziť riadenou synchronizáciou.

Existuje pomerne veľké množstvo programových systémov, ktoré umožňujú explicitné iniciovanie aktualizácia so zdieľanými údajmi. Najznámejší zástupcovia sú systémy CVS a SVN. Niektoré firmy tvoria vlastné systémy na manažment verzií špecifických dát. Jedným z príkladov je Microsoft SharePoint, ktorý priamo podporuje dokumenty vyprodukované balíkom Microsoft Office.

Rozšírením týchto systémov je ukladanie priebežných revízií produktov. Revízia je v tomto prípade chápaná ako výsledok explicitne iniciovanej aktualizácie.

2.3.3 Interaktívna aktualizácia a prístup k informáciám

Základom tohto modelu je možnosť viacerých kolaborantov kolektívne editovať určitú informáciu. Podstata je že pri modifikácií je k dispozícií vždy aktuálna verzia. Kolaborácia môže prebiehať synchrónne, aj asynchrónne, zároveň je zabezpečená perzistencia informácie.

Cieľom interaktívnej kolaborácie je informácia. Niektoré príbuzné modely a systémy ale pridávajú aj možnosť zmeny práce s informáciou.

Predpokladom na interaktívnu aktualizáciu a prístup k informáciám je vhodný softvér a cieľové prostredie. Dnešným trendom je sprístupňovanie systémov prostredníctvom Internetu. Informačným kanálom sa stáva hlavne web.

Výhody tohto prístupu sú:

- *Voľný prístup k materiálom a platformová nezávislosť.* Na prístup k informáciám je potrebný len internetový prehliadač a prístup k internetu/intranetu. Web je štandardom pre výmenu informácií, preto je podporovaný veľkým množstvom prostredí.
- *Aktuálna verzia.* Centrálna správa zabezpečuje, že je dostupná vždy aktuálna verzia.
- *Prepájanie pomocou liniek.* Dokumenty je jednoduché prepájať pomocou hyperlinkových odkazov. Takto je vytváraný bohatý informačný kontext, ktorý prepája informáciu s príbuznými informáciami, staršími verziami a i.
- *Zvýrazňovanie štruktúry informácie*. Informácia zobrazená s dôrazom na jej štruktúru zvyšuje jej prezentačnú hodnotu. Umožňuje konverziu do iného formátu a taktiež uľahčuje

orientáciu. Tento cieľ je v prostredí webu dosahovaný pomocou značkovacích jazykov HTML a XML.

Nevýhodou kolaborácie prostredníctvom webu je prevažne obmedzené používateľské rozhranie. Tento problém však ustupuje do pozadia s nástupom moderných webových technológií a štandardov (AJAX, JavaScript, CSS).

Pre úspech webovej kolaborácie je potrebné vyriešiť problém s používateľským rozhraním. To najčastejšie odradí používateľa od aktualizácie kolaboratívneho média.

2.4 Aplikácie interaktívnej kolaborácie

Nasledujúce rozdelenie definuje rôzne oblasti aplikácie. Tieto oblasti sú definované rozdielnymi spôsobmi a rozsahmi prístupu a cieľovou skupinou používateľov.

- *Jednotlivci*, ktorí využívajú kolaboratívne technológie na tvorbu, organizáciu a ukladanie informácií. Ak je kolaboratívny systém prístupný prostredníctvom siete je tak zabezpečená aj uloženie a záloha na inom počítači.
- *Dočasné kolaboratívne skupiny.* Cieľom týchto skupín je kolaborácia s využitím centralizovaných serverových riešení za cieľom realizácie špecifických projektov.
- *Záujmové skupiny.* Cieľom týchto skupín je vedenie interaktívnej diskusie a tvorba projektov. Využívajú serverové riešenia a nasadzované sú na internet aby bol zabezpečený prístup čo najväčšej komunity.
- *Akademické skupiny.* Cieľom je zverejňovanie a zveľaďovanie informácií, ako aj diskusia, a realizácia akademických projektov a pomoc pri výučbe.
- *Firemné aplikácie*. Cieľom je plánovanie, riadenie, tvorba dokumentácie, ako aj samotná tvorba projektov.

3 Otvorená webová kolaborácia - Wiki

Otvorená webová kolaborácia je trend v tvorbe a zveľaďovaní informácií. Koncept, ktorý stojí za úspechom veľkej časti projektov tvorených prostredníctvom otvorenej kolaboratívnej interakcie, je wiki.

Wiki je podľa [5] voľne rozšíriteľná kolekcia prepojených webových stránok. Z implementačného pohľadu je to hypertextový systém na ukladanie a modifikáciu informácií – databáza informácií a ich prepojení.

[6] pojednáva o vlastnostiach wiki. Vlastnosti, ktoré predurčujú wiki ako vhodné otvorené kolaboratívne médium sú

- *Wiki je otvorené*. Akýkoľvek člen komunity môže editovať a pridávať obsah tak aby vyhovoval jeho potrebám.
- *Organické.* Štruktúra a obsah je otvorený pre editovanie a vývoj.
- *Konvergujúce*. Základom celej koncepcie wiki je agilný prístup k publikovaniu a editácií informácie. Informácia je v tomto procese neustále zveľaďovaná a je hodnota narastá. Výsledkom je stabilný dokument, ktorý má vysokú informačnú hodnotu.
- *Dôveryhodné.* Koncepcia wiki sprístupňuje všetky informácie. Akákoľvek zmena je sledovaná a hodnotená používateľmi komunity. Ak sa nájde nevyhovujúca, alebo nesprávna informácia je pozmenená členmi komunity.

3.1 Charakteristika wiki ako webovej aplikácie

Wiki aplikácie majú podobné správanie. Kritéria, ktoré sa kladú na wiki sú

- *Editácia je rýchlo prístupná a editačné nástroje majú jednoduché rozhranie.* Obsah je ukladaný nie priamo v prezentovanom značkovacom jazyku ale v jednoduchšom formáte. Tento formát pripomína prirodzenú formu zápisu neformátovaného textu. Wiki je orientované na obsah a nie na vzhľad, pre je tento zápis postačujúci. Novšie systémy využívajú moderné technológie a sprístupňujú WYSIWYG[1] rozhranie.
- *Využívanie jednoduchého navigačného modelu*. Stránky obsahujú priamo prepojenia definujúce ich informačný kontext. Príbuzné stránky sú takto veľmi rýchlo dostupné „click

1 What You See Is What You Get. Skratka, ktorá označuje editačné rozhranie, ktoré je vzhľadovo rovnaké ako prezentačné prostredie.

away“.

- *Hocikto z definovanej komunity môže zmeniť informáciu.*
- *Rýchle získavanie informácií*. Toto kritérium zahŕňa konvencie používané vo wiki. Najznámejšia je, že meno hypertextovej linky na lokálnu stránku je zhodne s titulkom stránky[2]. Iný spôsob rýchleho získania informácie je implementácie vstavaného vyhľadávania.

Tieto kritéria nepriamo definujú nasledujúce vlastnosti wiki

- *Viacero možných prístupových bodov k danej stránke.*
- *História stránky spolu so zoznamom zmien.*
- *Jednoduchá kolaborácia.*

3.2 Čim je prezentačná časť wiki iná od klasických webových stránok

Hlavným rozdielom oproti konvenčným webovým stránkam, aj keď nie záväzným, je silný dôraz na obsah a nie na formu. Jednou z elementárnych časti sú šablóny, ktoré transformujú stránku z internej reprezentácie na HTML stránku. Toto určuje homogénnosť vzhľadu a konvencie pre zobrazované stránky.

Normálne stránky majú problém aj s odkazmi na neexistujúce stránky. Tento jav je známy pod pojmom „broken link“. Koncepcia wiki priamo podporuje vytváranie odkazov na neexistujúce lokálne stránky. Vo výslednej stránke sú ale tieto linky zvýraznené a po kliknutí priamo sprístupňujú rozhranie umožňujúce vytvoriť danú stránku.

Zvýraznenie sa robí zmenenou farbou hypertextovej linky, alebo pridaním symbolu, ktorý nepatrí do daného kontextu (ikona, '?' pred linkou, a i.).

Podobným spôsobom sa zvýrazňujú aj linky na externé stránky.

Okrem samotných informačných stránok sprístupňujú wiki systémy aj prehľadové informácie. Cieľom zverejnenia týchto informácií je sledovanie zmien na wiki. Sledovanie zmien je dôležitá činnosť pri odhaľovaní potenciálnych útokov na wiki. Vhodné typy zmien zároveň indikujú evolúciu informácií uložených vo wiki a z určitého pohľadu je toto merítko kvality wiki.

2 V prvých wiki systémoch sa linky zapisovali priamo ako mená stránok. Jediné odlíšenie od okolitého textu je že v prípade viacslovného pomenovania sa neuvádzajú medzery a slová sa oddeľujú zmenou veľkostí prvého písmena v slove – tzv. „Camel case“ convencia : Meno „Linka odkazujúca na stránku“ je následne zapísaná ako LinkaOdkazujúcaNaStránku.

Informácie, ktoré sú sprístupňované ako doplnkové zahŕňajú posledné zmeny spolu s pôvodcami, informácie o aktivite pôvodcov zmien, zmeny v organizácií. Prostredníctvom informácií o posledných zmenách môžu používatelia hodnotiť prácu členov komunity. Posledné zmeny sú dôležitým nástrojom pri rozpoznávaní útokov a hodnotenie dôveryhodnosti používateľa a relevantnosti informácie.

Posledné zmeny sú často krát exportované do formátov, ktoré umožňujú výmenu informácií medzi stránkami (RSS, Atom). Takto sa zároveň propagujú najaktuálnejšie informácie dostupné v rámci wiki.

3.3 Pohľad používateľa na publikáciu vo wiki

Wiki predpokladá motiváciu zo strany používateľa. Avšak k ideálnemu stavu sa pridávajú rôzne psychologické a sociologické neduhy. Na jednej strane sú problémy spojené s tým, že ľudia sa obávajú publikácie na wiki. Problémom je strach z publikácie neúplnej, ale nie celkom presnej informácie a následnej negatívnej odozvy zo strany komunity. Na druhej stráne je problém spojený práve s vkladaním a publikovaním nevhodných, nepresných informácií. Oba javy sú v určitých medziach pre wiki prospešné. Akonáhle ale tieto hranice opustia stávajú sa veľkým problémom pre wiki.

V malej miere tieto javy iniciujú vývoj informácie. Obava pred publikáciou informácie vytvorila požiadavku na vytvorenie alternatívneho kolaboratívneho média, ktoré by fungovalo paralelne s wiki. V praxi sa používajú diskusné nástroje. Používateľ pomocou diskusie vie ohodnotiť informácie, ktoré chce publikovať. Ak je cieľom používateľa informácie získať, prostredníctvom diskusného fóra vie definovať požiadavky a to núti ľudí zamyslieť sa nad publikáciou a zohľadniť tam fakty a požiadavky uvádzané v diskusií.

Rušivé a nepresné zmeny v prezentovaných informáciách vedú ku konfliktom v diskusnej časti a opäť iniciujú zohľadnenie týchto zmien pri publikovaní nových revízií stránok.

3.4 Role používateľov vo wiki

Publikácia vo wiki môže konvergovať k rôznym nežiadúcim javom. Preto je potrebné definovať určité obmedzenia, týkajúce sa bežných používateľov. Toto zavádza potrebu vytvorenia viacerých rolí používateľov v systéme.

Konvenčný spôsob je privilegovať používateľov na vykonávanie určitých akcií. Existujúce

wiki systémy podporujúce viacero používateľov definujú často krát rolu bežný používateľ a administrátor.

Bežný používateľ môže

- vytvárať stránky
- aktualizovať obsah (pridávať prílohy – obrázky, dokumenty)
- participovať na diskusií vo wiki

Operácie, ktoré môžu ohroziť wiki a môže ich vytvárať len administrátor sú

- mazanie stránok
- zmena prístupových práv ku stránkam
- blokovanie používateľov

Niektoré wiki systémy diferencujú viacero rolí v systéme. Pričom tie sa nemusia vzťahovať na celé wiki, ale aj na jeho časti. Tým je možné vytvárať administrátorov pre určité stránky a sekcie.

Existuje viacero návrhov práv vo wiki. Jedným z neštandardných nich je správa prístupových práv odzrkadľujúca vývoj informácií vo wiki [7].

3.5 Podpora verifikácie a validácie vo wiki

Verifikácia a validácia vo wiki je v súčasných systémoch implementovaná prostredníctvom pomocných prostriedkov. Ide o stránky, ktoré obsahujú informácie o editačnom procese stránky a zvýrazňujú tam určité body, ktoré môžu editovať nevhodnú zmenu.

Ide o stránky zobrazujúce posledné zmeny vo wiki a históriu stránky. Stránka s poslednými zmenami je prehľadová stránka, ktorá umožňuje prezerať si históriu zmien stránok, avšak pre celý systém.

Pre potreby odhaľovania problémov je potrebná aj analýza správania používateľa. Preto je možné získať aj správanie sa používateľa (uskutočnené zmeny) spolu s novými informáciami, ktoré získali status blokované.

3.6 Útoky spojené s otvorenosťou wiki

[8] definuje typy útokov. Tie je možné zovšeobecniť na nasledujúce typy.

- Mazanie relevantných informácií
- Vkladanie nerelevantného obsahu
- Pozmeňovanie správnych informácií, alebo ich vkladanie jednostranných interpretácií.
- Pridávanie liniek – SPAM
- Porušenie pravidiel určeného konkrétnym wiki
- Dementovanie správnych a hodnotných informácií
- Zmeny zaradenia informácií a premenovávanie stránok

Riešenie problémov s pridávaním a mazaním obsahu sa rieši zablokovaním pôvodcu zmien akonáhle sa zistí že to nie je len omyl, ale cielená činnosť.

Porušovanie pravidiel je riešené zamedzením prístupu pre porušovateľa v prípade, keď je jasné, že používateľ nechce zmeniť prístup k editovaniu obsahu wiki.

3.6.1 Motívy útokov

Motivácia k útokom je rozdielna pre rôzne typy útočníkov. [8] rozlišuje dva typy útočníkov.

- *Troll (originál)*. Voľný preklad hovorí o používateľovi zanieteného pre nesprávne, alebo radikálne názory. Tento pojem sa používa pri diskusných fórach ako označenie človeka, ktorý vkladá provokujúci obsah. V ďalšom texte sa budem na tento pojem odkazovať ako Provokatér.
- *Vandal.* Tento typ útočníka spôsobuje škodu vedome. Rozdiel oproti provokatérovi je v tom, že tento používateľ sa snaží ostať anonymný a nie je ochotný a schopný obhajovať zmeny, ktoré uskutočnil.

Útoky sú motivované:

- *Pohŕdaním celým projektom, na ktorý je vedený útok*
- *Skúšaním porušovania pravidiel*. Útočníci skúšajú, čo všetko im prejde.
- *Snahou o nahnevanie*, alebo skompromitovanie tvorcu projektu, na ktorý sa vedie útok
- *Odoprenie pozornosti od inej veci*
- *Dojmom, že ide len o zábavu.*
- *Finančná motivácia*. V prípade Spamu útočník získava peniaze za šírenie obťažujúcej reklamy.

- *Snaha o vyskúšanie vlastností systému*. Ošetrenie tohto problému je realizované pomocou tzv. Sandbox stránok. Ide o špeciálne stránky, ktoré sú určené na skúšanie vlastností systému.
- *Pripisovanie a zväčšovanie zásluh, prikrášľovanie reality, zmena objektívnych informácií na informácie, prezentujúce osobný názor.* Tento problém je aktuálny pre stránky obsahujúce životopisy osôb, ako aj horúce spoločenské témy.
- *Tzv. „edit wars“*. Tento pojem označuje jav, keď používateľ ruší zmeny druhého používateľa, zatiaľ čo ten ich opakuje. Počas trvania tohto útoku iný používatelia nemôžu uskutočniť zmeny, kedže „bojujúci“ používatelia nie sú ochotní hodnotiť a zahŕňať uskutočnené inými používateľmi.

3.6.2 Ciele útokov

Cieľavedomé útoky na wiki sú vedené prevažne na stránky, ktoré obsahujú kontroverzné informácie. Útočníci sa taktiež zameriavajú na veľké wiki systémy, ku ktorým pristupuje veľké množstvo ľudí zo skupiny, ktorú chce útočník osloviť.

Malé systémy väčšinou nemajú veľký problém s útokmi. Ich ochrana je často krát založená len na tom, že sú menej známe a viditeľné. Útoky na malé wiki systémy sú prevažne realizované automaticky s cieľom zničiť obsah wiki.

3.6.3 Rozsah útokov

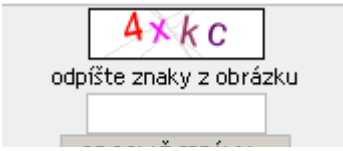
Na celé systémy a podsystémy sú smerované prevažne automatizované útoky. Automatizované útoky nie sú vhodné na modifikáciu jednotlivých informácií. Podporné systémy nerealizujú priamy prístup ku informáciám, ale slúžia na podporu kolaborácie. Sú to napríklad diskusné fóra a mailing listy. Na jednotlivé stránky sú útoky vedené prevažne jednotlivcom alebo skupinou.

3.6.4 Programové útoky

Určité typy útokov sú realizované pomocou programov. Takýto útok je charakteristický skôr rozsahom ako vkladanými informáciami. Útočný program prechádza stránku a hľadá prepojenia na iné stránky s tým získava databázu stránok na danom wiki. Následne iniciuje ich editáciu a uloží obsah.

Riešením sú testy, ktoré sú schopné rozpoznať človeka od stroja. Jeden z prístupov aplikovaných v praxi je realizovaný pomocou tvorby testov, ktoré človek prejde, ale stroj nie (CAPTCHA[3] [11]).

Na internete sa je táto ochrana prevažne realizovaná vloženia obrázku s chaoticky napísaným textom, ktorý je potrebné zaznamenať do formulára spojeného s editačným formulárom. Príklad tohto ochranného prvku je na obrázku.



Obrázok 6: Ochrana pred automatizovaným vkladaním - obrázok s textom (blog.sme.sk)

Text uvádzaný na obrázku je vykreslený takým spôsobom aby nebolo možné ho ľahko rozpoznať programovo. Projekt CAPTCHA sa neobmedzuje na text, ale zobrazuje aj rôzne abstraktné obrazce a od používateľa žiada vytvorenie prepojení vstupov. Viac príkladov testov je možné nájsť na stránke projektu.

Ľahko pozorovateľnou charakteristikou programových útokov je rýchly sled zmien. Ochranou môže byť čakanie určitú dobu, počas ktorej používateľ/program nemôže zmeniť danú stránku. Normálny používateľ toto obmedzenie, v prípade nastavenia vhodne dlhého čakania, nepocíti. Keďže nie je schopný v rýchlom slede pridávať a modifikovať obsah na wiki.

3.6.5 Zmeny informácií v prospech editujúcich

Tento jav sa prejavuje keď používateľ, ktorý edituje stránku chce zlepšiť svoje referencie a to aj za cenu vloženia nesprávnych informácií. Tieto typy útokov nie je možné ľahko vystopovať a je potrebná spoluúčasť komunity.

3.6.6 Spam

V súčasných wiki systémoch ochrana pred spamom nie je dobre integrovaná. Ochrana sa obmedzuje na filtrovanie obsahu na základe výskytu určitých fráz. Iným spôsobom je podpora riadenia prístupu k jednotlivým stránkam. To je realizované jednoduchými právami umožňujúcimi editovanie stránky, alebo komplexným systém na riadenie prístupu.

3 Completely Automated Public Turing Test to Tell Computers and Humans Apart (www.captcha.net)

Komplexné systémy na riadenie prístupu sú používane na blokovanie IP adries sietí/počítačov, ktoré sú známe tým, že zasielajú spam. Ďalším parametrom prístupu môže byť používateľské meno, DNS názov, emailová adresa.

Jedna komunita je často krát prepojená s viacerými wiki. Tieto wiki systémy bývajú prepojené. Prepojenie je realizované automaticky pomocou kanálov RSS a iných podobných systémov. Krátke informácie z wiki systému sú automaticky zobrazované na wiki, ktoré využíva RSS. V prípade, že je wiki, ktoré má takýto systém, napadnuté, sú útočné informácie zverejnené aj na ostatných wiki. Ochrana pred spamom je preto potrebná aj na tomto mieste v systéme.

Jednou z možných ochrán pred spamom je zabránenie jeho šírenia prostredníctvom internetu. Myšlienkový predpoklad je, že spamer demotivovaný neúspechom šírenia prestane vytvárať spamové odkazy. Iné riešenia využívajú analýzy a vyhľadávanie charakteristík spamu.

3.6.6.1 Zabránenie šírenia linkového spamu

Čiastkovým riešením je zabránenie šírenia spamu prostredníctvom indexovacích služieb. Google zaviedol za týmto účelom rozšírenie tágu na vytváranie liniek „A“ [9].

```
<a href="http://www.externaadresa.com" rel="nofollow">...</a>
```

Na podobnom princípe je založené publikovanie informácií o stránkach, ktoré sú pre prehliadače relevantné, alebo nevhodné na indexovanie. Ide o súbory robots.txt. Tie sú súčasťou odporúčania Robots Exclusion Standard [10]. Tieto súbory sú implicitne sťahované keď prehliadač vytvára index stránok a obsahujú informácie o tom, ktoré súbory sú vhodné na katalogizáciu a ktoré nie. Nasledujúci príklad ukazuje obsah súboru robots.txt

```
User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp
Disallow: /~joe/
```

Okrem smerovania liniek a explicitného zoznamu stránok je možné zakázať indexovanie stránok pomocou značenia v tágoch „meta“. Nasledujúca ukážka označí stránku tak, aby ju vyhľadávač neindexoval do databázy a zároveň aby neindexoval linky, ktoré sú na tejto stránke

obsiahnuté.

```
<meta name="robots" content="noindex,nofollow" />
```

Uvedené spôsoby sú len odporúčania, ktorých sa držia tvorcovia indexovacích strojov na internete. Zároveň nie je žiadúce blokovať všetky prepojenia. Preto je potrebné zohľadniť využitie týchto blokovacích mechanizmov. To vraví prehliadačom, aby nenasledovali odkaz na danú stránku. Toto riešenie však nebráni vkladaniu spamu, ale bráni len jeho rozširovaniu. Prevencia pred vkladaním vyžaduje jeho rozpoznanie.

Horeuvedené riešenie s linkami vo wiki je preventívne, a nedokáže dostatočne dynamicky riešiť problém s linkovým spamom.

3.6.6.2 Metódy na rozpoznanie spamu aplikovateľné na stránky vo wiki

Riešenia zahŕňajú dva prístupy. Jedným je zablokovanie zakázaného obsahu, ďalšie je filtrovanie zakázaných častí obsahu. Blokovanie je realizované na základe viacerých prvkov. Každý prvok je možné identifikovať niektorou technikou.

Statické filtrovanie

Táto technika vyhľadáva v obsahu špecifické výrazy (regulárne výrazy). Pri výskyte niektorého z výrazov v databáze je obsah označený ako nevhodný. Okrem samotného obsahu sa vyhodnocujú aj informácie, ktoré nie sú priamou časťou obsahu. V prípade emailovej komunikácie sa vyhodnocujú hlavičky správ a v prípade internetovej adresy doména a IP adresa. V prípade útoku sú tieto informácie často krát pozmenené tak aby útočník ukryl svoju identitu.

Nevýhodou tohto prístupu je vysoká náročnosť na údržbu zoznamu blokovaných výrazov a charakteristík. V prípade použitia tejto techniky je aj vysoká pravdepodobnosť nesprávneho vyhodnotenia obsahu. Príkladom je stránka o sexuálnej výchove. Aj keď obsahuje slovo „Sex“ v danom kontexte táto informácia nie je zakázaná.

Útočníci tento typ filtra obchádzajú rôznymi zmenami vo výrazoch a dopĺňaním informácií, ktoré napokon nie sú zobrazené, napríklad komentáre:

```
Viagra -> V1agra
```

```
Viagra -> Via<!---->gra
```

Riešením je analýza tzv. bielych znakov. V internetovom kontexte ide o symboly, ktoré sa v konečnom dôsledku nezobrazia používateľovi po vyrenderovaní internetovým prehliadačom. Vo wiki je potrebné zohľadniť špecifiká použitého značkovacieho jazyka, alebo jazyk obmedziť o prvky, ktoré umožňujú tento typ útoku.

Analýzu výrazov je potrebné rozšíriť aj o analýzu zastupiteľných znakov. Ide o znaky, ktoré sú po vyrenderovaní browserom ekvivalentné s pôvodným zápisom a znaky, ktoré opticky pripomínajú pôvodné znaky.

i -> 1 a -> @ o -> 0 Z -> 2	A -> A b -> b

Tabuľka 1: Príklady zastupiteľných a ekvivalentných znakov

Pri analýze je potrebné zohľadniť aj znakové kódovanie textu, keďže národné znaky sú rôzne reprezentované v jednotlivých znakových sadách.

Statické filtrovanie je v pôvodnej verzií zvlášť vhodné na filtrovanie adries a URL identifikátorov. Akákoľvek zmena v názve zmení referencovaný zdroj a to je nežiadúce. Priamo v obsahu stránky sa už tento typ spamu vyskytuje v menšej miere.

Štatistické filtrovanie

Tento typ filtrovania je založený na tom, že niektoré výrazy sa nachádzajú častejšie v určitej skupine dokumentov. Analýzou vyskytujúcich sa výrazov je možné zaradiť s určitou pravdepodobnosťou k určitej skupine dokumentov, napríklad spamu.

Štatistické filtrovanie je schopné adaptovať sa na nové typy správ a výrazov. Priama administrácia nie je vyžadovaná. Ak je zapnuté učiaci režim je potrebné označiť príslušnosť správ k danej skupine a pomocou algoritmu sa získajú a spracujú charakteristiky, ktoré umožnia presnejšie rozpoznať skupinu dokumentu.

Výhodou štatistického filtrovania oproti statickému je, že vyhodnocujú vyhľadané výrazy s dôrazom na kontext. Napríklad ak je cieľom zablokovať nevhodný obsah o drogách a v obsahu sa nachádza slovo identifikujúci drogu predpokladá sa, že ide o blokovaný obsah. V tomto prípade sa obsahu priradí hodnota asociovaná s výskytom charakteristických výrazov. Ale ak obsah obsahuje informácie o prevencií pred šírením drog a s tým asociované výrazy, hodnotenie klesá. Pod určitou kritickou adresou je tento dokument označený ako neutrálny.

Útočníci sa pokúšajú obísť tento filter tak, že do obsahu vkladajú veľké množstvo výrazov, ktoré znižujú príslušnosť k skupine reprezentujúcej blokovaný obsah a robia obsah neutrálny.

4 Návrh algoritmov a postupov na rozpoznanie útoku na wiki a jeho zamedzenie

Algoritmy a postupy uvedené v tejto časti vychádzajú z myšlienok uvedených v [13] a [14]. Nasledujúce postupy a algoritmy je možné rozdeliť do nasledujúcich kategórií:

- *Taktiky na zamedzenie vkladania nevhodného obsahu.*
- *Blokovanie pôvodcov spamu.*
- *Analýza a vyhľadanie nevhodného obsahu.*
- *Blokovanie a odstraňovanie nevhodného obsahu.*
- *Manuálne hlásenie spamu a výskytu nevhodného obsahu.*

4.1 Taktiky na zamedzenie vkladania nevhodného obsahu

Prvá skupina zahŕňa ochranné prvky prítomné priamo na stránke. Často krát ide o prvky, ktoré zabraňujú automatizovaným útokom. Vhodné je použiť projekt CAPTCHA. Implementácia je realizovaná vložením obrázku, obsah ktorého človek musí odpísať do pripraveného formulára. Na strane servera sa následne overí, či sa tieto texty zhodujú. Takýto prístup pridáva ďalší ovládací prvok, ktorý môže znamenať menší komfort pre používateľa. Jedným z návrhov je spojenie ovládacích prvkov (odoslanie formulára obsahujúceho aktualizovanú verziu) spolu s CAPTCHA prvkom. Príklad je zobrazený na obrázku 7.

Obrázok 7: Príklad kombinácie prvku CAPTCHA s ovládacím prvkom formulára (http://www.myseattle.com/image/save_page.png)

4.2 Analýza a rozpoznanie nevhodného obsahu

4.2.1 Metódy na analýzu prítomnosti explicitného spamu

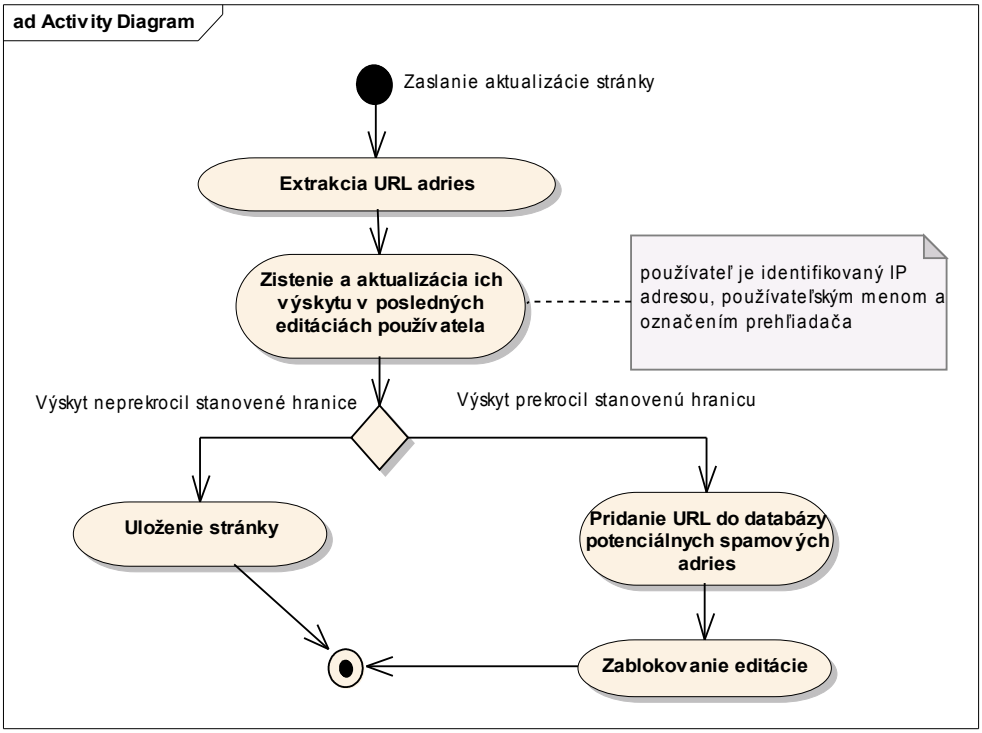
Tieto metódy vychádzajú z riešení, ktoré sa používajú na ochranu pred spamom v iných oblastiach ako wiki. Dobre vyvinuté systémy na ochranu spamu sa používajú na filtrovanie elektronickej pošty. Tieto systémy rozpoznávajú explicitný spam. Explicitný spam priamo obsahuje výrazy a stránky, ktoré pôsobia rušivo a nie sú prepojené s pôvodnou informáciou.

Na analýzu je možné využiť existujúce knižnice na analýzu a rozpoznanie spamu. Výstupom analýzy spamu je získanie viacerých metrík a súhrnnej hodnoty. Tieto hodnoty je následne možné využiť na určenie či má byť publikácia zablokovaná.

Pre wiki systémy je možné využiť externú a internú databázu URL adries indikujúcich spam. Jednou z externých databáz je [15]. Táto databáza je jednoducho dostupná prostredníctvom webu. Určenie, či daná linka je spam je realizované DNS požiadavkou na špecifickú doménu. V prípade správneho určenia je URL adresa v zozname blokovaných adries.

4.2.2 Analýza správania používateľov

Vo wiki je možné definovať vzory nevhodného správania v čase, ktoré je potom možné na strane server blokovať. Ide o frekvenciu aktualizácií stránok, frekvenciu výskytu kľúčových slov a frekvenciu výskytu opakujúcich sa a neopakujúcich sa URL adries. Po prekročení stanovených limitov je editácia zablokovaná. Postup pri implementácií je znázornený na diagramoch 8 a 9.



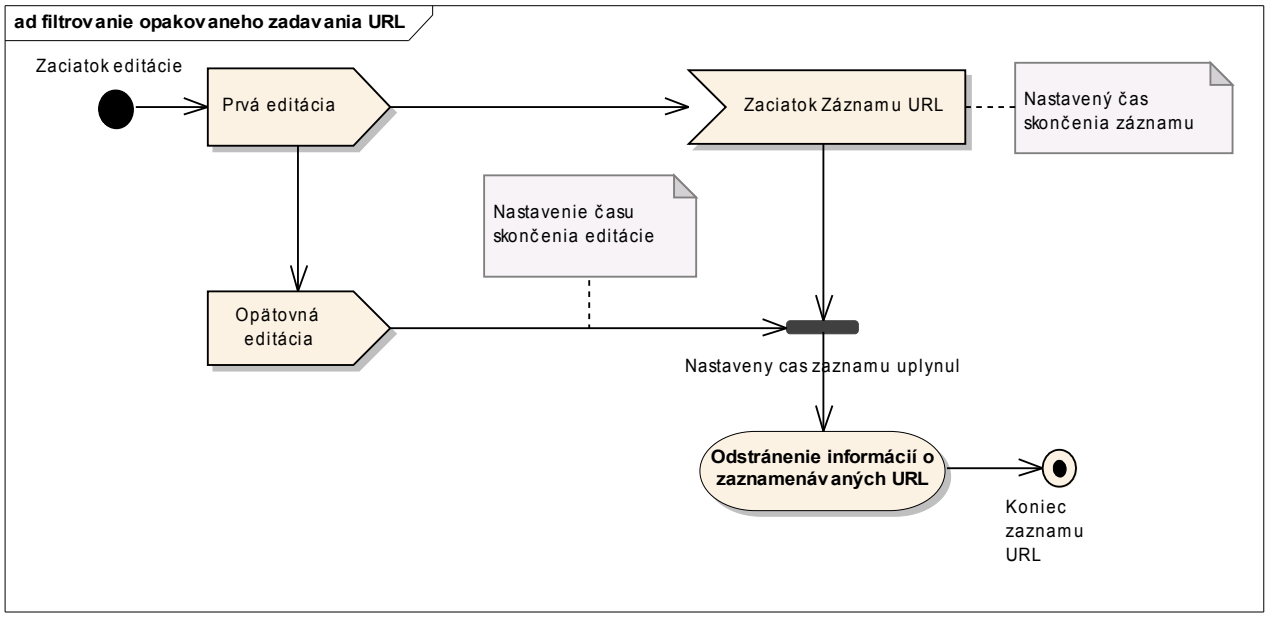
Obrázok 8: Vyhodnotenie opakovaných vkladaní URL adries

Môj návrh definuje časový interval počas ktorého je záznam realizovaný. Dôležitým faktorom je aj informácia, či ide o linky na interné stránky v danom wiki, alebo externé adresy. Mojím ďalším rozšírením tohto postupu je určenie samostatnej hranice výskytu opakovaných liniek pre interné stránky a zvlášť pre externé adresy. Problém môže nastať, ak sú odkazy na jednu centrálnu stránku žiadúce. Toto je možné vyriešiť definovaním výnimiek v adresách. Príkladom takéhoto centrálneho linkovania môže byť prepojenie so stránkou obsahujúcou licenciu, ktorá sa vzťahuje na obsah stránky. Právo definovať výnimky v hodnotení jednotlivých adries má administrátor daného wiki.

Analýza správania môže hodnotiť nie len vkladanie liniek, ale aj zvýšené hodnotenie článkov pomocou konvenčného spamového filtra. Princíp je podobný ako pri blokovaní URL. Hodnotenie prebieha v určitom časovom rámci. Po uplynutí tohto rámca sú informácie o správaní vymazané. Pri hodnotení sa zohľadňuje interval vrátený spamovým filtrom. Interval je v tomto prípade rozdelený na 3 časti. Najväčšia pravdepodobnosť spôsobí okamžité zablokovanie a toto je zaznamenané pre daného používateľa. Blokovanie však ešte neprebieha. V prípade zvýšenej príslušnosti ku spamu je táto pravdepodobnosť len zaznamenaná. V prípade nízkej príslušnosti ku spamu nie je zaznamenávaná žiadna informácia o používateľovej akcií.

Zablokovanie používateľa nastáva ak počas sledovaného obdobia prekročí počet takýchto

záznamov určitý počet a súhrnné hodnotenie príslušnosti ku spamu.



Obrázok 9: Začiatok a koniec záznamu URL

4.3 Autorizácia / zamedzenie prístupu

Z podstaty wiki vyplýva, že pre evolúciu informácie je potrebné čo najvoľnejšie prostredie. To znamená, že informácia by mala byť čo najľahšie editovateľná čo najširším okruhom používateľov.

Tento prístup však v súčasnom prostredí nie je vhodný. Útok na wiki je často krát vedený používateľom, ktorý chce ostať v anonymite. Na základe tohto faktu je možné vytvoriť viacero bezpečnostných politík pre jednotlivé operácie vo wiki.

Jedným typom politiky je politika, ktorá zamedzí prístup k editačným operáciám vo wiki neregistrovaným používateľom. Toto riešenie je implementačné najjednoduchšie. Všetky položky v histórií stránky sú prepojené s daným používateľom. V prípade, že registrovaný používateľ začne vandalizovať stránky, je možné jednoducho vystopovať operácie, ktoré vykonal a opraviť vzniknuté škody. Štandardná reakcia smerovaná na používateľa je varovanie, obmedzenie práv pre vykonávanie konkrétnych operácií, poprípade celkové zamedzenie prístupu k editačným operáciám vo wiki.

Problémom tejto politiky môžu byť falošné registrácie. Pre obmedzenie počtu falošných registrácií je potrebné technicky obmedziť registráciu. Príkladom môže byť dočasné označenie používateľovho prehliadača. Takto označený prehliadač už nemá povolený prístup k opätovnej

registrácií. Okrem označovania klienta je možné zaznamenať aj iné údaje – IP adresu, browser, operačný systém – a na základe týchto údajov môže byť zakázaná opätovná registrácia.

Iným riešením je vytvorenie systému odporúčaní. Priama registrácia nie je povolená, registrácia používateľov je uskutočniteľná až po prijatí pozvania od už registrovaného používateľa. Týmto sa čiastočne rieši problém s prístupom vandalov k editačným operáciám. Toto riešenie zabraňuje prístupu nových používateľov, keďže na získanie registrácie je potrebné poznať už registrovaného používateľa. Otvorenosť wiki je v prípade tohto riešenia do značnej miery obmedzená. Výhodou je ale homogénnosť komunity, ktorá sa vytvára okolo wiki, keďže sa predpokladá, že používateľ, ktorý dostane pozvanie sa zaoberá a zaujíma podobnými témami ako používateľ, ktorý mu udelil pozvanie.

Doteraz uvedené spôsoby registrácie používateľov vyžadovali explicitné získanie prístupu od iných používateľov, alebo registráciu a získanie určitého stupňa dôveryhodnosti, ktorá by sa odrazila na editačných právach. Nasledujúci návrh funguje opačným spôsobom. Táto bezpečnostná politika je založená na predpoklade, že žiadny používateľ pri prvom prístupe k editačným operáciám nie je vandal. Systém sleduje používateľovo správanie sa v systéme. V prípade, že je odhalený vandalizmus spôsobený týmto používateľom, je mu zamedzený prístup k editačným operáciám.

Zistiť, či používateľ je pôvodcom vandalizmu vo wiki, je iniciované inými mechanizmami uvádzanými v predchádzajúcej sekcií. Zamedzenie je realizované blokovaním konkrétnych IP adries, poprípade rozsahov a sieti. To či používateľ mal zamedzený prístup je možné určiť aj podľa označenia browsera (prostredníctvom cookie), ktoré získa pri prvom zablokovaní.

4.4 Stránky so zvýšenou editačnou aktivitou

Stránky, ktoré sa často menia zneprehľadňujú históriu dokumentu. Časté zmeny indikujú nestabilitu informácie. Keď je stránka v takomto stave, nie je vhodné zobrazovať aktuálne verzie, ktoré nemusia obsahovať vhodné informácie.

Jedným z možných návrhov riešenia je uvedený na [14]. Stránky, na ktorých sa uskutočnil určitý počet zmien v krátkom čase získajú status „často editované“. Stránky s týmto statusom majú definované odlišné správanie sa.

Informácie v histórií sa neukazujú priamo ale celá editačná aktivita je zahrnutá v jednej položke histórie. Následne je možné zobraziť detaily tejto editačnej aktivity. Dané riešenie

predpokladá, že kým má stránka status „často editovaná“ neuchovávajú sa verzie. Kým má stránka tento status nie je zobrazovaná aktuálna verzia, ale je zobrazená stabilná verzia. V tomto prípade je stabilná verzia tá, ktorá je uložená ako posledná pred označením stránky ako často editovanej.

Môj návrh predpokladá uloženie revízií, pokiaľ má stránka tento status. Avšak po zrušení tohto statusu sú tieto zmeny spojené do jednej revízie.

4.4.1 Identifikácia a opatrenia proti „edit wars“

Tento jav indikuje, že používatelia wiki si navzájom rušia svoje príspevky. Toto je nezdravý jav, keďže obaja používatelia sú presvedčený o tom, že práve ich príspevok je ten správny a nechcú uznať jeho nepresnosť, alebo nesúhlas iných používateľov. Preto sa snažia stále vracať svoje zrušené aktualizácie. Počas tohto procesu získava stránka status „často editovaná“

Jedným z riešení je dočasné zablokovanie editácie stránky. Počas zakázanej editácie používatelia výmenu názorov presmerujú na diskusné fórum, kde sa pomocou ostatnej časti komunity hľadá kompromisné riešenie.

Určenie, či je potrebné zablokovať operácie je možné definovaním pravidiel, ktoré určujú aké typy zmien boli uskutočnené počas obdobia, keď stránka nadobudla status „často editovaná“. Môj návrh predpokladá zvýšený výskyt návratov k posledným verziám v histórií stránky. V prípade, že počet zmien obsahuje určený pomer medzi novými verziami a návratmi ku starým je stránka zablokovaná s odôvodnením, že ide o „edit war“. Informácia, že vznikol takýto stav je zobrazená na príslušnej stránke.

4.5 Získavanie informácií o relevantnosti informácií z pohľadu používateľa

Wiki vyžaduje vďaka svojej otvorenosti intenzívnu kontrolu od používateľov. Pasívny prvok, ktorý zaznamenáva názor používateľa na stránku je preto veľmi žiadaný.

Pre používateľa je najjednoduchšie rozpoznať spam. V prípade zisteného spamu používateľ môže označiť stránku ako spam. Pre túto potrebu je potrebné rozšíriť ovládanie funkčnej časti wiki o tlačidlo, ktoré označí stránku ako spam. V tomto prípade je revízia označená ako spam a je administrátor tento fakt je zaznamenaný.

Pre presnejšiu analýzu je potrebné získavať od používateľa viac informácií. Viacero prvkov, ale už pôsobí rušivo a nesplnilo by svoj účel. Vhodným miestom na vyjadrenie sa ku kvalite sú

doplnkové média – diskusné fórum, mailing list. Pre potreby automatizovaného spracovania je potrebné rozšíriť funkcionalitu wiki systému o analýzu a získavanie informácií o relevantnosti z týchto zdrojov. Jednoduchým rozšírením je pridanie typu príspevku do webovej diskusie. Je potrebné rozoznávať typy príspevkov na príspevky týkajúce sa relevantnosti stránky a ostatné. Zadávanie príspevkov týkajúcich sa relevantnosti by malo byť rozšírené o prvky na ohodnotenie relevantnosti.

4.6 Možné spôsoby na ochranu wiki

Wiki je orientované na obsah stránok. Útoky sú často krát vedené práve na stránky. Problematické sú zvlášť zmeny, ktoré realizujú zmenu obsahu na iný, ktorý vyzerá veľmi podobne ako relevantný obsah. Preto je potrebné hľadať riešenia orientované na obsah a nie na vyhľadávanie konkrétnych prvkov indikujúcich spam.

V prípade, že sa kontext novej stránky výrazne odlišuje od pôvodnej/stabilnej verzie, je zvýšená pravdepodobnosť, že nová informácia je nerelevantná. Porovnanie je možné uskutočniť viacerými spôsobmi.

4.6.1.1 Porovnanie automatizovaného kategorizovania novej verzie informácie s pôvodnou

Jedným je kategorizácia informácie. Predpokladom je, že máme algoritmus, ktorý vie dokument na základe jeho obsahu zaradiť s určitou pravdepodobnosťou do viacerých kategórií. Následne nová informácia by mala mať s určitou presnosťou rovnaké zaradenie ako pôvodná informácia.

Na kategorizáciu dokumentov je možné použiť viacero spôsobov. Pritom je možné použiť štatistickú analýzu výskytu slov, vyhľadávať sémantické prepojenia, alebo použiť hybridný prístup. Jedným z príkladov algoritmu na kategorizáciu je [17].

Tieto algoritmy avšak pracujú nad skupinou dokumentov, ktoré rozdelia, preto je problematické použiť ich na zaradenie jedného dokumentu do už existujúcej štruktúry. Preto je vhodnejšie používať iné metódy katalogizácie. Jedným z možných smerovaní tohto návrhu ochrany sú sémantické wiki systémy. V súčasnosti neexistujú dostatočne vyvinuté implementácie sémantického wiki. V týchto dokumentoch je možné využiť metainformácie na katalogizáciu dokumentu.

4.6.1.2 Využitie kontextových odkazov

Táto metóda využíva na analýzy kontextové odkazy vo stránke. Kontextový odkaz je v tomto prípade uvažovaný nie ako hyperlink, ale ako kľúčové slovo, ktoré identifikuje skupinu dokumentov, ktoré sú kontextovo prepojené s informáciou. Môj návrh predpokladá, že nová a pôvodná informácia sú kontextovo prepojené s približne rovnakou skupinou informácií.

Prvou časťou analýzy informácie je vydolovanie kľúčových slov z pôvodnej a novej verzie informácie. Následne je potrebné vyhľadať prepojenie s inými stránkami. Prirodzený predpoklad je použitie databázy stránok a ich kľúčových slov. Pre niektoré typy informácií neexistujú databázy obsahujúce dokumenty asociované s kľúčovými slovami. Zároveň tieto databázy trpia neaktuálnosťou oproti zmenám vo wiki. Predpokladom je využitie častejšie aktualizovaných databáz. V mojom návrhu predpokladám využitie už existujúcich indexovacích služieb a vyhľadávačov.

Kľúčové slová získané z jednotlivých dokumentov sú vyhľadávané pomocou vyhľadávača. Výstupom je množina odkazov prepojených s daným kľúčovým slovom. Kľúčové slová sami o sebe nehovoria nič o kontexte, v ktorom sa vyskytujú. Vhodnejší prístup je zadávanie podmnožín kľúčových slov do vyhľadávača. Po získaní výsledkov z pôvodnej informácie a novej informácie sú výsledné množiny kontextových odkazov porovnávané. Výstupom tohto algoritmu je zhodnosť dokumentov.

Ak zoberieme dve množiny odkazov, tak výstup je pomer počtu spoločných odkazov k celkovému počtu v odkazov v mohutnejšej množine. Problém môže nastať, ak napríklad v nasledujúcom príklade

```
auto, spotreba, rýchlosť           -> www.skoda.cz/oktavia.html
spotreba, spoľahlivosť, auto, cena -> www.skoda.cz/fabia.html
```

dve skupiny slov určujú podobný kontext, ale vyhľadávač vrátil rôzne adresy. Toto je možné ošetriť tým, že vrátené adresy budú vyhodnocované ako zhodné ak sa zhodujú v určitej časti adresy URL.

5 Opis implementovaného systému

Cieľom je implementovať jednoduchý wiki systém. Pri návrhu sa kladie dôraz na jednoduchosť používania a na možnosť implementácie ochranných mechanizmov pred zadávaním nevhodného vstupu. Výstupom je aj implementácia referenčného mechanizmu a presná definícia rozhrania.

Pri vývoji je použitá aplikačný rámec Spring, preto vzory použité v aplikácií sú volené tak, aby bolo možné v maximálnej miere využiť vlastnosti tohto rámca a vyhnúť sa jeho nedostatkom. Zároveň som sa snažil rozdeliť systém do vrstiev podľa paradigmy MVC.

5.1 Špecifikácia systému

Implementované wiki musí umožňovať

- Zobrazovanie stránok s dôrazom na jednotné zobrazenie.
- Sprístupnenie editačnej funkcionality – možnosť vytvárať nové stránky a editovať už existujúce. Táto časť je najdôležitejšia časť wiki systému.
- Možnosť prezerania revízií stránky.
- Možnosť návratu k pôvodnej revízií stránky.
- Vkladanie a sprístupnenie príloh – súborov.
- Návrh musí poskytnúť rozhranie na dodatočnú implementáciu verifikačných mechanizmov, ktoré sú aplikované na údaje spracovávané počas vytvárania novej stránky a editácie existujúcej.

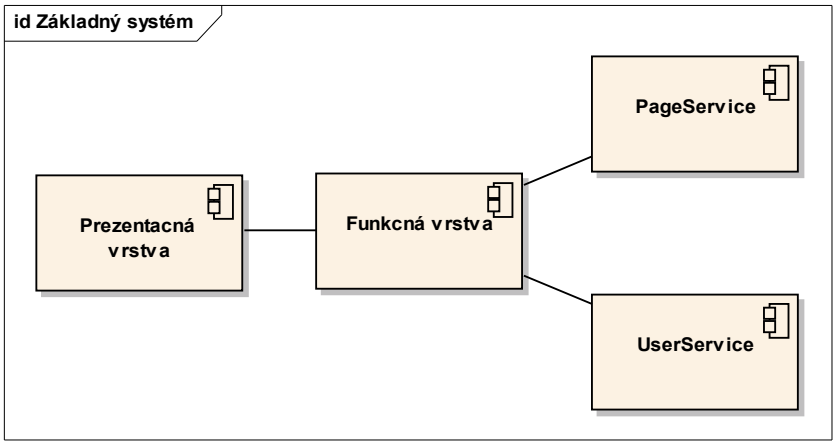
5.2 Použité technológie

Wiki som sa rozhodol implementovať v jazyku Java. Java bola vybraná pre vysokú dostupnosť programových komponentov. Voľba bola podmienená aj mojimi skúsenosťami s použitou technológiou. Ako webový kontajner bol použitý server Apache Tomcat.

Úložisko dát je implementovaná nad databázou PostgreSQL v kombinácií so súborovým systémom. Prístup k databáze je realizovaný nie priamo, ale prostredníctvom aplikačného rámca Hibernate. Objektové mapovanie databázy umožňuje dynamickú zmenu reprezentácie dát a rozširovanie dátového modelu.

5.3 Architektúra základného systému

Základná architektúra pripomína všeobecnú architektúru wiki.



Obrázok 10: Prepojenie komponentov základného systému.

Prezentačná vrstva zodpovedá zobrazenie vstupných údajov a prvotné spracovanie a validáciu vstupných údajov. Prezentačná vrstva zabezpečuje konzistentnosť vzhľadu. Vytvára používateľské rozhranie na prácu so systémom.

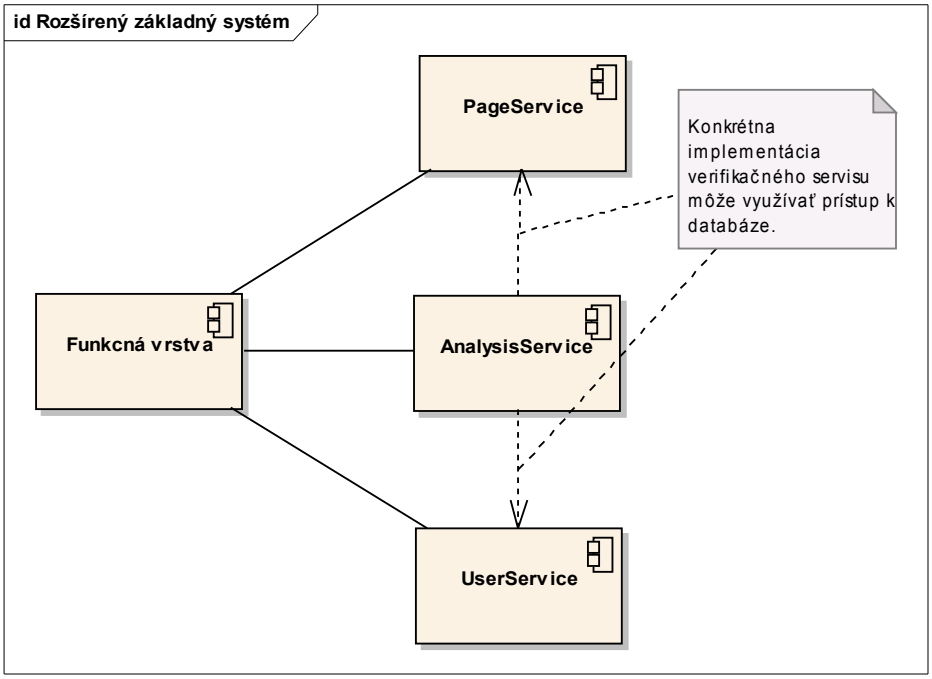
Funkčná vrstva zabezpečuje spracovanie vstupných údajov a iniciovanie príslušnej akcie v systéme. Akcia je delegovaná na príslušnú metódu v PageService, alebo UserService.

Servisné komponenty PageService a UserService sú zodpovedné za realizáciu príslušnej funkcie nad databázou.

5.4 Rozšírenie funkčnej vrstvy

Overovanie a validácia prebieha na funkčnej vrstve. Verifikácia je delegovaná externým servisom zodpovedným za validáciu. Obrázok 11 znázorňuje prepojenie komponentov s verifikačným servisom.

AnalysisService je zodpovedný za verifikáciu informácií o nových revíziách a verziách stránok. Ako vstup príma údaje, ktoré má verifikovať a údaje o zdroji zmien. Vstupné rozhranie je definované flexibilne, aby bolo možné údaje predávané verifikačnému servisu pridávať a meniť podľa potreby.



Obrázok 11: Prepojenie funkčnej vrstvy s AnalysisService

V systéme je definovaný generický AnalysisService. Ten slúži na delegovanie verifikácie ďalším servisom. Takýmto spôsobom je možné zreťaziť verifikáciu a uplatniť na vstupné dáta viac verifikačných mechanizmov.

6 Zhodnotenie a záver

Analyzoval som otvorené webové kolaboratívne prostredia wiki, pričom som sa zameral na návrh riešenia problémov spojených s otvorenosťou systému. Prebral som viacero typov útokov na wiki a ich možné riešenie.

V práci som som sa zameral na analýzu možných spôsobov zabezpečenia wiki a postupov na rozpoznanie útoku na wiki. Postupy uvádzané v tejto práci je možné využiť aj v iných webových kolaboratívnych systémoch, ako sú blogy a internetové fóra. Navrhnuté riešenia analyzujú správanie používateľa a vytvára tak jeho profil. Tento môže byť neskôr použitý na dodatočné hodnotenie kvality článku podľa profilov používateľa.

Výstupom projektu je veľmi jednoduchá wiki aplikácia umožňujúca pridávanie a editovanie stránok. V rámci aplikácie sú využité technológie CAPTCHA a spam filter prebraný z mailových programov.

Súčasťou návrhu sú aj návrhy porovnávania stránok pomocou výslednej katalogizácie a kľučových slov prostredníctvom konvenčných a nekonvenčných databáz. Cieľom ďalšej práce môže byť vytvorenie jednoduchých implementácií a overenie funkčnosti mnou navrhovaného riešenia.

Postupy na analýzu správania používateľa sú veľmi podobné postupom používaným v prostredí aplikácie „Business Inteligence“ a „Data Mining“, kde sú podobné postupy použité na predpovedanie správania sa klienta a určovanie rizikovosti. Zámerom ďalšej práce môže byť podrobnejšie preskúmanie týchto metód a ich aplikácia na ochranu otvorených kolaboratívnych prostredí.

Wiki sa postupne transformuje na platformu vytvárajúcu sémantické prepojenia informácií. Preto navrhujem, aby ďalšia práca analyzovala ochranu dát, ako aj hodnotenie relevantnosti v semantických wiki.

Literatúra

[1] Metacollab: General theory of collaboration. [ONLINE] Dostupné na http://collaboration.wikia.com/wiki/General_theory_of_collaboration

[2] Wikipedia: Collaborative software. [ONLINE] Dostupné na http://en.wikipedia.org/wiki/Collaborative_software

[3] C2: Why Wiki Works Not. [ONLINE] Dostupné na http://c2.com/cgi/wiki/wiki?WhyWikiWorksNot

[4] The Google Pagerank Algorithm and How It Works. [ONLINE] Dostupné na http://www.iprcom.com/papers/pagerank/

[5] Leuf, Bo. The Wiki Way: quick collaboration on the Web / Bo Leuf, Ward Cunningham, Addison-Wesley, 5. vydanie, 2005, kap. 1, s. 14

[6] C2: Wiki Design Principles. [ONLINE] Dostupné na http://c2.com/cgi/wiki?WikiDesignPrinciples

[7] Burrow, A. L. : Negotiating access within Wiki: a system to construct and maintain a taxonomy of access rules, Proceedings of the fifteenth ACM conference on Hypertext and hypermedia, Santa Cruz, CA, USA, 2004, s. 77. – 86. [ONLINE] Dostupné na http://portal.acm.org/citation.cfm?doid=1012807.1012831

[8] Wikipedia: Wiki Vandalism. [ONLINE] Dostupné na http://en.wikipedia.org/wiki/Wiki_vandalism

[9] Google Blog: Preventing comment spam. [ONLINE] Dostupné na http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html

[10] Wikipedia: Robots Exclusion Standard. [ONLINE] Dostupné na http://en.wikipedia.org/wiki/Robots.txt

[11] CAPTCHA. [ONLINE] Dostupné na http://www.captcha.net/

[12] Wikipedia: Stopping e-mail abuse. [ONLINE] Dostupné na http://en.wikipedia.org/wiki/Stopping_e-mail_abuse

[13] Metawiki: Anti-vandalism ideas. [ONLINE] Dostupné na http://meta.wikimedia.org/wiki/Anti-vandalism_ideas

[14] Metawiki: Spam Filter. [ONLINE] Dostupné na

http://meta.wikimedia.org/wiki/Spam_Filter

[15] SURBL: Spam URI Realtime Blocklists. [ONLINE] Dostupné na http://www.surbl.org/

[16] Metawiki: Automated edit war squashing. [ONLINE] Dostupné na http://meta.wikimedia.org/wiki/Automatic_edit_war_squashing

[17] Repiský, V. - Homola, M.: A Similarity-based approach in term weighting for text categorization, Journal of Electorical Engineering, vol. 56, no. 12/s, 2005, s. 94-97

PRÍLOHA A Obsah elektronického média

Adresár	*Popis*
/bin-dist	Obsahuje binárnu distribúciu systému
/bin-dist/wiki	Tento adresár obsahuje aplikáciu. Počas inštalácie je potrebné prekopírovať tento adresár do adresára webapps programu Tomcat.
/ddl	Obsahuje schému databázy
/programy	Obsahuje programy potrebné na spustenie systému.
/src-dist	Obsahuje zdrojové kódy programu

PRÍLOHA B Technická dokumentácia

Projekt je vypracovaný v jazyku Java.

Popis použitej notácie

Systém je implementovaný ako webová aplikácia s využitím aplikačného rámca Spring. Spring definuje komponenty s ohraničenou funkcionalitou a vytvára medzi nimi prepojenia podľa funkcionality, ktorú vyžadujú. Pri tvorbe technickej dokumentácie preto nebudem uvádzať úplné diagramy tried a v niektorých prípadoch uvediem prepojenia komponentov tak ako sú realizované v aplikačnom rámci Spring.

Pre jednoduchosť sa vyhnem presnému opisu rozhraní a budem uvádzať len metódy, ktoré sú potrebné na objasnenie práve opisovanej časti. Zároveň sa budem sústrediť na opis správania funkcie v danom kontexte.

Pri opise jednotlivých komponentov je využitý diagram prepojení s ostatnými komponentami v rámci Spring.

Na tvorbu UML diagramov bol použitý program Sparx Enterprise Architect.

Konvencie používane v systéme

Názov stránky – URL

Názov stránky vo wiki je presne asociovaný s URL adresou, konkrétne s poslednou časťou. Stránky s daným titulom získame výmenou všetkých medzier v názve znakom „_“. Pri spätnej konverzií (z URL adresy na meno) sa ešte nastavuje prvé písmeno na veľké.

```
/stranka.html                 ->   Stranka
/Nejaka_ina_stranka.html      ->   Nejaka ina stranka
/%C4%BE%C5%A1%C4%8D.html      ->   Ľšč
```

Systém predpokladá kódovanie URL v UTF-8. V prípade špeciálnych znakov (znaky s diakritikou) sú akceptované zakódované znaky pomocou notácie „%XX“ kde XX je kód znaku (http://en.wikipedia.org/wiki/URL_Encoding). V prípade viac bajtových znakov je týchto sekvencií uvádzaných viac. Stránky s diakritickými znakmi v názve nie sú transliterované pri

prepise do URL adresy.

Revízia stránky

Revízia je identifikovaná dátumom zmeny. Vonkajšie rozhranie dostáva túto hodnotu vo formáte

```
yyMMddhhmmss
```

V tomto formáte je „yy“ rok, „MM“ mesiac, „dd“ deň v mesiaci, „hh“ hodina, „mm“ minúta a „ss“ sekunda. Tento dátum reprezentuje začiatok editácie stránky.

Opis obsahu jednotlivých balíkov

sk.kmit.wiki.analysis

Tento balík obsahuje rozhranie a implementáciu WikiAnalysisService. V súčasnosti sú implementované dva servisy.

- WikiAnalysisServiceImpl – implementácia zreťazenia servisov. Verifikuje vstupné dáta pomocou servisov, ktoré sú mu predané počas konfigurácie.
- JasenFacadeService – servis, ktorý využíva Jasen Spam Filter na overenie, či je správa spam. V konštruktore je nastavovaná hraničná hodnota.
- WikiAnalysisResult – Objekt, ktorý reprezentuje výsledok verifikácie.

sk.kmit.wiki.beans

Tento balík obsahuje objekty, ktorými sú predávané údaje z formulárov. Súčasťou sú validátory. Každý beán je využívaný v niektorom z kontrolérov.

sk.kmit.wiki.controller

V tomto balíku sú obsiahnuté kontroléry, ktoré zodpovedajú za spracovanie požiadaviek. V systéme sú definované 4:

- *PageViewController* – riadi zobrazenie obsahu stránky. Vstupné údaje sú obsiahnuté v objekte *sk.kmit.wiki.beans.PageViewCommand.*

- *PageCreateController* – vytvorenie stránky. Vstupné údaje sú obsiahnuté v objekte *sk.kmit.wiki.beans.PageEditForm.*
- *PageEditController* – editácia stránky. Vstupné údaje sú obsiahnuté v objekte *sk.kmit.wiki.beans.PageEditForm.*
- *PageHistoryController* – zobrazenie histórie stránky. Vstupné údaje sú obsiahnute v objekte *sk.kmit.wiki.beans.PagerCommand.*

sk.kmit.wiki.dao

V tomto balíku sú obsiahnuté triedy na manipuláciu s databázou.

sk.kmit.wiki.domain

V tomto balíku sú dátové triedy, prostredníctvom ktorých sa pracuje s databázou. Na prácu je použitý aplikačný rámec Hibernate.

sk.kmit.wiki.exception

Tento balík obsahuje výnimky, ktoré môžu nastať v systéme. Definuje dva typy výnimiek:

- Runtime – nie sú explicitne odchytávané: *NotImplemented, WikiRuntimeException, WikiInternalException*
- *WikiException, WikiDbException, WikiPageAlreadyExists, WikiPageNotFound,*

sk.kmit.wiki.service

Obsahuje servisné triedy pre realizáciu akcií nad databázou. V servisných triedach je manažované transakčné spracovanie.

sk.kmit.wiki.util

Obsahuje podporné triedy pre prácu vo wiki.

sk.kmit.wiki.view

Obsahuje dodatočné triedy pre prezentačnú vrstvu.

Opis funkčných komponentov

Prezentačná vrstva

Prezentačná vrstva je riešená pomocou jazyka Velocity (http://jakarta.apache.org/velocity). Tento jazyk bol zvolený pre svoju jednoduchosť a vysoký výkon. Ďalším faktorom bola možnosť definovať šablóny, ktoré je možné vkladať do seba a tým je zabezpečená vzhľadová konzistentnosť.

Prepojenie s funkčnými komponentami je realizované prostredníctvom aplikačného rámca Spring Web MVC. V tomto rámci je zobrazenie stránky iniciované predaním údajov ktoré sa majú vykresliť spolu so šablónou v jazyku Velocity, ktorá ich má vykresliť. Toto umožnilo jasne ohraničiť prezentačnú vrstvu od funkčnej.

Základnou časťou mojej implementácie je súbor *layout.vm*. Ten obsahuje základnú šablónu stránky do ktorej sú vkladané ďalšie šablóny zodpovedné za vyrenderovanie potrebného výstupu. Hlavná šablóna vyžaduje nasledujúce predané predané z kontroléra:

action	Obsahuje textový popis akcie, ktorá je asociovaná s daným formulárom. Možné hodnoty sú: • „edit“ - editácia stránky • „create“ - vytvorenie stránky • „history“ - zobrazenie histórie • „view“ - zobrazenie stránky • „attachment“ - práca s prílohami V prípade inej akcie, je potrebné uviesť prázdny reťazec, alebo názov akcie. Tá je následne v prezentačnej vrstve ignorovaná. Hodnota tejto premennej ovplyvnuje zobrazenie kontroliek na prácu so stránkou.
title	Obsahuje titulok aktuálne zobrazovanej stránky.
rev	Obsahuje revíziu stránky.
url	Obsahuje URL stránky

Nastavenie týchto parametrov ovplyvňuje zobrazenie informačných hlásení a sprístupňovanie ovládacích prvkov vo wiki.

V hlavnej šablóne je definovaná premenná *screen_content.* Táto premenná je nahradená vnorenou šablónou. Opis vnorených šablón je možné nájsť v opise kontrolérov v elektronickej programátorskej dokumentácií.

Prezentačná vrstva je rozšírená o samostatné triedy zodpovedné spracovanie iného ako HTML výstupu. Konkrétne ide o triedu CaptchaView. Tá zobrazuje obrázky s textom, ktorý má byť rozpoznaný používateľom.

Servisy v systéme

- *PageService*. Zabezpečuje prácu so stránkami. Sprístupňuje rozhranie pre vytváranie a modifikáciu stránok. Súčasťou sú aj metódy na zistenie, či stránka existuje a metódy pre prácu so súborovými prílohami.
- *WikiAnalysisService*. Zabezpečuje vyhodnocovanie správania a je zodpovedný za blokovanie stránok a záznam informácií o blokovaní.

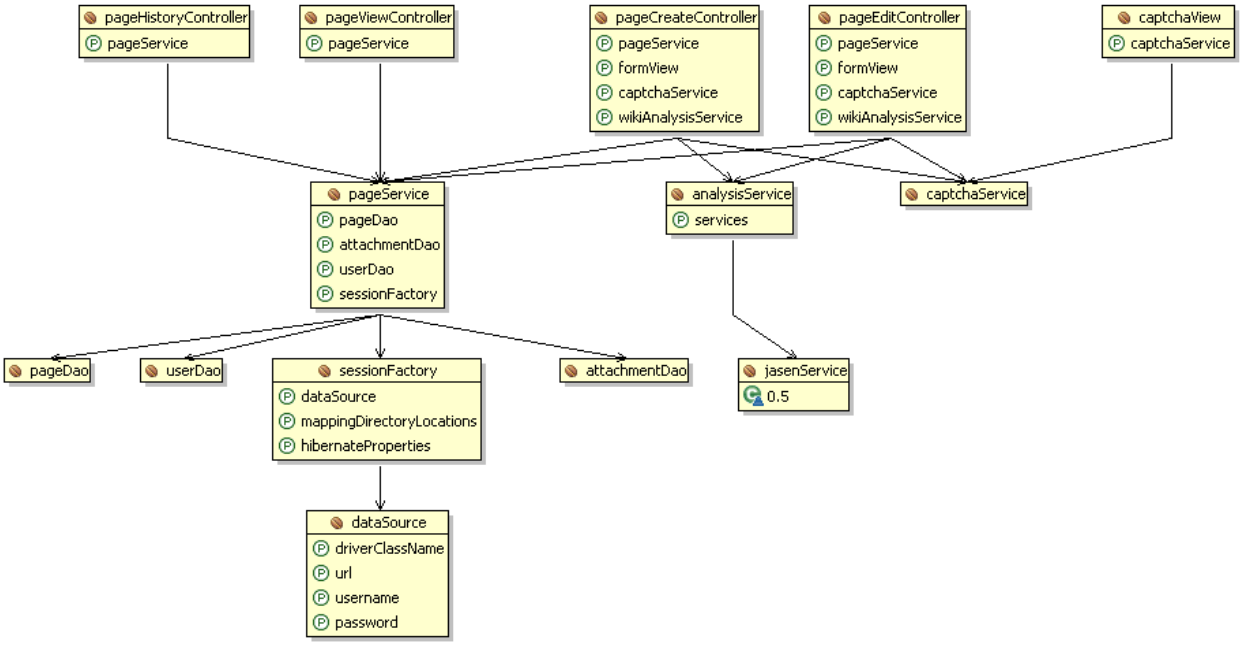
Samotné servisy realizujú určitú množinu operácií nad databázou. Funkcie nad databázou sú implementované v samostatných komponentoch, ktoré sú prístupné z viacerých servisných komponentov. Využitý je DAO návrhový vzor. Tento vzor bol nutnosť, kvôli vzniku možných cyklických závislostí, ktoré rámec Spring neviem implicitne riešiť. Nutnosťou bolo aj transakčné spracovanie operácií, ktoré prebiehajú nad viacerými DAO objektami. V systéme je manažment transakcií riešený na úrovni servisov. Nasledujúce DAO objekty sú implementované v systéme:

- *PageDao*. Zapúzdruje funkcie pre prácu so stránkami a ich históriou.
- *AttachmentDao*. Zapúzdruje funkcie pre prácu s priloženými súbormi.
- *UserDao*. Zapúzdruje funkcie pre prácu s používateľmi systému.

Súčasťou systému sú pomocné komponenty, ktoré sa nevyužívajú priamo na prácu so stránkami, ale sprístupňujú podporné funkcie za účelom zabezpečenia aplikácie a vykonávanie podporných funkcií a to CaptchaService a CaptchaView.

Prepojenie komponentov v systéme pomocou Spring rámca

Prepojenie komponentov je znázornené na obrázku 13.



Obrázok 12: Prepojenie komponentov v systéme

Komponenty zodpovedné za vytvorenie stránky a novej revízie

Verifikácia prostredníctvom CAPTCHA je spracovávaná komponentom CaptchaService. Na implementáciu tohto komponentu bol použitý projekt jCAPTCHA (jcaptcha.sourceforge.net).

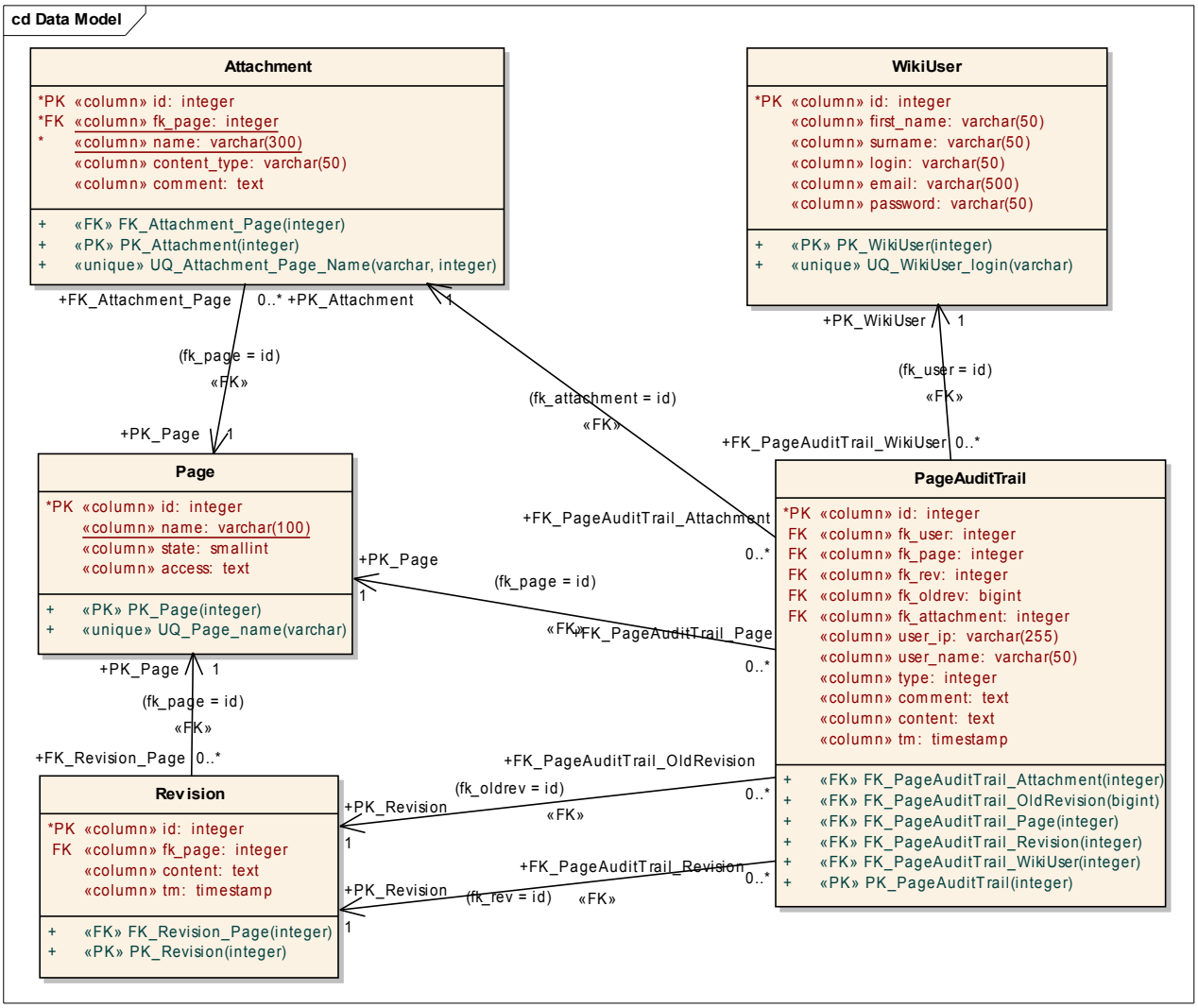
Spracovanie je rozdelené na 2 časti a to ukladanie samotnej stránky a ukladanie príloh ku stránke. Pred uložením stránky sú prílohy ukladané do dočasnej pamäte. Prílohy sú uložené v tejto pamäti počas trvania sedenia. Štandardná konfigurácia dĺžky sedenia je 30 minút a je možné ju upresniť v konfigurácií webového kontajnera. Presný popis je uvedený v používateľskej príručke.

Verifikácia pomocou CAPTCHA

CAPTCHA verifikácia je iniciovaná pri zobrazení formulára. Je vygenerované náhodné číslo a to je predané prezentačnej vrstve a umiestnená ako skrytá položka vo formulári. Tá si vyžiada obrázok, ktorý má v URL zakomponované číslo a následne je podľa tohto čísla vygenerovaný CAPTCHA obrázok (chalenge request). Po zadaní a odoslaní s vyplnenou odpoveďou na CAPTCHA obrázok je odpoveď overená prostredníctvom servisu dostupného v aplikačnom kontexte.

Dátový model

Dátový model zohľadňuje aj neimplementované vlastnosti. Mapovanie do tried mapuje tabuľku na konkrétnu triedu. Jediná výnimka je PageAuditTrail.



Page

stĺpec	*typ*	*popis*
id	Integer	Primárny kľúč tabuľky
name	Varchar(100)	Meno stránky
state	Smallint	Stav stránky – súvisí so zmenami stavu : Často editovaná stránka
access	Text	Definícia prístupových práv ku stránke

Indexy a obmedzenia

- PK_Page(id) – primárny kľúč, asociovaná sekvencia Page_Id_Seq
- UQ_Page_Name(meno) – obmedzenie na meno – musí byť unikátne

Revision

stĺpec	*typ*	*popis*
Id	Integer	Primárny klúč revízie
Fk_page	Integer	Prepojenie so stránkou
Content	Text	Obsah stránky
Tm	Timestamp	Čas vytvorenia novej revízie

Indexy a obmedzenia

- PK_Revision(id) – primárny kľúč, asociovaná sekvencia Revision_Id_Seq
- FK_Revision_Page(Fk_page) – cudzí kľúč odkazujúci na stránku, propagácia je povolená len pre operáciu delete.

Attachment

stĺpec	*typ*	*popis*
Id	Integer	Primárny kľúč
Fk_page	Integer	Prepojenie so stránkou
Name	Varchar(300)	Lokálne meno prílohy
Content_type	Varchar(50)	Typ prílohy
Comment	Text	Komentár ku prílohe

Indexy a obmedzenia

- PK_Attachment(Id) – primárny kľúč, asociovaná sekvencia Attachment_Id_Seq
- FK_Revision_Page(Fk_page) – cudzí kľúč odkazujúci na stránku, propagácia pre operáciu delete.
- UQ_Attachment_Page_Name(Name, Fk_page) – zabezpečuje unikátnosť mena prílohy pre stránku

PageAuditTrail

Táto tabuľka slúži na ukladanie histórie stránky. Obsahuje informácie o operáciách, ktoré

boli vykonávané nad stránkou.

stĺpec	*typ*	*popis*
Id	Integer	Primárny kľúč
Fk_user	Integer	Prihlásený používateľ
Fk_page	Integer	Stránka nad ktorou sa uskutočnili zmeny
Fk_rev	Integer	Revízia, ktorej sa týka zmena
Fk_oldrev	Integer	Revízia, ku ktorej sa uskutočnil návrat.
Fk_attachment	Integer	Príloha, ktorá bola pridaná
User_ip	Varchar(255)	IP adresa používateľa
User_name	Varchar(50)	Meno používateľa, ktorý spôsobil vytvorenie položky
Type	Integer	Typ položky histórie
Comment	Text	Komentár
Content	Text	Pomocný obsah – bližšia informácia o type položky v histórií
Tm	Timestamp	Čas vytvorenia položky v histórií stránky

Indexy a obmedzenia

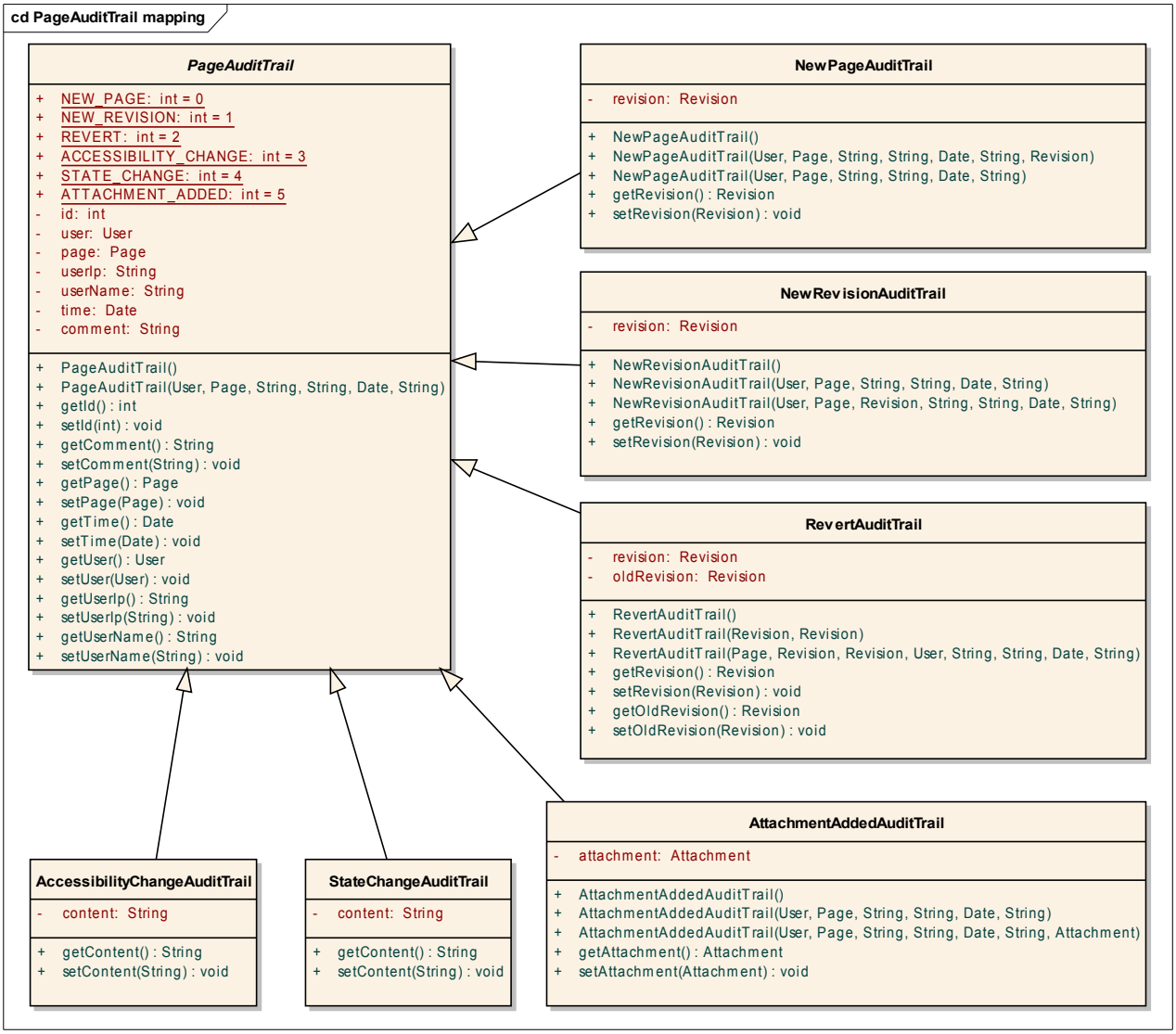
- PK_PageAuditTrail(id) – primárny kľúč, asociovaná sekvencia Pageaudittrail_id_seq.
- FK_PageAuditTrail_Attachment(Fk_attachment) – Prepojenie s tab. Attachment, propagácia operácie delete.
- FK_PageAuditTrail_OldRevision(Fk_oldrev) – Prepojenie s tab. Revision, propagácia operácie delete.
- FK_PageAuditTrail_Page(Fk_page) – Prepojenie s tab. Page, propagácia operácie delete.
- FK_PageAuditTrail_Revision(Fk_revision) – Prepojenie s tab. Revision, propagácia operácie delete.
- FK_PageAuditTrail_WikiUser(Fk_wikiuser) – Prepojenie s tab. WikiUser, propagácia operácie delete.

Mapovanie špecifických typov PageAuditTrail do tried

Pri mapovaní dedičnosti bola použitá stratégia „tabuľka-hierarchia tried“[4]. Diskriminátor typu triedy je stĺpec *Type*. Hierarchia tried je znázornená na obrázku 15.

4 Table per class hierarchy

Hodnota diskriminátora	*Mapovaná trieda*
0	sk.kmit.wiki.domain.NewPageAuditTrail
1	sk.kmit.wiki.domain.NewRevisionAuditTrail
2	sk.kmit.wiki.domain.RevertAuditTrail
3	sk.kmit.wiki.domain.AccessibilityChangedAuditTrail
4	sk.kmit.wiki.domain.StateChangeAuditTrail
5	sk.kmit.wiki.domain.AttachmentAddedAuditTrail



Obrázok 15: Hierarchia tried odvodených od PageAuditTrail

WikiUser

stĺpec	*typ*	*popis*
Id	integer	Primárny kľúč
First_name	Varchar(50)	Krstné meno používateľa
Surname	Varchar(50)	Priezvisko používateľa
Login	Varchar(50)	Prihlasovacie meno používateľa
Email	Varchar(500)	Email používateľa
Password	Varchar(50)	Zašifrované heslo

Indexy a obmedzenia

- PK_Wikiuser(Id) – primárny kľúč.
- UQ_Wikiuser(Login) – obmedzenie na unikátnosť prihlasovacieho mena v aplikácií.

PRILOHA B Používateľská príručka

Keďže ide o webovú používateľská príručka opisuje postup inštalácie a štruktúru konfiguračných súborov. Súčasťou je aj popis ovládania samotnej webovej aplikácie.

Systémové požiadavky

Systém vyžaduje hardvér potrebný na prevádzku webového kontajnera Tomcat v kooperácií s databázou PostgreSQL.

Softvérové požiadavky

Java SDK verzie 1.5
Webový kontajner Tomcat verzie 5.5.15
Databáza PostgreSQL verzie 8.1.

Inštalácia systému

Je potrebné nainštalovať programy Tomcat a PostgreSQL. Ďalší inštalačný postup predpokladá prítomnosť týchto programov. Predpokladom je aj existencia databázy v PostgreSQL.

1. *Nahratie schémy do databázy.* V programe pgAdmin sa pripojte na databázu. V menu tools je nástroj Query Tool. Po spustení sa zobrazí okno. V menu file treba vybrať položku open file a otvoriť súbor *ddl.sql* dodaný na priloženom elektronickom médiu. Následne treba načítaný skript spustiť. To je možné vykonať stlačením klávesy F5, alebo prostredníctvom menu : Query -> Execute.
2. *Nastavenie parametrov spojenia.* V adresári *bin-dist/wiki/WEB-INF/props* je potrebné editovať obsah súboru wiki.properties. Hodnoty treba nastaviť podľa nasledujúceho príkladu:

```
database.url = jdbc:postgresql://<adresa DB>:5432/<db>
database.username = <uzivatel na db>
database.password = <heslo uzivatela na db>
```

<adresa DB> je IP adresa, alebo meno počítača, na ktorom beží databázový systém PostgreSQL.

3. *Skopírovanie adresára wiki do adresára webapps programu Tomcat.*

Po uskutočnení týchto krokov a spustení programu Tomcat je systém, v prípade štandartných nastavení, dostupný na adrese http://localhost:8080/wiki/Index.html

Používanie systému

Do systému sa pristupuje cez webové rozhranie, čiže je potrebné mať nainštalovaný akýkoľvek prehliadač webových stránok, či už Internet Explorer, Mozilla, Opera, KHTML, MSN Explorer, Netscape Navigator, Firefox a podobne na príslušnom operačnom systéme.

Obrázok 16: Pohľad na prvú obrazovku

Každá stránka (obrázok 16)obsahuje hlavné menu (vľavo) a vedľajšie menu (vpravo hore), ktoré sa prispôsobuj, podľa aktuálne zobrazenej stránky. Vedľajšie menu môže obsahovať 3 položky „Zobrazenie“, „Edituj“ a „História“.

Zobrazenie príspevku

Zobrazenie príspevku (obrázok 17) je jednoduché a prehľadné. Obsahuje nadpis príspevku a číslo poslednej revízie. Samostatný obsah príspevku je tvorený rôznymi štýlmi textu, podľa voľby užívateľa. Obsahuje aktívne prepojenia na ďalšie sekcie, ktoré sú rozdelené do troch druhov:

- Existujúce príspevky – sú zvýraznené modrou farbou s ponukou zobrazenia.

- Zatiaľ neexistujúce príspevky – sú zvýraznené červenou farbou, s ponukou možnosti vytvorenia tejto sekcie, následného editovania a uloženia.
- Externé príspevky – sú zvýraznené modrou farbou so znakom .

Obrázok 17: Zobrazenie stránky v systéme

Editácia príspevku

Editácia príspevku je jednoduchá a intuitívna. Využíva spôsoby a označenia podobné väčšine editorom textu. Po úprave obsahu textu, po vyplnení komentára pre históriu a mena autora je potrebné zadať kontrolný reťazec. Zmeny v príspevku nie je možné uložiť pokiaľ nebude kontrolný reťazec zadaný správne.

Obrázok 18: Editácia príspevku

História príspevku

História príspevku zobrazuje verzie príspevkov, ktoré boli postupne menené. Po kliknutí na „Čas zmeny“ je možné otvoriť danú verziu príspevku, editovať ju a uložiť ako aktuálnu.

Obrázok 19: Zobrazenie histórie príspevku