

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

GENERÁTOR VĚDECKÝCH WEBOVÝCH PORTÁLŮ

BAKALÁŘSKÁ PRÁCE

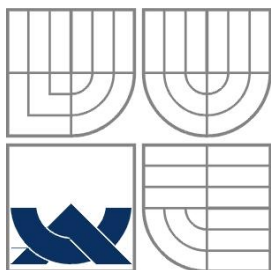
BACHELOR'S THESIS

AUTOR PRÁCE

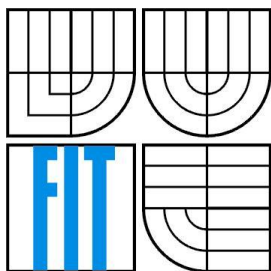
AUTHOR

PETER PEKAROVIČ

BRNO 2014



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

GENERÁTOR VĚDECKÝCH WEBOVÝCH PORTÁLŮ

SCIENTIFIC WEB PORTAL GENERATOR

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

PETER PEKAROVIČ

VEDOUCÍ PRÁCE
SUPERVISOR

doc. RNDr. PAVEL SMRŽ, Ph.D.

BRNO 2014

Abstrakt

Tato práce se zabývá implementací vědeckého webového portálu s možností automatického vytváření obsahu portálu. Popisuje možnosti využití upozorňovacích služeb a získávání publikací pomocí těchto služeb, analýzu klíčových slov a kategorizaci publikací. V práci je navrhnutý a implementovaný systém pro obsluhu vědeckého portálu i samotný portál.

Abstract

This thesis is concerned with implementation of scientific web portal and possibility of automatic generating of portal content. It describes possibilities of usage of alert services to acquire publications, analysis of keywords and categorization of publications. Thesis contains design and implementation of system for portal administration and portal itself.

Klíčová slova

webový portál, upozorňovací služby, vědecká publikace, analýza klíčových slov, kategorizace publikace

Keywords

web portal, alert services, scientific publication, keyword analysis, categorization of publication

Citace

Peter Pekarovič: Generátor vědeckých webových portálů, bakalářská práce, Brno, FIT VUT v Brně, 2014

Generátor vědeckých webových portálů

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana doc. RNDr. Pavla Smrža, Ph.D. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Peter Pekarovič
21. května 2014

Poděkování

Tímto bych chtěl poděkovat panu doc. RNDr. Pavlovi Smržovi, Ph.D. za užitečné rady a připomínky při vypracovávání této práce. Dále bych chtěl poděkovat rodině a přítelkyni za podporu při vypracovávání práce a kolegům ze společnosti AVG Technologies za odbornou pomoc při implementaci.

© Peter Pekarovič, 2014.

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
1 Úvod.....	3
2 Webový portál.....	4
2.1 Vedecké webové portály.....	4
3 Upozorňovacie služby.....	5
3.1 Typy upozornení.....	5
3.1.1 Uložené vyhľadávanie	5
3.1.2 Tabuľka obsahu	6
3.1.3 Citovanie.....	6
4 Návrh.....	7
4.1 Návrh administrátorského systému.....	8
4.1.1 Grafické užívateľské rozhranie.....	8
4.1.2 Získavanie publikácií.....	8
4.1.3 Operácie s publikáciami.....	10
4.1.4 Synchronizácia publikácií.....	12
4.2 Návrh klientskeho systému.....	13
4.2.1 Grafické užívateľské rozhranie.....	14
4.2.2 Inicializácia portálu.....	14
4.2.3 Vyhľadávanie publikácií.....	14
4.2.4 Zobrazenie výsledkov vyhľadávania	16
5 Implementácia.....	17
5.1 Použité jazyky, prostredie a nástroje	17
5.2 Databáza	17
5.3 Implementácia administrátorského systému	18
5.3.1 Databázový model administrátorského systému.....	18
5.3.2 Grafické užívateľské rozhranie.....	19
5.3.3 Získavanie publikácií.....	21
5.3.4 Operácie s publikáciami.....	24
5.3.5 Synchronizácia.....	27
5.4 Implementácia klientskeho systému	28
5.4.1 Databázový model klientskeho systému.....	29
5.4.2 Grafické užívateľské rozhranie.....	30
5.4.3 Zobrazovanie informácií z databázy	32
5.4.4 Vyhľadávanie publikácií.....	35

6	Testovanie	36
7	Záver	37
7.1	Zhodnotenie práce	37
7.2	Možné rozšírenia	37
7.2.1	Kvalitnejší PDF parser.....	37
7.2.2	Efektívnejšie získavanie kľúčových slov.....	37
7.2.3	Podpora viacerých zdrojov	38
7.2.4	Automatizované testy	38
7.2.5	Spracovanie chýb.....	38
7.2.6	Vytváranie užívateľských účtov	38
7.2.7	Implementovanie modulov Statistics, Help a About	38
A	Obsah CD.....	41
B	Manuál	42
B.1	Príprava databázy.....	42
B.2	Príprava webového serveru.....	42
B.3	Demonštračné dáta.....	43

1 Úvod

Žijeme v dobe, v ktorej sú informácie čoraz viac šírené prostredníctvom internetu. Tieto informácie sú často rozptýlené na mnohých miestach a používateľ internetu je nútený ich vyhľadávať vyčerpávajúcim a časovo náročným spôsobom.

Túto situáciu sa snažia riešiť webové portály, ktoré hromadia informácie z mnohých zdrojov a ponúkajú ich v jednotnej forme užívateľom portálu. Existujú portály, ako napríklad webový portál Yahoo! [1], ktoré sa snažia zaujať veľmi široké spektrum užívateľov a ponúkajú im obsah z rôznych oblastí záujmu. V opozícii stoja portály orientované na špecifické publikum, zamerané na jedinú oblasť záujmu. Príkladom takéhoto portálu je motoristický portál Automobil Revue [2].

Cieľom tejto bakalárskej práce je navrhnuť a implementovať vedecký webový portál s podporou automatického získavania obsahu, ktorý je ponúkaný užívateľom. Obsah portálu tvoria publikácie z rôznych vedeckých disciplín, ktoré portál získava z relevantných zdrojov. Využíva popritom upozorňovacie služby, ktoré poskytujú vydavateľstvá, vedecké časopisy alebo iné vedecké portály.

V práci je navrhnuté a implementované prostredie pre administrátora portálu, ktorý prostredníctvom administrátorského systému získané publikácie prijíma alebo odmieta, v prípade prijatých publikácií systém pomáha publikáciu kategorizovať a navrhuje administrátorovi kľúčové slová, ktoré publikáciu charakterizujú. Spolu s ďalšími získanými informáciami o publikácii administrátor publikáciu ukladá do databáze portálu.

Takisto sa v práci nachádza návrh a implementácia klientskeho systému, ktorý publikácie zobrazuje užívateľom, umožňuje im publikácie vyhľadávať, filtrovať a prezerat' ich elektronickú podobu.

Úvod do problematiky webových portálov a upozorňovacích služieb pokrýva druhá (kap. 2) a tretia kapitola (kap. 3) práce. Tieto kapitoly poskytujú teoretický základ potrebný na návrh a implementáciu vedeckého webového portálu.

Návrh portálu, ktorý sa nachádza v štvrtej kapitole (kap. 4), je rozdelený na dve hlavné časti, a to návrh administrátorského systému (kap. 4.1) a návrh klientskeho systému (kap. 4.2).

Implementácia návrhu je umiestnená v piatej kapitole (kap. 5). Za ňou nasleduje kapitola (kap. 6), ktorá popisuje testovanie oboch systémov.

V poslednej kapitole (kap. 7) sa nachádza zhrnutie výsledkov práce a jej možné rozšírenia do budúcnosti.

2 Webový portál

Webovým portálom môžeme najčastejšie nazvať webovú stránku, ktorá zberá informácie z rôznych zdrojov a prezentuje ich užívateľovi jednotným spôsobom.

Myšlienka portálu je nazberané informácie dávať dokopy a vytvoriť jediný bod prístupu (angl. single point of access) k informáciám, vytvoriť akúsi knižnicu kategorizovaného a užívateľmi prispôsobeného obsahu [3]. Úlohou portálu by malo byť:

- Automaticky poskytovať informácie užívateľom vzhľadom na ich rolu
- Odporúčať užívateľom ďalšie informácie a ponúkať možnosť si získavané informácie prispôbovať svojim potrebám

Prispôbovanie obsahu portálu je dôležitá súčasť portálov a každý užívateľ by mal byť schopný dostávať len tie informácie z portálu, ktoré skutočne potrebuje. Portály preto umožňujú užívateľom vytvárať účty, kde si môžu modifikovať nastavenia portálu.

Portály by takisto mali disponovať efektívnym vyhľadávacím modulom, ktorý umožňuje užívateľom prehľadávať obsah portálu podľa užívateľských kritérií [4].

Väčšina dnešných portálov poskytuje informácie iba o úzkej skupine oblastí a sústredia sa na špecifickú skupinu užívateľov.

Na základe tohto môžeme portály rozdeliť na:

1. Horizontálne
2. Vertikálne

Horizontálne portály sa sústredia na celú internetovú komunitu užívateľov. Tieto webové stránky, mnohokrát nazývané aj ako „megaportály“, obvyčajne disponujú vyhľadávacími modulmi a poskytujú užívateľom možnosť prispôbovať si obsah tým, že im ponúkajú široké spektrum služieb (informácie o počasí, novinky zo sveta, trendy na trhoch a podobne) [3].

Takýmito portálmi sú napríklad portály Yahoo! a Lycos.

Vertikálne portály poskytujú informácie a služby užšej špecifickej komunite užívateľov a väčšinou poskytujú informácie iba o jednej oblasti záujmu. Takýmito portálmi sú napríklad vedecké webové portály, portály o automobiloch, o móde a iné.

Príkladom takýchto portálov je napríklad motoristický portál Automobil Revue.

2.1 Vedecké webové portály

Vedecké webové portály sú vertikálne portály, ktoré poskytujú užívateľom, typicky vedcom alebo študentom, informácie zo sveta vedy. Poskytujú im jednoduchý prístup k najnovším informáciám z rôznych vedeckých oblastí.

Populárnymi vedeckými portálmi sú napríklad BioMedNet, PubMed, ScienceDirect a AgNIC.

3 Upozorňovacie služby

„Vo všeobecnosti upozornenia zodpovedajú notifikáciám, keď niečo, čo vás zaujíma bolo práve publikované alebo pridané do databázy alebo na web stránku.” [5] Mnohé vedecké databázy a časopisy poskytujú automatizované upozorňovacie služby, ktoré notifikujú užívateľov, ako napríklad výskumných pracovníkov, o nových publikáciách, citáciách a tieto notifikácie posielajú užívateľom elektronickou poštou alebo pomocou RSS kanálu.

Upozorňovacie služby môžu byť prispôbosené podľa potrieb užívateľa a pomáhajú užívateľovi držať krok s aktuálnymi výskumami a literatúrou v rôznych oblastiach záujmu. Definícia, terminológia a frekvencia notifikovania sa líši od vydavateľa k vydavateľovi.

Upozorňovanie by malo byť automatické, znižovať potrebu pravidelne navštevovať webové stránky a vedecké databázy. Viacero poskytovateľov upozorňovacích služieb vyžaduje, aby užívateľ mal založený profil u poskytovateľa. Existuje viacero typov upozornení:

- Uložené vyhľadávanie (angl. Saved search alert)
- Tabuľka obsahu (angl. Table of contents alert – TOC)
- Citovanie (angl. Citation alert)

3.1 Typy upozornení

3.1.1 Uložené vyhľadávanie

„Uložené vyhľadávanie je notifikácia užívateľa o nových článkoch, ktoré sú výsledkom vyhľadávacieho dotazu.” [6]

Uložené vyhľadávanie je typ upozornenia, kde užívateľ vytvára vyhľadávací dotaz, ktorý je opakovane spúšťaný nad vedeckou databázou. Vyhľadávací dotaz je spúšťaný v pravidelných cykloch na dennej, týždennej alebo mesačnej frekvencii.

Výsledky dotazu sú posielané užívateľovi na email alebo na RSS kanál. Samotným obsahom výsledku sú dáta o publikáciách, ako napríklad abstrakt, názov publikácie, dáta o citáciách alebo odkazy na samotné publikácie. Výsledky sa líšia u každého poskytovateľa tejto služby.

Rôzne databázy poskytujú rozdielne obmedzenia na počet vyhľadávacích dotazov dostupných užívateľovi.

Na vytvorenie vyhľadávacieho dotazu sú väčšinou potrebné tieto kroky [7] :

1. Pripojiť sa k databáze cez webovú stránku poskytovateľa
2. Vytvorenie profilu resp. užívateľského účtu
3. Vytvorenie dotazu podľa špecifikácie od poskytovateľa
4. Zvolenie frekvencie opakovania dotazu a formu zasielania výsledkov

3.1.2 Tabuľka obsahu

Mnoho výskumných oblastí disponujú kľúčovými časopismi, ktoré publikujú najdôležitejšie a najzaujímavejšie publikácie. Ak užívateľ začne využívať upozornenia o tabuľke obsahu, tak tabuľka obsahu je odosielaná užívateľovi vždy, keď je publikované nové číslo časopisu. [\[8\]](#)

Tabuľka obsahu je spoľahlivá metóda, ako užívateľa informovať o aktuálnych výskumoch na akademickej pôde.

Dáta môžu byť užívateľovi posielané na RSS kanál alebo prostredníctvom emailu.

3.1.3 Citovanie

Táto služba upozorňuje užívateľa o citáciách vzťahujúcich sa buď ku konkrétnemu autorovi alebo ku konkrétnej publikácii.

Ak je autor citovaný alebo sa nachádza v použitej literatúre publikácie, tak užívateľovi je poslaná notifikácia s odkazom na danú publikáciu.

Ak je publikácia citovaná v inej publikácii alebo sa nachádza v použitej literatúre, tak je podobne ako pri autorovi, užívateľovi zaslaná notifikácia na danú publikáciu. [\[9\]](#)

4 Návrh

Navrhujeme systém, ktorý dokáže vyhľadávať vedecké publikácie v externých zdrojoch a tieto publikácie ukladať do databáze. Zdrojom publikácií sú vydavateľstvá, vedecké časopisy a podobne. Následne je schopný tieto publikácie zobrazovať v pokročilej forme prezentácie a dokáže publikácie filtrovať na základe rôznych užívateľských kritérií. Zobrazovanie publikácií a získavanie publikácií môžeme rozdeliť do dvoch logických jednotiek resp. podsystémov, a to:

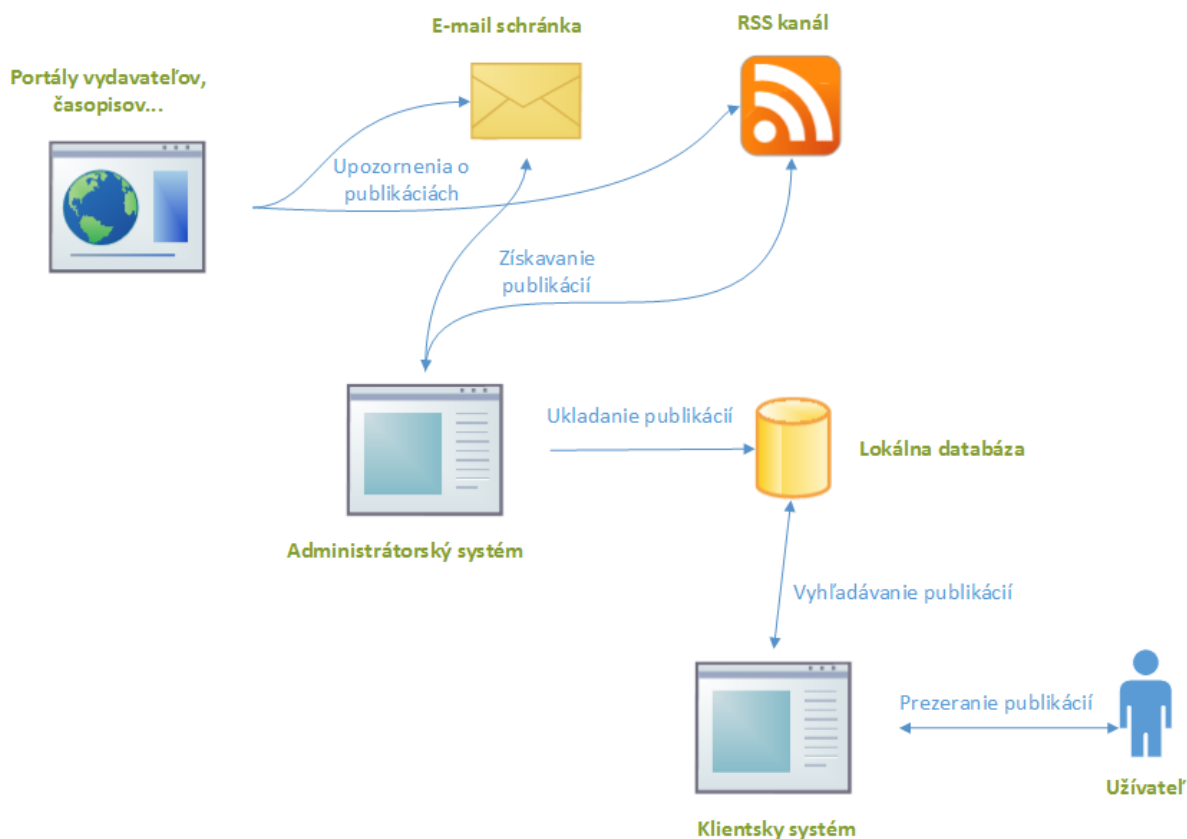
1. Administrátorský systém
2. Klientsky systém

Administrátorský systém slúži na správu portálu a vytváranie obsahu. Systém komunikuje s externými zdrojmi publikácií, z ktorých publikácie získava a s lokálnou databázou portálu, kde publikácie ukladá. Táto databáza je spoločná aj pre administrátorský systém a aj pre klientsky systém.

K administrátorskému systému má prístup iba správca portálu a nie je dostupný užívateľom portálu.

Klientsky systém slúži na zobrazovanie obsahu portálu a je verejne dostupný užívateľom. Užívatelia sú schopní vyhľadávať publikácie z lokálnej databázy a ak je to možné, tak aj prezerat obsah publikácií v elektronickej podobe, či publikáciu stiahnuť.

Komunikáciu medzi oboma systémami a externými zdrojmi popisuje nasledujúci obrázok.



Obr. 4.1 Komunikácia medzi systémami portálu a externými zdrojmi

Portály vydavateľstiev, časopisov a iné zdroje publikácií, ktoré poskytujú upozorňovacie služby (kap. 3), zasielajú upozornenia o nových publikáciách odoberateľom formou elektronickej pošty alebo na RSS kanál. Odtiaľ administrátorský systém upozornenia získava a ich spracovaním získava samotné publikácie, ktoré na podnet od administrátora ukladá do databáze portálu. Klientsky systém na podnet od užívateľa portálu v tejto databáze publikácie vyhľadáva a následne zobrazuje užívateľovi.

4.1 Návrh administrátorského systému

Administrátorský systém portálu slúži na spravovanie portálu, osobitne na generovanie obsahu portálu, a to prostredníctvom automatického získavania publikácií pre daný portál. Užívateľ tohto systému je administrátor, ktorý systém obsluhuje.

Administrátor prostredníctvom systému určuje zdroje, z ktorých bude systém publikácie čerpať a spúšťa proces, ktorý publikácie získava. Získané publikácie sú ponúknuté administrátorovi a ten za pomoci systému nimi naplňuje databázu portálu.

Systém poskytuje všetky prostriedky potrebné od získania publikácií z externých zdrojov až po efektívne uloženie do databáze. Tie sú po procese synchronizácie zobrazované v klientskej časti portálu.

4.1.1 Grafické užívateľské rozhranie

Systém používa grafické užívateľské rozhranie (GUI) v anglickom jazyku. Po vstupe do systému si administrátor vyberá z horizontálneho menu, umiestneného v hornej časti, rôzne moduly systému.

Administrátorovi sú k dispozícii nasledujúce moduly:

1. Modul na získavanie publikácií
2. Modul na synchronizácie
3. Nastavenia

Po vstupe do modulu na získavanie publikácií je administrátorovi k dispozícii tlačidlo, ktorým odštartuje proces získavania informácií. Výsledky sú zobrazené vo forme tabuľky. Jednotlivé záznamy obsahujú tlačidlá, pomocou ktorých administrátor spúšťa akcie ako prijatie článku alebo odmietnutie článku.

Prijatie článku prebieha vo viacerých krokoch formou modálneho dialógového okna.

Po vstupe do modulu na synchronizácie je administrátorovi k dispozícii tabuľka zobrazujúca synchronizačné balíky (kap. 4.1.4.1). Detail jednotlivých balíkov sa administrátorovi zobrazuje formou modálneho dialógového okna.

V module nastavení sú administrátorovi zobrazené tabuľky s dátami potrebnými na získavanie publikácií. Tieto dáta sa dajú editovať, vytvárať nové alebo mazať. Tieto úpravy prebiehajú v dialógových oknách.

4.1.2 Získavanie publikácií

Obsah portálu je tvorený z publikácií, ktoré systém získava automaticky, a to tak, že odoberá publikácie z rôznych vedeckých portálov, vydavateľstiev a časopisov. Následne systém za pomoci

administrátora tieto publikácie spracováva, získava publikáciu v elektronickej podobe a ukladá do databáze portálu.

Systém využíva upozorňovacie služby (kap. 3), poskytované vydavateľmi, webovými portálmi a inými využívanými zdrojmi, ktoré odoberateľa informujú o výskyte nových publikácií.

Publikácie sú odoberateľovi odosielané najčastejšie dvoma kanálmi:

- RSS (Really Simple Syndication)
- E-mail

Automatické získavanie publikácií prebieha na vyžiadanie od administrátora portálu, ktorý túto akciu systému iniciuje.

4.1.2.1 RSS

Niektoré portály, vydavatelia alebo časopisy umožňujú vytvoriť odoberateľom vlastný RSS kanál, kde sú odoberatelia informovaní o publikáciách vybraných na základe kritérií, ktoré si užívateľ zvolil (kap. 3.1.1).

Informácie o publikácii sú uložené vo formáte RSS, ktorý je dialektom jazyka XML. Systém využíva ten fakt, že RSS zodpovedá špecifikácií XML 1.0 a špecifikácií RSS 2.0, ktorá prikazuje, aby XML obsahovalo povinné elementy *Title*, *Link* a *Description*.

Systém získa z elementu *Title* názov publikácie, z elementu *Link* odkaz na elektronicкую podobu publikácie a z elementu *Description* popis publikácie. Ostatné elementy systém zobrazí administrátorovi, ktorý ich môže použiť napríklad na získanie informácií o autoroch publikácie, roku vydania publikácie a podobne.

4.1.2.2 E-mail

Informovanie odoberateľa zasielaním nových publikácií formou elektronickej pošty je najčastejšou využívanou metódou vydavateľstiev a vyplýva z povahy upozorňovacích služieb. Nevýhodou je fakt, že každý poskytovateľ tejto služby používa vlastný formát správy a nie je možné vytvoriť univerzálny spôsob na získavanie informácií o publikáciách z rôznych zdrojov.

Systém preto využíva službu Google Scholar [10], ktorá vyhľadáva publikácie u širokého spektra vydavateľstiev, vedeckých portálov a podobne. Následne Google Scholar informuje pomocou upozorňovacích služieb odoberateľov o týchto publikáciách.

Systém spracováva informácie o publikáciách od Google Scholar a tieto publikácie sú ponúknuté administrátorovi portálu.

4.1.2.3 Definovanie zdrojov

Systém nezískava žiadne publikácie až do doby, kedy administrátor nevytvorí a nedefinuje externé zdroje, z ktorých má systém publikácie čerpať. Je možné vytvoriť a definovať ľubovoľné RSS kanály, pri ktorých je nutné zadať URL k RSS kanálu a názov kanálu.

Aby systém mohol získavať publikácie získané prostredníctvom elektronickej pošty, systém umožňuje priradiť emailovú adresu do systému a vytvoriť emailový zdroj publikácií, kde je nutné zadať odosielať emailov a názov kanálu. Na tento email budú posielané publikácie službou Google Scholar.

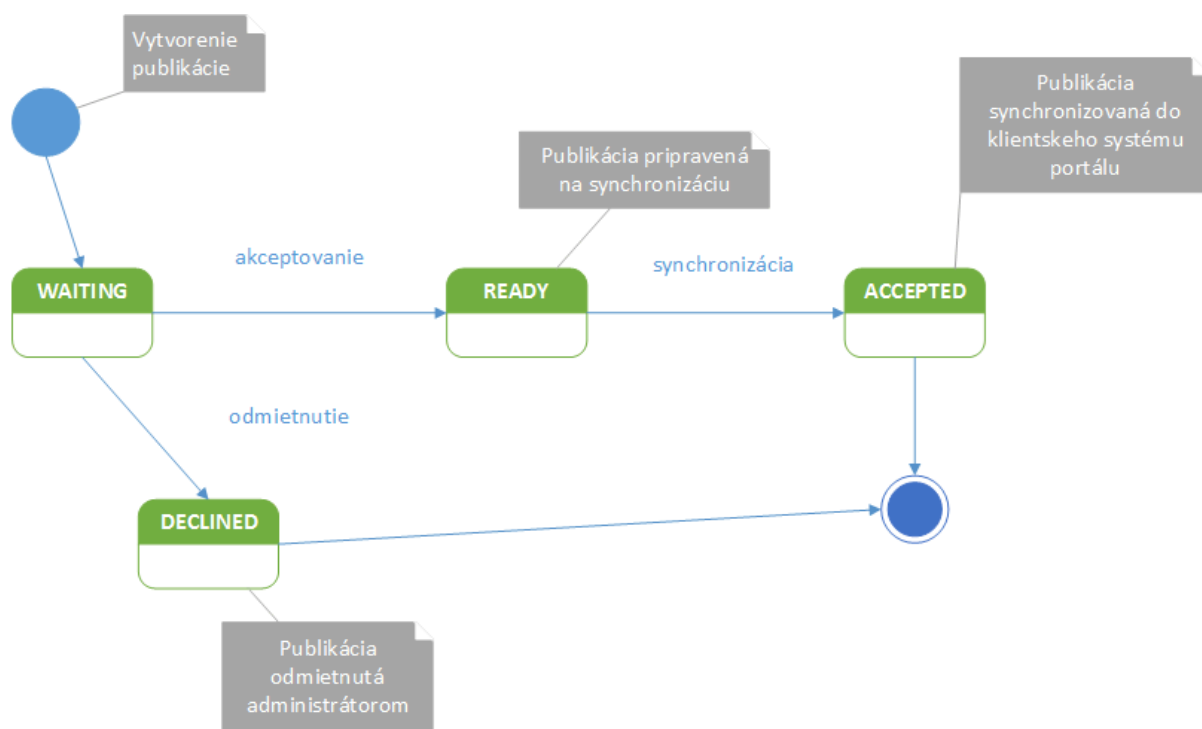
Systém podporuje spracovanie emailov z emailových účtov, ktoré má pod správou služba Gmail.

Tomuto všetkému predchádza samotné vytvorenie účtov na vedeckých portáloch, portáloch vydavateľstiev a podobne.

4.1.3 Operácie s publikáciami

S publikáciami, ktoré administrátorský systém získal z externých zdrojov, môže administrátor v zásade vykonať dve akcie. Prvou je akceptovanie publikácie, druhou odmietnutie publikácie. Publikácie sú ponúkané administrátorovi dovtedy, dokiaľ jednu z týchto akcií nevykoná.

V závislosti od vykonaných akcií sa publikácia môže nachádzať v rôznych stavoch, ktoré popisuje nasledujúci stavový diagram.



Obr. 4.2 Stavy publikácie

Pri vytvorení publikácie sa publikácia dostáva do stavu WAITING, kedy čaká na užívateľskú akciu. Zároveň to znamená, že publikácia sa zobrazuje v systéme administrátorovi.

Ak administrátor publikáciu akceptuje, nasleduje séria úkonov, ktoré musí administrátor vykonať, aby bola publikácia zobrazená v klientskej časti portálu, teda v klientskom systéme. Jednotlivé úkony sú popísané v nasledujúcich kapitolách. Po tom ako administrátor vykoná všetky úkony okrem samotnej synchronizácie, ktorá je vykonávaná separátne, tak sa publikácia dostáva do statusu READY a je pripravená na synchronizáciu, a to tak, že je pridaná do synchronizačného balíku.

Ak administrátor publikáciu odmietne, stav publikácie sa zmení na DECLINED a s publikáciou už nie je možné urobiť žiaden úkon.

4.1.3.1 Prijatie publikácie

Získanie metainformácií o publikácii

Po tom ako sa administrátor rozhodne publikáciu akceptovať, systém za pomoci administrátora získa základne informácie (tzv. metainformácie) o publikácii. Toto sa deje vo viacerých krokoch, kedy systém je v roli hľadača informácií a administrátor celý tento beh koriguje a informácie podľa potreby upravuje do správnej syntaktickej, či sémantickej podoby.

Získanie elektronickej podoby publikácie vo formáte PDF

Každá publikácia, ktorá je ponúknutá administrátorovi na schválenie obsahuje odkaz na originálny zdroj publikácie alebo priamo na elektronicкую podobu publikácie. Systém sa snaží získať adresu URL elektronickej podoby publikácie a keď sa mu to nepodarí, tak administrátor má možnosť túto informáciu doplniť, a tým pádom umožniť systému pokračovať s hľadaním ďalších informácií o publikácii. Získanie PDF publikácie je nutná podmienka, a preto portál je schopný pracovať iba s publikáciami, ktoré existujú vo forme PDF. Jedná sa hlavne o neplatené publikácie.

Získanie kľúčových slov publikácie

Systém získava kľúčové slová publikácie tromi rôznymi spôsobmi. Prvým spôsobom je získavanie metainformácií, ktoré obsahuje každý PDF súbor. Medzi týmito informáciami sú aj kľúčové slová obsahu súboru. Táto informácia však býva zastúpená v metainformáciách PDF súboru iba minoritne.

Druhým a tretím spôsobom je získavanie kľúčových slov zo samotného obsahu PDF súboru. Na túto činnosť využíva systém konvertor, ktorý konvertuje PDF súbor do reťazca a systém potom s ním pracuje ako s reťazcom.

Pri druhom spôsobe sa systém snaží nájsť kľúčové slová v úvode publikácie, a to tak, že hľadá slovo „keyword“ a jeho rôzne formy. Ak takéto slovo nájde, snaží sa získať zoznam slov, ktoré sa nachádzajú ihneď za týmto slovom. Tento zoznam pokladá systém za kľúčové slová.

Tretí spôsob je hľadanie početnosti slov zastúpených v texte publikácie. PDF súbor vo forme reťazca sa upravuje do vhodnej podoby týmito krokmi:

1. Reťazec sa prevedie do kódovania UTF-8
2. Vymažú sa interpunkčné znamienka, symboly, čísla a nepotrebné biele znaky
3. Reťazec sa rozdelí na jednotlivé slová

Získanie početností slov prebieha nasledujúcimi fázami:

1. Slová sa prevedú na základný tvar (angl. stemming)
2. Zhodné slová sa spočítajú
3. Spočítane slová sa prevedú naspäť do pôvodnej formy

Prevod základného slova naspäť do pôvodnej formy sa deje tak, že sa spočíta početnosť všetkých slov, ktoré majú zhodný základ a vyberie sa kľúčové slovo s najväčšou početnosťou.

Po týchto všetkých krokoch je k dispozícii zoznam slov s početnosťou, na ktoré sa ešte aplikuje filter, ktorý odstráni predložky, citoslovce a iné nevhodné slová. Sú to slová s nulovou informačnou hodnotou.

Z kľúčových slov, ktoré systém získal pomocou troch vyššie popísaných spôsobov sa odstránia duplikáty a zoradené podľa početnosti, sú kľúčové slová ponúknuté užívateľovi. Užívateľ môže tieto slová upravovať, odstrániť, poprípade doplniť zoznam o vlastné kľúčové slová.

Kategorizácia publikácie

Každá publikácia je zaradená do užívateľom definovaných kategórií. Kategorizácia publikácie sa deje automaticky na základe získaných kľúčových slov alebo sú kategórie manuálne doplnené užívateľom.

Systém v databáze uchováva záznamy, ktoré vyjadrujú vzťahy medzi slovotvorným základom kľúčových slov a kategóriami. Pri každom pridaní článku do portálu sa posilní vzťah všetkých slovotvorných základov kľúčových slov s kategóriami daného článku. Automatické kategorizovanie prebieha v nasledujúcich krokoch:

1. Kľúčové slová článku systém prevedie na slovotvorné základy
2. Systém zistí, či slovotvorné základy existujú v databáze systému
3. Ak existujú, tak systém vyberie všetky kategórie, ktoré sa k slovotvorným základom vzťahujú
 - a. Kategórie sa zoradia podľa frekvencie výskytu a sily vzťahu, ktoré majú so slovotvornými základmi
 - b. Kategórie sú ponúknuté administrátorovi, ktorý ich má možnosť použiť pri kategorizácii článku, upraviť alebo nepoužiť
4. Ak neexistujú, tak administrátor článok kategorizuje manuálne

Z daného návrhu vyplýva, že prvé články pridávané do systému bude administrátor kategorizovať manuálne a s každým ďalším článkom sa systém bude učiť kategorizovať články automaticky a s väčšou presnosťou.

Príprava publikácie na synchronizáciu

Ak sa administrátor rozhodol ponúknuť publikáciu prijať a v spolupráci so systémom získal o nej metainformácie, tak publikácia je pripravená k synchronizácii do klientskeho systému. Publikácia je zaradená do synchronizačného balíku (kap. [4.1.4.1](#)).

4.1.4 Synchronizácia publikácií

Publikácie, ktoré administrátor akceptoval, nie sú automaticky pridávané do databázy, z ktorej je portál tvorený, resp. nie sú zobrazované, ani nijakým spôsobom zobraziteľné v klientskom systéme portálu.

Toto platí až do momentu, kedy sú články synchronizované z administrátorského systému do klientskeho systému. Synchronizácia neprebieha v pravidelných intervaloch automaticky, ale je nutné ju vyvolať manuálne z administrátorského systému.

4.1.4.1 Synchronizačné balíky

V systéme sa vytvárajú synchronizačné balíky, ktoré obsahujú množinu článkov, ktoré administrátor prijal. Každý balík, ktorý už neprešiel úspešnou synchronizáciou (synchronizačný balík sa nenachádza v stave PROCESSED), má možnosť byť synchronizovaný do klientskeho systému.

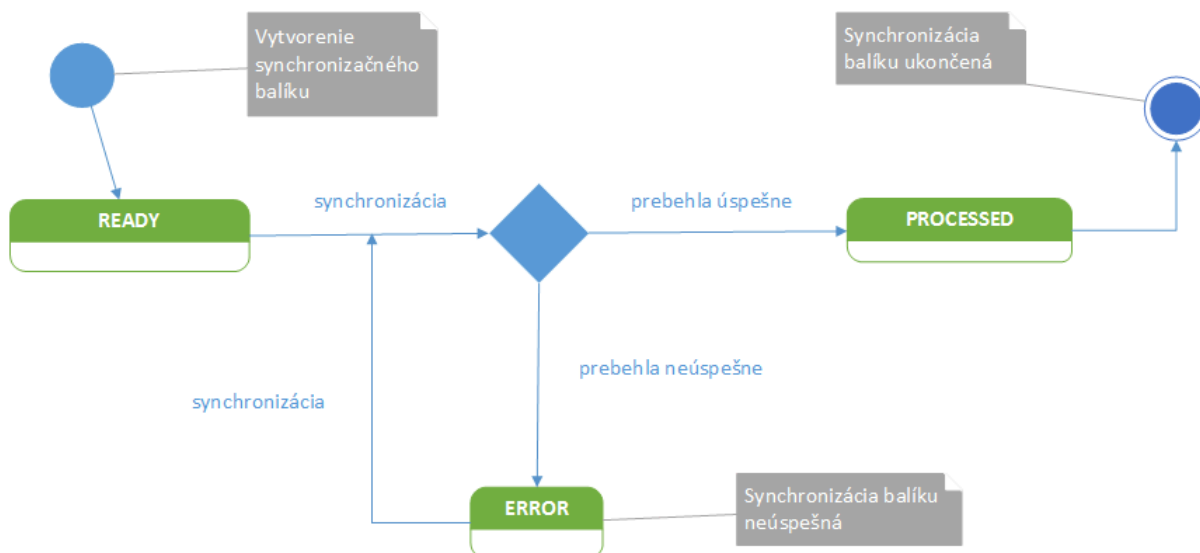
Balíky sa vytvárajú iba v prípade, že administrátor akceptoval publikáciu, a to na základe splnenia všetkých nasledujúcich podmienok:

1. V prípade, že neexistuje žiadny balík, ktorého dátum vytvorenia je menší ako 7 dní od dátumu, kedy je publikácia akceptovaná

2. V prípade, že takýto balík existuje a prešiel úspešnou alebo neúspešnou synchronizáciou (balík sa nachádza v stave PROCESSED, resp. ERROR)

Z predchádzajúcich tvrdení vyplýva, že nie je možné vytvoriť synchronizačný balík, ktorý neobsahuje žiadny článok a takisto nie je možné, aby v rozmedzí 7 dní existoval viac ako jeden balík, ktorý ešte neprešiel synchronizáciou. Pri takto frekventovanom vytváraní balíkov je pravdepodobnejšia úspešnosť synchronizácie publikácie z administrátorského systému do klientskeho systému.

Stavový diagram nižšie popisuje zmenu stavov synchronizačného balíku od momentu vytvorenia až po moment úspešnej synchronizácie.



Obr. 4.3 Stavy synchronizačného balíku

Po vytvorení balíku sa balík nachádza v stave **READY**, to znamená, že nič nebráni tomu, aby balík prešiel synchronizáciou. Samotná synchronizácia balíku prebieha tak, že administrátor v detaile vybraného synchronizačného balíku vyvolá synchronizáciu.

Ak synchronizácia prebehne úspešne, balík sa dostáva do koncového stavu **PROCESSED**. V tomto momente sú publikácie dostupné v klientskom systéme a synchronizačný balík už nie je možné znovu synchronizovať. Publikácie sa dostávajú do stavu **ACCEPTED**. Stavový diagram publikácií je popísaný v inej kapitole (obr. 4.2).

Ak synchronizácia prebehne neúspešne, balík sa dostáva do stavu **ERROR**. Žiadna z publikácií nachádzajúcich sa v tomto balíku nie je dostupná klientskemu systému. Synchronizačný balík nesie v sebe informáciu o tom, ktorá publikácia spôsobila zlyhanie synchronizácie.

Synchronizačný balík v stave **ERROR** je možné znovu synchronizovať, a tým sa pokúsiť o úspešnú synchronizáciu.

4.2 Návrh klientskeho systému

Portál bude disponovať okrem administrátorského systému, ktorý slúži na správu portálu a generovanie obsahu portálu aj systémom klientskym, ktorý bude slúžiť užívateľom ako centrum vedeckých publikácií. Klientsky systém bude slúžiť na zobrazovanie publikácií z lokálnej databázy portálu.

Užívateľ bude schopný publikácie vyhľadávať a vyhľadávanie si prispôsobovať podľa svojich potrieb. Výsledky budú užívateľovi prezentované v pokročilej forme, užívateľ bude môcť na výsledky aplikovať filtre, bude schopný zobrazovať detaily výsledkov a prezerat' elektronickú formu publikácie v prípadoch, kedy to bude možné.

4.2.1 Grafické užívateľské rozhranie

Systém disponuje grafickým užívateľským rozhraním v anglickom jazyku. Hlavnou časťou rozhrania je vyhľadávacie formulárové pole, ktoré prijíma textový vstup od užívateľa. Obsahom tohto vstupu je zoznam kľúčových slov, podľa ktorých sú užívateľovi zobrazené relevantné výsledky z databáze. Pod formulárovým poľom sa nachádza priestor na dopĺňovanie informácii ohľadom zadávaného vstupu od užívateľa. V ľavej hornej časti rozhrania sa nachádza zoznam kategórií, pod ktoré spadajú jednotlivé publikácie. V pravej hornej časti sa nachádza zoznam autorov, ktorí publikovali najviac článkov v danej oblasti výskumu.

Výsledky sú prezentované užívateľovi pod vyhľadávacím polom, a to vo forme tabuľky, ktorá obsahuje záznamy o jednotlivých publikáciách. V danej tabuľke sú zobrazené iba základné informácie o publikáciách.

Systém zobrazuje detaily o publikáciách alebo autoroch formou modálnych okien. Po zobrazení výsledkov sú užívateľovi odomknuté filtre, ktoré môže aplikovať na výsledky. Filtre sa nachádzajú v ľavej dolnej časti portálu a sú prezentované formou zaškrŕavacích políčk, s ktorými môže užívateľ manipulovať.

4.2.2 Inicializácia portálu

Pri vstupe užívateľa na portál, je portál inicializovaný. Nie je predvolená žiadna kategória, systém vyhľadáva publikácie vo všetkých kategóriách.

Na portáli sú zobrazené aktuálne pridané publikácie, systém ponúka užívateľovi zoznam kategórií, v ktorých môže užívateľ vyhľadávať a zoznam najpopulárnejších autorov publikácií. Popularita je meraná počtom publikácií, ktoré autor napísal.

4.2.3 Vyhľadávanie publikácií

Vyhľadávanie publikácie prebieha od momentu, kedy užívateľ zadáva vstup do formulárového poľa. Systém reaguje na udalosť pustenía klávesy (angl. keyup event) a asynchrónne kontaktuje server. Užívateľovi sa zobrazuje animácia načítavania dát a v momente, ako klient dostane odpoveď od serveru, sú užívateľovi zobrazené výsledky.

4.2.3.1 Vyhľadávanie podľa kľúčových slov

Systém chápe vstup od užívateľa ako zoznam kľúčových slov oddelených navzájom od seba čiarkami. Po odoslaní vstupu na server, systém spracováva vstup v týchto krokoch:

1. Zistí, či vstup nepatrí do kľúčových slov jazyka SQL, obrana proti SQL injection
2. Odstráni nadbytočné biele znaky
3. Rozdelí vstup na zoznam kľúčových slov
4. Kľúčové slová upraví na základný tvar

5. Odstráni duplicitné základné tvary kľúčových slov
6. Vytvorí SQL dotaz do databáze, ktorý vyhľadá všetky publikácie, ktoré obsahujú minimálne jedno z týchto kľúčových slov, uvažuje sa zvolená kategória
7. Systém vráti obmedzený počet publikácií užívateľovi ako výsledok

4.2.3.2 Vyhľadávanie podľa kategórií

Užívateľ môže v paneli s kategóriami zvoliť kategóriu, v ktorej chce vyhľadávať publikácie. Pri zvolení kategórie, systém asynchrónne kontaktuje server a znovu načíta posledné publikácie pridané do databáze portálu. Výsledkom sú publikácie, ktoré patria do užívateľom vybranej kategórie. Paralelne s tým, systém načítava aj najpopulárnejších autorov danej kategórie. Systém graficky užívateľa upozorňuje, akú kategóriu má práve aktívnu.

Pri vyhľadávaní publikácií počas aktivovanej kategórie, systém výsledky obmedzuje iba na danú kategóriu. Portál umožňuje vyhľadávať súčasne iba v jednej kategórii.

4.2.3.3 Automatické dopĺňovanie kľúčových slov

Užívateľ má možnosť aktivovať alebo deaktivovať automatické dopĺňovanie kľúčových slov. Táto možnosť je nastaviteľná v kontextovom menu pri vyhľadávacom formulári. Pri inicializácii systému je táto možnosť prednastavená ako aktívna.

Ak užívateľ zadáva vstup do vyhľadávacieho poľa a má aktivované automatické dopĺňovanie kľúčových slov, tak pri každom pustení klávesy sa asynchrónne prenáša aktuálny vstup na server a server vracia zoznam potenciálnych kľúčových slov, ktoré užívateľ plánoval zadať do vstupu.

Server vždy hľadá potenciálne kľúčové slova len pre jedno zadávané kľúčové slovo. Ak je na vstupe viacero kľúčových slov, ktoré sú oddelené čiarkami, serveru je odosielaný iba posledný prvok tohto zoznamu. Systém vracia len také kľúčové slová, ktoré vyhovujú aktívnej kategórii a v obmedzenom počte.

Po tom ako sú kľúčové slová navrhnuté užívateľovi, tak užívateľ má možnosť navrhované kľúčové slovo použiť a rozšíriť oň formulárový vstup, táto akcia nespúšťa opätovné volanie servera.

4.2.3.4 Odporúčanie inej kategórie

Užívateľ má možnosť aktivovať alebo deaktivovať ponúkание iných kategórií. Pri inicializácii systému je táto možnosť aktivovaná.

Ak je táto možnosť aktivovaná a užívateľ má zvolenú kategóriu, tak v prípade, že zadávané kľúčové slovo obsahujú publikácie, ktoré nepatria do aktívnej kategórie, je užívateľovi ponúknutá zmena kategórie.

Zoznam odporúčaných kategórií je zobrazený užívateľovi a ten má možnosť si jednu z kategórií zvoliť. Po zvolení kategórie sa znovu načítajú najpopulárnejší autori v danej kategórii ako aj najnovšie pridané publikácie do databázy týkajúce sa danej kategórie. Vstup vo formulári sa nemení a užívateľovi sú ponúknuté potenciálne kľúčové slová v závislosti na vstupe.

4.2.4 Zobrazenie výsledkov vyhľadávania

Po tom ako užívateľ potvrdí vyhľadávací formulár, tak na základe zadaných kľúčových slov (kap. [4.2.3.1](#)), systém vracia užívateľovi tabuľku obsahujúcu publikácie.

Informácie o publikáciách sú stručné a slúžia užívateľovi na zorientovanie sa v publikáciách. Systém zobrazuje názov publikácie, abstrakt a autorov publikácie. Užívateľ je schopný zobrazit' detail publikácie alebo detail autora publikácie. Takisto má možnosť výsledky filtrovať.

4.2.4.1 Filtrovanie výsledkov

Po tom, ako sú zobrazené výsledky vyhľadávania, je užívateľovi zobrazený panel s atribútmi publikácií.

Užívateľovi je poskytnutý zoznam všetkých autorov vyhľadaných publikácií a takisto zoznam rokov vydania publikácií. Užívateľ má možnosť dodatočne filtrovať výsledok odobraním autorov alebo obmedziť publikácie podľa dátumu vydania. Po upravení filtrov, užívateľ potvrdí aplikovanie filtrov na výsledok. Následne je pôvodný vyhľadávací dotaz upravený o tieto zmeny a poslaný znovu na server. Server vracia výsledky užívateľovi, ktoré sa dajú opätovne filtrovať.

4.2.4.2 Detail publikácie

Výsledná tabuľka s publikáciami obsahuje len základne informácie o každej publikácii. Užívateľ má možnosť po zvolení publikácie vidieť detail publikácie.

Detail publikácie obsahuje všetky základné informácie poskytnuté vo výslednej tabuľke ako aj rozširujúce informácie o publikácii. Užívateľovi sa zobrazuje počet videní publikácie, kategórie publikácie a kľúčové slova publikácie. Zobrazovaný je aj akčný panel, ktorý umožňuje užívateľovi prezerat' obsah publikácie v elektronickej podobe.

Z detailu publikácie je možné prejsť na detail autora zvolením autora zo zoznamu autorov publikácie.

4.2.4.3 Detail autora

Portál umožňuje prezerat' profily jednotlivých autorov publikácií. Detail autora je možné prezerat' zvolením autora vo výslednej tabuľke alebo zvolením autora v detaile publikácie.

Detail obsahuje fotografiu autora, vedecké oblasti, v ktorých autor publikuje, počet publikácií autora, ktoré sa nachádzajú v databáze portálu a zoznam týchto publikácií.

Zo zoznamu publikácie je možné jednu z publikácií zvoliť a prejsť na jej detail.

5 Implementácia

Portál je implementovaný na základe návrhu systému a samotná implementácia bola rozdelená na dva bloky, a to administrátorský a klientsky systém.

Na implementáciu systémov sú využité techniky objektovo orientovaného programovania. Každý zo systémov majú vlastný adresárový strom so zdrojovými kódmi a súčasťou implementácie sú aj moduly, ktoré používajú oba systémy. Zdrojové kódy týchto modulov sú uložené v špeciálnej adresárovej štruktúre. Štruktúru adresárov je možné vidieť v prílohe. (kap. [A](#))

5.1 Použité jazyky, prostredie a nástroje

Pri vytváraní portálu používame programovacie jazyky vhodné na vytváranie webových aplikácií. Portál je vytvorený formou dynamickej webovej stránky vo formáte HTML 5 za použitia multiplatformného skriptovacieho jazyka PHP 5. PHP skripty sú interpretované na strane serveru a užívateľovi je prenášaný až výsledok ich činnosti. Na vytváranie interaktívnych prvkov portálu, efektov a animácií používame jazyk JavaScript s knižnicou jQuery, ktorého interpret je obvykle spúšťaný na strane klienta (v prehliadači). Súčasťou implementácie je častá asynchrónna komunikácia s PHP serverom, na ktorú je využívaná technológia AJAX. Dizajn portálu je vytvorený pomocou jazyka CSS.

Na získavanie publikácií alebo vytváranie chybových logov aplikácia využíva jazyk XML. Na prácu s relačnými databázami využívame dotazovací jazyk SQL a databázový systém MySQL.

Portál bol implementovaný za použitia vývojového prostredia NetBeans IDE a nástroja na správu databáze phpMyAdmin.

5.2 Databáza

Portál využíva databázu vytvorenú na základe relačného modelu dát. Na správne fungovanie systému bolo potrebné vytvoriť databázu na ukladanie publikácií, ich kľúčových slov, autorov a iných atribútov o publikáciách. Na prácu s databázami používa triedu *DBAdapter*, ktorá poskytuje základné rozhranie pre prácu s databázou. Trieda využíva PHP knižnicu *mysqli*, ktorá poskytuje API na komunikáciu s databázovým systémom MySQL.

Najčastejšie používané metódy triedy *DBAdapter* sú:

1. *function __construct(\$host,\$user,\$pass,\$dbname)*
 - konštruktor triedy, parametrami metódy sú prístupové údaje do databázy
2. *public function Connect()*
 - metóda slúži na pripojenie k databáze
3. *public function ExecuteQuery(\$sql_query = 0)*
 - vykonáva SQL príkazy nad databázou
4. *public function GetResult()*

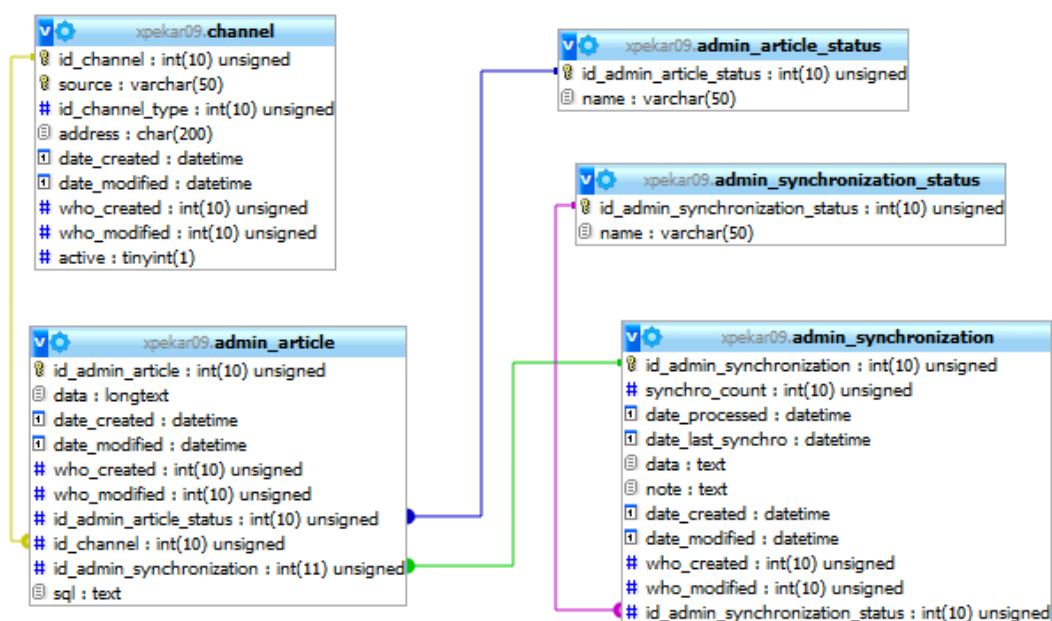
- vracia výsledok databázového dotazu

5.3 Implementácia administrátorského systému

Systém je implementovaný podľa návrhu (kap. 4.1).

5.3.1 Databázový model administrátorského systému

Nižšie je uvedený model databáze, ktorý využíva administrátorský systém.



Obr. 5.1 Databázový model administrátorského systému

Kardinality tabuliek:

- *admin_article* : *admin_article_status* – N:1
- *admin_article* : *admin_synchronization* – N:1
- *admin_synchronization* : *admin_synchronization_status* – N:1
- *admin_article* : *channel* – N:1

Tabuľky *admin_synchronization_status* a *admin_article_status* obsahujú záznamy o používaných stavoch publikácií a synchronizačných balíkov. Primárne kľúče tabuliek sú cudzími kľúčmi vo vymenovaných tabuľkách, keďže textová hodnota nie je vhodná hodnota cudzieho kľúču.

5.3.1.1 Tabuľka *admin_article*

V tabuľke *admin_article* sú uložené publikácie, ktoré sú ponúkané administrátorovi portálu na pridanie do portálu alebo odmietnutie. Každý článok sa nachádza v nejakom stave, ktorý popisuje atribút *id_admin_article_status*, ktorý je zároveň cudzím kľúčom. Cudzí kľúč *id_channel* popisuje, z akého kanálu článok pochádza a atribút *id_admin_synchronization* obsahuje informáciu o tom, v akom synchronizačnom balíku sa publikácia nachádza. Tento atribút je nepovinný a je vyplnený iba v prípade, že publikácia bola prijatá administrátorom (kap. 5.3.4.2).

Atribút *data* obsahuje HTML kód, ktorý popisuje, ako bude publikácia zobrazená administrátorovi portálu.

Atribúty *date_created*, *date_modified*, *who_created*, *who_modified* popisujú, ako a kým bol záznam tabuľky vytváraný alebo modifikovaný.

Atribút *sql* obsahuje SQL skript, ktorý vola procedúru, ktorá danú publikáciu uloží do databáze klientskeho systému, z ktorej sa zobrazujú články užívateľom portálu.

5.3.1.2 Tabuľka *admin_synchronization*

V tabuľke *admin_synchronization* sú uložené záznamy o synchronizačných balíkoch (kap. 4.1.4.1).

Každý balík sa nachádza v stave, ktorý popisuje atribút *id_admin_synchronization_status*. Atribút *data* je zatiaľ nevyužitý, no predpokladá sa, že môže mať vyžitie po implementácii rozšírení portálu (kap. 7.2).

Každý synchronizačný balík obsahuje komentár o tom, ako prebehla synchronizácia. Táto informácia je uložená v atribúte *note*. Hodnota „OK“ znamená, že obsah synchronizačného balíku bol synchronizovaný do databáze portálu. Iná hodnota popisuje chybu, ku ktorej došlo počas synchronizácie.

Synchronizácia balíku môže byť v prípade neúspechu vykonaná viackrát, počet pokusov vyjadruje atribút *synchro_count*. Čas posledného pokusu o synchronizáciu popisuje atribút *date_last_synchro*. Úspešná synchronizácia má vyplnený dátum synchronizácie v atribúte *date_processed*.

5.3.2 Grafické užívateľské rozhranie

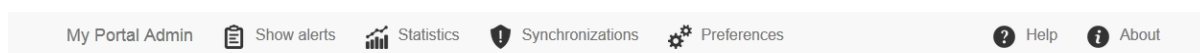
Grafické užívateľské rozhranie je implementované vzhľadom na návrh (kap. 4.1.1).

Dizajn administrátorského systému a interakcia systému s užívateľom boli implementované za použitia voľne dostupného framework-u Bootstrap [11].

Štýly administrátorského systému sa nachádzajú v adresári `Portal/admin_system/css`.

5.3.2.1 Navigačný panel

Pri vstupe do administrátorského systému je administrátorovi zobrazený navigačný panel administrátorského systému uvedený na nasledujúcom obrázku.



Obr. 5.2 Navigácia administrátorského systému

Administrátor má k dispozícii 6 modulov, ktoré môže využívať. Implementované moduly sú iba 3, a to:

- Show Alerts (kap. [5.3.3](#))
- Synchronizations (kap. [5.3.5](#))
- Preferences (kap. [4.1.2.3](#))

Moduly Statistics, Help a About môžu byť využité pri implementácii rozšírení portálu (kap. [7.2](#)). Pri implementácii navigačného panelu je využitý Bootstrap komponent *Navbar* [\[12\]](#).

5.3.2.2 Tabuľky

Administrátorovi sú v systéme ponúkané záznamy z databáze väčšinou formou tabuliek. Tabuľky sú zobrazované vo všetkých implementovaných moduloch.

Na implementáciu tabuliek sú využité štýly, ktoré ponúka rozhranie Bootstrap [\[13\]](#). V module Show Alerts je funkcionality a štýl tabuľky implementovaná použitím voľne dostupného JQuery pluginu *DataTables*. [\[14\]](#)

DataTables umožňuje záznamy v tabuľke zoradovať, vyhľadávať v tabuľke, obmedzovať počet zobrazovaných záznamov ako aj záznamy rozdeľovať do viacerých stránok tabuľky.

Ukážky tabuliek použitých v systéme vidíme na nasledujúcich obrázkoch.

Source	Type	Link/ E-mail account
Google Scholar	 Email	scholaralerts-noreply@google.com
Springer		http://www.springer.com/physics?SGWID=5-10100-820-0-0
IEEE		http://ieeexplore.ieee.org/rss/TOC4816218.XML
Science Direct		http://rss.sciencedirect.com/getMessage?registrationId=FIECGLMCGKELJIFEHIFFGJFJGOJFKPMENMHIMMGHE

Obr. 5.3 Tabuľka v štýle Bootstrap

Search:
Show entries
Showing 1 to 10 of 176 entries
< Previous
Next >

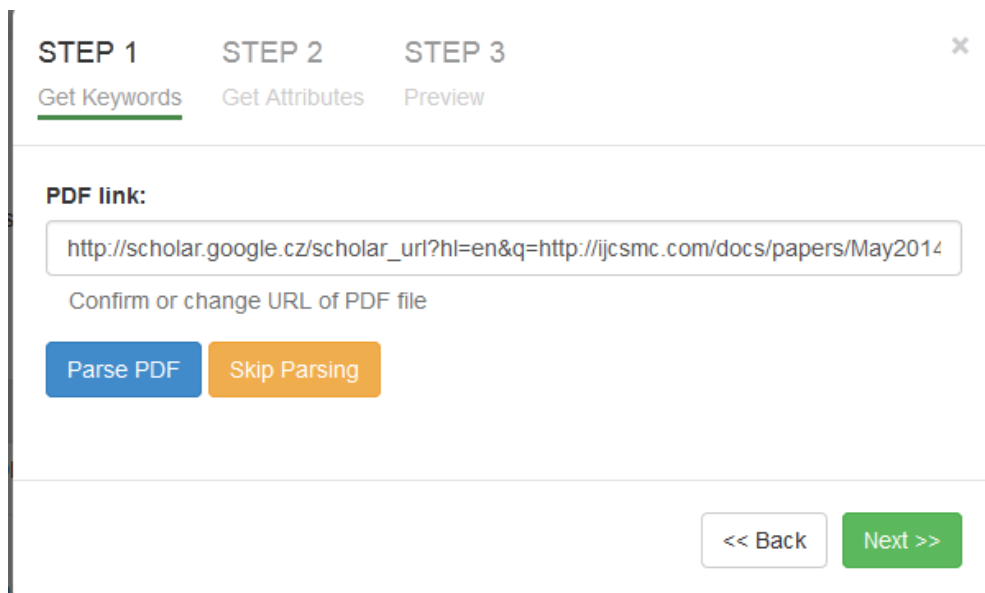
Alert	About
A MISBEHAVIOR DETECTION SCHEME ESTABLISHMENT IN DELAY TOLERANT NETWORKS LINK http://scholar.google.cz/scholar_url?hl=en&q=http://ijcsmc.com/docs/papers/May2014/V3I5201430.pdf&sa=X&scisig=AAGBfm1IDFg1gzuddyDg7iGa4coA7ICLR&oi=scholaralt DESCRIPTION Page 1. B.Sivakumar et al, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.5, May- 2014, pg. 76-82 ... International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology ... More/Less Details	Google Scholar
Decline Accept	

Obr. 5.4 Tabuľka v štýle DataTables

5.3.2.3 Dialógové okná

Systém komunikuje s administrátorom formou dialógov. Dialógy sú využívané vo všetkých implementovaných moduloch.

Dialógy sú užívateľovi zobrazované vždy na jeho podnet. Dizajn a funkcionálnosť dialógov je implementovaná využitím Javascript-ového komponentu *Modal*, ktoré poskytuje framework Bootstrap [15]. Ukážku dialógového okna môžeme vidieť nižšie.



The image shows a modal dialog box with a title bar containing three steps: 'STEP 1' (active), 'STEP 2', and 'STEP 3'. Below the title bar, the first step is labeled 'Get Keywords'. The main content area has a label 'PDF link:' followed by a text input field containing the URL 'http://scholar.google.cz/scholar_url?hl=en&q=http://ijcsmc.com/docs/papers/May2014'. Below the input field is the text 'Confirm or change URL of PDF file'. At the bottom of the main area are two buttons: 'Parse PDF' (blue) and 'Skip Parsing' (orange). At the bottom right of the dialog are two navigation buttons: '<< Back' and 'Next >>' (green).

Obr. 5.5 Dialógové okno zobrazované pri prijatí publikácie

5.3.2.4 Tlačidlá

Tlačidlá sú vstupným bodom na interakciu so systémom. Sú na ne naviazané rôzne udalosti, na ktoré systém reaguje. Tieto udalosti obsluhuje prehliadač a využíva interpret jazyka Javascript. Tlačidlá sú implementované pomocou rozhrania Bootstrap [16].

Ukážku tlačidiel môžeme vidieť na predchádzajúcom obrázku.

5.3.3 Získavanie publikácií

Získavanie a zobrazovanie získaných publikácií prebieha v module Show Alerts. V predchádzajúcej kapitole (kap. 4.1.2) je popísaný spôsob získavania publikácií a návrh, podľa ktorého je modul implementovaný.

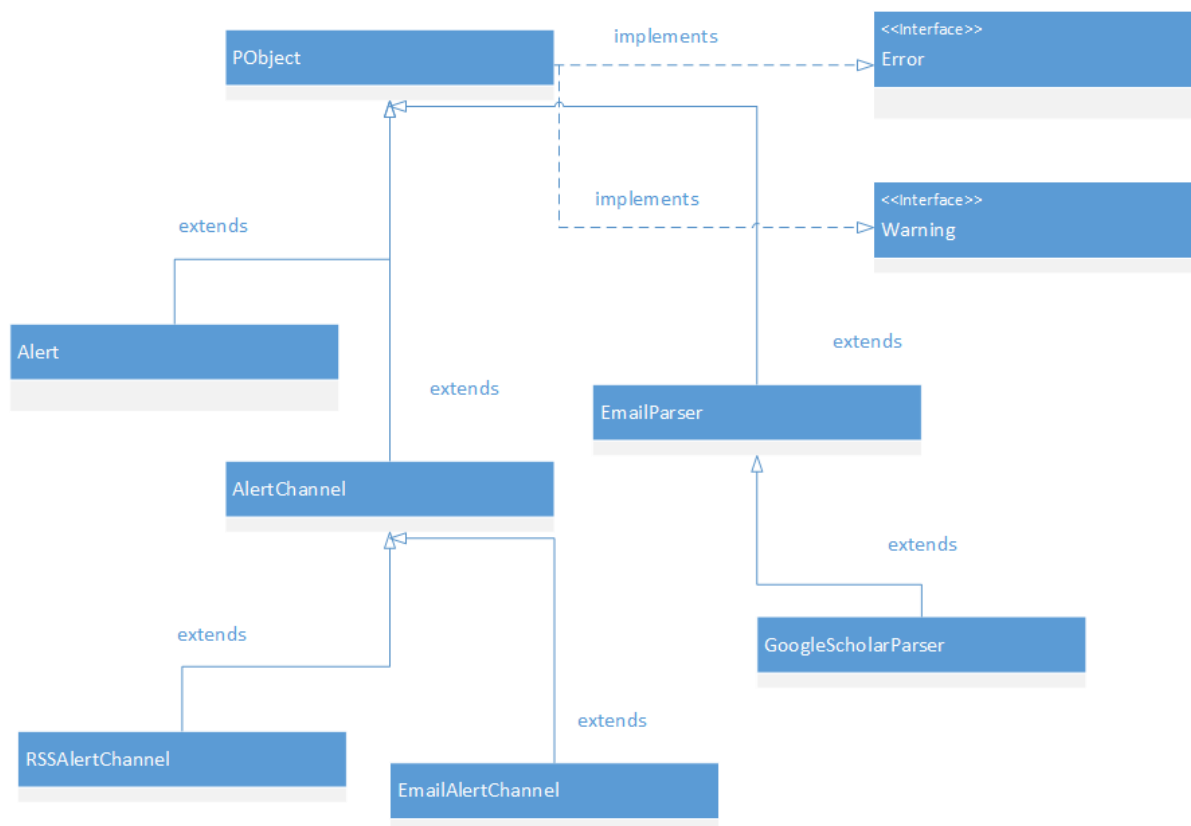
Systém získava publikácie z dvoch typov externých zdrojov.

- RSS kanál
- E-mail

Následne po ich získaní ich zobrazuje administrátorovi.

5.3.3.1 Objektový návrh

Navrhol som viacero tried, ktoré aplikácia využíva na získavanie publikácií a ich zobrazovanie. Nižšie môžeme vidieť diagram tried. Diagram popisuje iba názvy a vzťahy medzi triedami.



Obr. 5.6 Objektový návrh administrátorského systému

Trieda *PObject* zastrešuje všetky triedy používané v implementácii portálu. Implementuje chybové rozhranie *Error* a rozhranie *Warning*. Všetky inštancie objektov v portáli majú možnosť ukladať chybové logy na server alebo vypisovať varovania.

5.3.3.2 Čerpanie z externých zdrojov

V module Preferences je implementovaná podpora pre výber e-mailu, na ktorý sú posielané upozornenia a takisto na pridávanie kanálov, z ktorých systém upozornenia o nových publikáciách čerpá.

Abstraktná trieda *AlertChannel* reprezentuje kanál, cez ktorý získava systém publikácie. Ako bolo spomenuté vyššie (kap. 5.3.3), tak kanál môže byť dvojitého typu. Publikácie získavané z RSS kanálu spracováva trieda *RSSAlertChannel* a publikácie získané z e-mailového účtu spracováva trieda *EmailAlertChannel*.

Triedy *RSSAlertChannel* a *EmailAlertChannel* vznikli špecializáciou triedy *AlertChannel*. Trieda *AlertChannel* im poskytuje spoločné rozhranie, a to konkrétne metódy:

- *abstract protected function Commit(\$db)*

- metóda prevádza upozornenia o publikáciách do HTML podoby a ukladá ich do databázy, volá metódu *AddToDatabase* triedy *Alert* (kap. [5.3.3.3](#))
- *abstract protected function Parse()*
 - získava upozornenia z kanálov a vytvára upozornenia o publikáciách, inštancie triedy *Alert* (kap. [5.3.3.3](#))

V prípade získavania publikácií z e-mailového účtu nastáva problém, že externé zdroje, ktoré poskytujú upozorňovacie služby, používajú rôzne formáty e-mailu. Na spracovávanie e-mailov slúži trieda *EmailParser*, ktorá obsahuje abstraktnú metódu *Parse* na spracovanie obsahu elektronickej pošty.

Implementovaná je podpora služby Google Scholar, ktorá získava nové publikácie a posiela ich formou elektronickej pošty. Na spracovávanie pošty je v tomto prípade použitá trieda *GoogleScholarParser*.

5.3.3.3 Upozornenie o novej publikácii

Trieda *Alert* je abstrakciou jedného upozornenia o novej publikácii. V atribútoch triedy uchováva zdroj publikácie, základne a rozšírené informácie o publikácii. Najpoužívanéjšie metódy v poradí v akom sú volané:

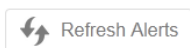
- *function __construct(\$source,\$input = NULL)*
 - konštruktor triedy, pri vytváraní dostáva inštancia triedy *dáta*, ktoré vydolovali triedy *RSSAlertChannel* a *EmailAlertChannel*
- *private function Parse()*
 - získané dáta upravuje trieda do vhodnej podoby
- *public function AddToDatabase (\$adapter,\$data)*
 - upozornenie o publikácii uloží do databázy, konkrétne do tabuľky *admin_article* (kap. [5.3.1.1](#)), záznam má nastavený stav na WAITING

5.3.3.4 Zobrazenie upozornení

Upozornenia o nových publikáciách sú ponúknuté administrátorovi na jeho vyžiadanie. Administrátor stlačí tlačidlo Refresh Alerts v module Show Alerts , po čom systém vykonáva v poradí tieto úkony:

1. Systém získa nové publikácie z kanálov a uloží ich do databázy.
2. Z tabuľky *admin_article* získa záznamy, ktoré sú v stave WAITING.
3. Získane záznamy sú zobrazené administrátorovi.

Na stlačenie na tlačidlo Refresh Alerts systém reaguje tak, že asynchrónne kontaktuje server, na ktorom sa spustí skript `Portal/admin_system/engine/new_alerts.php`. Zatiaľ čo klient čaká na odpoveď, užívateľovi sa zobrazuje animácia uvedená na obrázku nižšie.



Obr. 5.7 Animácia počas čakania na odpoveď od serveru

Každé upozornenie obsahuje abstrakt publikácie, názov publikácie a odkaz na zdroj publikácie. Z týchto informácií je administrátor schopný posúdiť obsah publikácie a na základe toho môže publikáciu prijať alebo odmietnuť. Vid' obrázok vyššie (obr. 5.4).

5.3.4 Operácie s publikáciami

5.3.4.1 Odmietnutie publikácie

Ak administrátor klikne na tlačidlo Decline, klient kontaktuje server a spustí sa skript `Portal/admin_system/engine/decline_alert.php`. Skript zmení status na danom upozornení na hodnotu DECLINED.

Systém rozpozná publikáciu podľa atribútu `id_record`, ktorý obsahuje riadok tabuľky s upozorneniami. Vid' následujúce HTML.

```
<tr class="odd" id_record="56" >...</tr>
```

5.3.4.2 Prijatie publikácie

Po kliknutí na tlačidlo Accept sa administrátorovi zobrazí dialóg s odkazom na elektronickú podobu publikácie. Odkaz je získavaný zo základných informácií, ktoré zobrazovalo upozornenie a je editovateľný.

Získavanie kľúčových slov

Na to, aby systém získal kľúčové slová je potrebné elektronickú podobu publikácie zanalyzovať. Na to je potrebné previesť PDF súbor na reťazec. Na túto činnosť program využíva voľne dostupnú PHP knižnicu `\SmalotPdfParser` od Sebastiena Malota [17].

Ak užívateľ iniciuje spracovanie PDF súboru, tak je kontaktovaný server, ktorý spúšťa skript `Portal/admin_system/engine/get_keywords.php` a vracia administrátorovi zoznam kľúčových slov.

Získanie kľúčových slov zastrešuje trieda `PDFParser`. Pri každom spracovávaní PDF súboru sa vytvorí inštancia tejto triedy a volajú sa jej metódy v tomto poradí:

1. `function __construct($file)`
 - o `$file` je odkaz na elektronickú podobu publikácie
 - o vytvorí sa inštancia triedy `\SmalotPdfParser`
2. `private function GetDestinationURL($url)`
 - o metóda, ktorá sa snaží získať efektívny odkaz na publikáciu vo formáte PDF

- URL, na ktorú pôvodný odkaz odkazuje je často presmerovaná na inú URL a pomocou tejto metódy získame priamy odkaz na elektronickú podobu publikácie, využíva sa PHP knižnica *curl* [18]
- 3. *public function Parse()*
 - získa metainformácie o PDF súbore a prevedie ho na reťazec
- 4. *private function ExtractKeywords()*
 - prevedie reťazec do formátu UTF-8
 - získa kľúčové slova publikácie pomocou 3 techník:
 - Z metainformácií PDF súboru pomocou metódy *ExtractMetaKeywords*
 - Z obsahu PDF pomocou metód *ExtractTagKeywords* a *ExtractTextKeywords*
- 5. *private function ExtractTagKeywords(\$input)*
 - v úvode publikácie hľadá slovo „keywords“ a identifikuje zoznam uvedených kľúčových slov oddelených čiarkou alebo bodkočiarkou
 - využíva regulárne výrazy a PHP funkciu *strip_tags* [19]
- 6. *private function ExtractTextKeywords(\$input)*
 - pomocou regulárnych výrazov odstráni z textu publikácie prevedeného do reťazca interpunkčné znamienka, symboly, čísla, rôzne značky a nahradí postupnosti bielych znakov iba jednou medzerou
 - rozdelí reťazec na slová využitím PHP funkcie *mb_split* [20]
 - slová prevedie na základný tvar použitím Porter Stemmer algoritmu, ktorý v PHP naprogramoval Richard Heyes [21] a je voľne dostupný, nachádza sa v súbore `Portal/admin_system/engine/classes/PorterStemmer.php`
 - zistí početnosť slov v základnom tvare a zoradí ich
 - slová v základnom tvare upraví naspať do pôvodného tvaru

Na získané kľúčové slová sa aplikuje filter, ktorý odstráni nežiaduce slová, ktorých zoznam sa nachádza v súbore `Portal/admin_system/other/stopwords.txt`.

Užívateľovi je ponúknutých 30 kľúčových slov, z toho 10 je zobrazených, ďalších 20 môže zobraziť kliknutím na tlačidlo More a taktiež môže vidieť, ktoré slová boli odstránené po aplikovaní filtra. Obrázok nižšie popisuje akou formou sú kľúčové slova zobrazené administrátorovi.

Metadata Keywords

Tag Keywords

node replication network set



Text Keywords

nodes



schemes



sensor



time



model



key



network



attacks



authentication



management



Add Keyword

Get More

Show Declined

Obr. 5.8 Navrhované kľúčové slová publikácie

Z ponúknutých kľúčových slov si administrátor môže vybrať, ktoré použije a takisto môže pridať vlastné kľúčové slová.

Kategorizácia publikácií

Kategorizovanie publikácií bolo implementované na základe návrhu (kap. [4.1.3.1](#)).

Skript `Portal/admin_system/engine/get_recommended_categories.php` získava kategórie z databáze a ponúka ich administrátorovi. Skript dostane od klienta zoznam kľúčových slov oddelených čiarkou, prevedie ich na základný tvar.

V databáze je vytvorený pohľad `vw_stem_categories`, ktorý každému slovu v základnom tvare priraduje zoznam kategórií, ktoré obsahujú publikácie, v ktorých bolo toto slovo použité. Databázový model klientskeho systému sa nachádza v jednej z ďalších kapitol (kap. [5.4.1](#)). Na nasledujúcom obrázku vidíme navrhované kategórie zobrazované administrátorovi.

Categories

matematika	☐	geologia	☒
adas	☒	dadad	☒
dfdfbiologia	☒		

Obr. 5.9 Kategorizovanie publikácie

Príprava na synchronizáciu

Ak publikáciu administrátor prijme, tak sa publikácia zaradí do synchronizačného balíku a elektronická podoba publikácie sa uloží na server do adresára `Portal/shared/pdf_articles`.

Na toto slúži metóda *SavePDF* triedy *PDFParser*, ktorá vytvorí súboru unikátny názov zložený z náhodných znakov a číslíc a súbor uloží.

Na pripravenie publikácie na synchronizáciu slúži databázová procedúra *update_admin_article*, ktorá je implementovaná formou transakcie a vykonáva úkony v tomto poradí:

1. Priradí publikácii, čiže záznamu v tabuľke *admin_article*, číslo synchronizačného balíku. Ak vhodný balík neexistuje, tak ho vytvorí
2. Do atribútu *sql* vloží skript, ktorý po spustení synchronizácie publikáciu vloží do lokálnej databáze klientskeho systému. (kap. [5.3.5](#))

5.3.5 Synchronizácia

Implementáciu synchronizácie podľa návrhu (kap. [4.1.4](#)) nájdeme v module *Synchronizations*. Ak administrátor zvolí tento modul, tak sa mu zobrazí tabuľka so synchronizačnými balíkmi. O túto funkcionality sa stará skript `Portal/admin_system/engine/get_synchronizations.php`.

Skript vypíše administrátorovi najnovších 100 synchronizačných balíkov, ktoré je možné vidieť na nasledujúcom obrázku.

Id	Date Created	Date Last Attempt	Date Processed	Attempts	Note	Status
28	2014-05-14 00:24:10	2014-05-14 00:24:25	2014-05-14 00:24:25	1	OK	PROCESSED
27	2014-05-03 18:42:02	2014-05-03 19:10:28	2014-05-03 19:10:28	2	OK	PROCESSED

Obr. 5.10 Synchronizačné balíky

Na jednotlivé synchronizačné balíky, ktoré sú v stave `READY` (kap. [4.1.4.1](#)), administrátor môže spustiť proces synchronizácie. Tento proces je implementovaný formou databázovej procedúry *synchronize_articles*, ktorá vykonáva úkony v tomto poradí:

1. Získa všetky publikácie, ktoré sú v danom synchronizačnom balíku

2. Pre jednotlivé publikácie dynamicky spúšťa SQL skripty uložené v atribúte *sql*
3. Po skončení synchronizácie, procedúra vyplní atribúty *date_last_synchro*, *date_processed*, *id_admin_synchronization_status*, *synchro_count* a *note*, ktoré sú vysvetlené v inej kapitole (kap. [5.3.1.2](#))
4. Nastaví synchronizovaným publikáciám status ACCEPTED

Ukladanie publikácií do databáze portálu

Počas synchronizácie sa spúšťajú SQL skripty pre jednotlivé záznamy v tabuľke *admin_article*. Skripty obsahujú databázovú procedúru *add_article*, ktorá ukladá publikácie do databázy klientskeho systému. Napĺňa tabuľky, opísané v databázovom modeli klientskeho systému v tomto poradí:

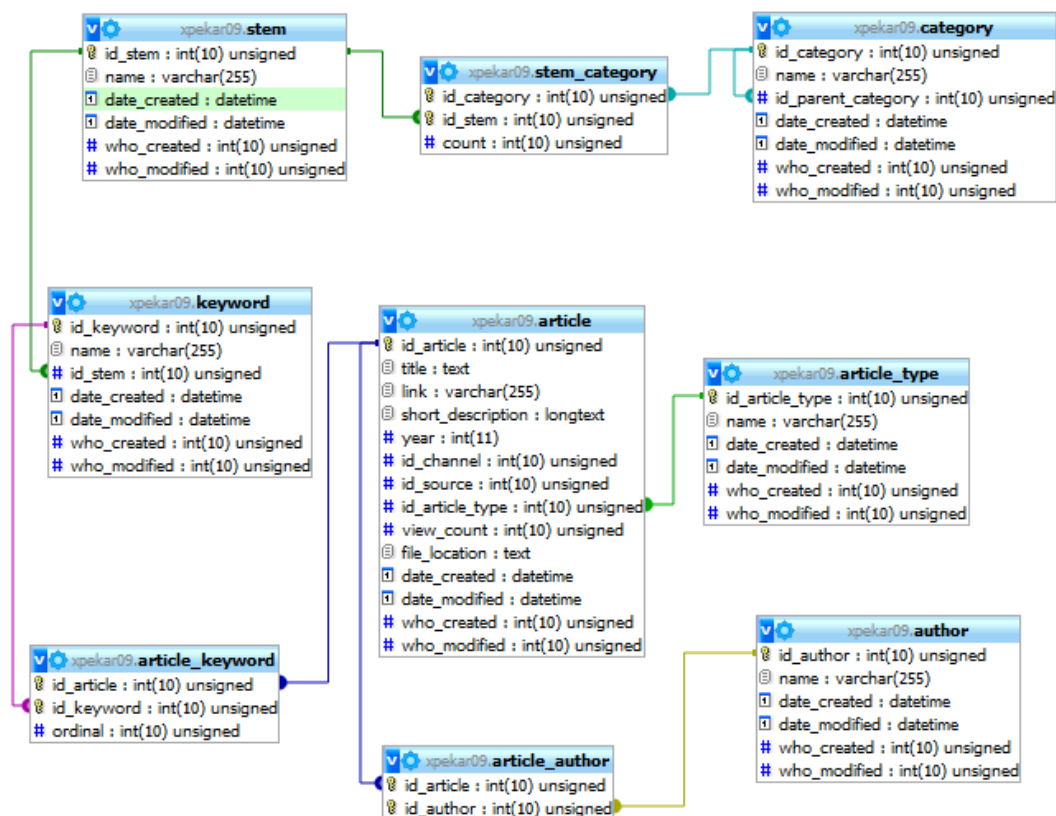
1. Do tabuľky *article* vloží záznam o publikácii so všetkými povinnými atribútmi
2. Do tabuľky *author* vloží záznamy o autoroch v prípade, že autori ešte v databáze neexistujú
3. Do tabuľky *article_author* vloží záznam o tom, kto publikáciu napísal
4. Do tabuľky *keyword* vloží kľúčové slovo, ak ešte v databáze neexistuje a do tabuľky *stem* vloží základný tvar kľúčového slova, ak tento záznam ešte neexistuje
5. Do tabuľky *article_keyword* vloží záznamy o tom, že dané kľúčové slová patria k danej publikácii
6. Do tabuľky *category* vloží kategórie, ak rovnaké záznamy neexistujú
7. Do tabuľky *stem_category* vloží záznam o tom, že dané základné tvary kľúčových slov patria k daným kategóriám

Procedúra beží ako transakcia.

5.4 Implementácia klientskeho systému

Systém je implementovaný podľa návrhu klientskeho systému (kap. [4.2](#)).

5.4.1 Databázový model klientskeho systému



Obr. 5.11 Databázový model klientskeho systému

Kardinality tabuliek:

- *article* : *keyword* – M : N (vznik tabuľky *article_keyword*)
- *article* : *author* – M : N (vznik tabuľky *article_author*)
- *article* : *article_type* – N : 1
- *stem* : *category* – M : N (vznik tabuľky *stem_category*)
- *stem* : *keyword* – 1 : N

Tabuľka *article_type* obsahuje záznamy o rôznych typoch publikácií. V implementácii táto tabuľka nie je využívaná a jej používanie je vhodné rozšírenie.

Tabuľky *article_keyword*, *stem_category* a *article_author* vyplynuli z konceptuálneho návrhu databázy. Jednotlivé kľúčové slová a autori publikácií sa môžu nachádzať vo viacerých publikáciách a naopak, jednotlivá publikácia môže obsahovať viacero kľúčových slov a môže byť publikovaná viacerými autormi.

Základný tvar kľúčového slova môže definovať viacero kategórií a naopak jedna kategória môže byť definovaná viacerými základnými tvarmi kľúčových slov.

Tabuľka *keyword* obsahuje záznamy kľúčových slov. Každé kľúčové slovo má svoj základný tvar, ktorý je umiestnený v tabuľke *stem*.

5.4.1.1 Tabuľka *article*

Tabuľka obsahuje informácie o publikáciách, ktorú sú zobrazované užívateľom portálu. Každá publikácia má svoj názov, odkaz na zdroj publikácie, abstrakt a rok vydania publikácie. Tieto informácie sú uložené v atribútoch *title*, *link*, *short_description* a *year*. Kľúčové slová publikácie sú uložené v tabuľke *article_keyword*, autory publikácie v tabuľke *article_author* a kategórie, do ktorých publikácia spadá v tabuľke *stem_category*.

Atribút *id_channel* hovorí o tom, z akého kanálu je publikácia získaná. Atribúty *id_source* a *id_article_type* sú v systéme nevyužívané a ich využívanie je vhodným rozšírením.

Atribút *view_count* určuje počet pozretí publikácie na portáli. Atribút *file_location* hovorí o fyzickom umiestnení súboru publikácie.

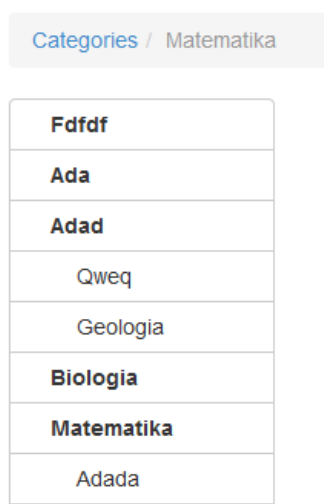
5.4.2 Grafické užívateľské rozhranie

Grafické užívateľské rozhranie je implementované podľa návrhu (kap. 4.2.1). Podobne ako GUI administrátorského systému, na dizajn a interakciu s užívateľom bol využitý framework Bootstrap. Štýly klientskeho systému sa nachádzajú v adresári `Portal/client_system/css`.

Na tabuľky, tlačidlá a dialógy boli použité rovnaké štýly a technológie ako pri administrátorskom systéme.

5.4.2.1 Panel s kategóriami

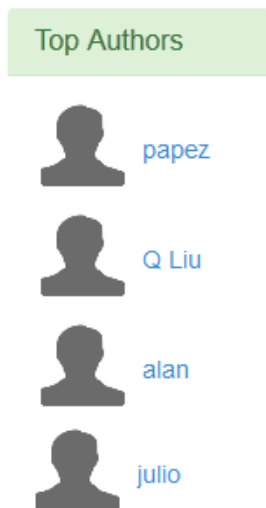
Kategórie publikácií sú zobrazované v ľavej časti portálu formou tlačidiel. Nad kategóriami sa nachádza aktuálne zvolená kategória vo forme striedky. Na implementáciu bol použitý komponent *Breadcrumbs*, ktorý poskytuje rozhranie Bootstrap [22]. Na obrázku nižšie môžeme vidieť, ako vyzerá panel s kategóriami v klientskom systéme.



Obr. 5.12 Panel s kategóriami

5.4.2.2 Panel s autormi

Autori sú zobrazení v pravej časti klientskeho systému. Panel je zobrazovaný pomocou komponenty *Panels*, ktorú poskytuje Bootstrap [23]. V paneli sa nachádza zoznam autorov. Každý autor má zobrazenú svoju fotku a svoje meno. Jednotlivé prvky zoznamu sú zobrazené pomocou komponenty *Media object* [24]. Ukážka panelu s autormi sa nachádza na nasledujúcom obrázku.

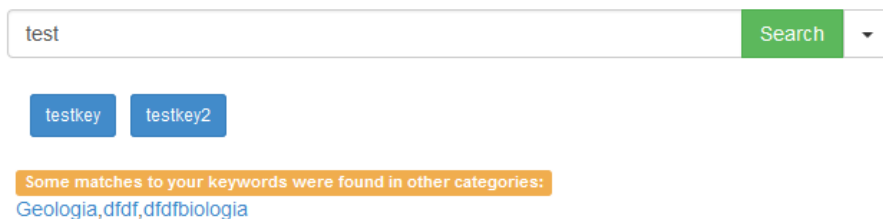


Obr. 5.13 Panel s autormi

5.4.2.3 Vyhľadávací formulár

Formulár na vyhľadávanie publikácií bol vytvorený za použitia Bootstrap komponentu *Input groups* [25]. Pod formulárom sa nachádza zoznam ponúkaných kľúčových slov, ktoré sú zobrazované formou tlačidiel a odporúčané kategórie vo forme odkazov, pri ktorých je využívaný komponent *Labels* [26].

Ukážka formuláru je na nasledujúcom obrázku.



Obr. 5.14 Vyhľadávací formulár

5.4.2.4 Panel s filtrami

Filtre na filtrovanie výsledkov sa nachádzajú v ľavej dolnej časti portálu. Pri implementácii boli využité komponenty *Labels* a *Input groups*, ktoré poskytuje rozhranie Bootstrap.

Ukážka panelu s filtrami sa nachádza na nasledujúcom obrázku.

The image shows a web interface for filtering data. It consists of three main sections, each with a yellow header and a list of items with checkboxes:

- Authors:** A yellow header box labeled "Authors" is followed by a list of four items: "L Russo", "S Rosa", "B Bona", and "M Matteucci". Each item has a small checkbox to its left, all of which are checked.
- Sources:** A yellow header box labeled "Sources" is followed by a list of one item: "teest". It has a checked checkbox to its left.
- Years:** A yellow header box labeled "Years" is followed by a list of one item: "2013". It has a checked checkbox to its left.

Below these sections is a horizontal line, and at the bottom is an orange button labeled "Apply Filters".

Obr. 5.15 Panel s filtrami

5.4.3 Zobrazovanie informácií z databázy

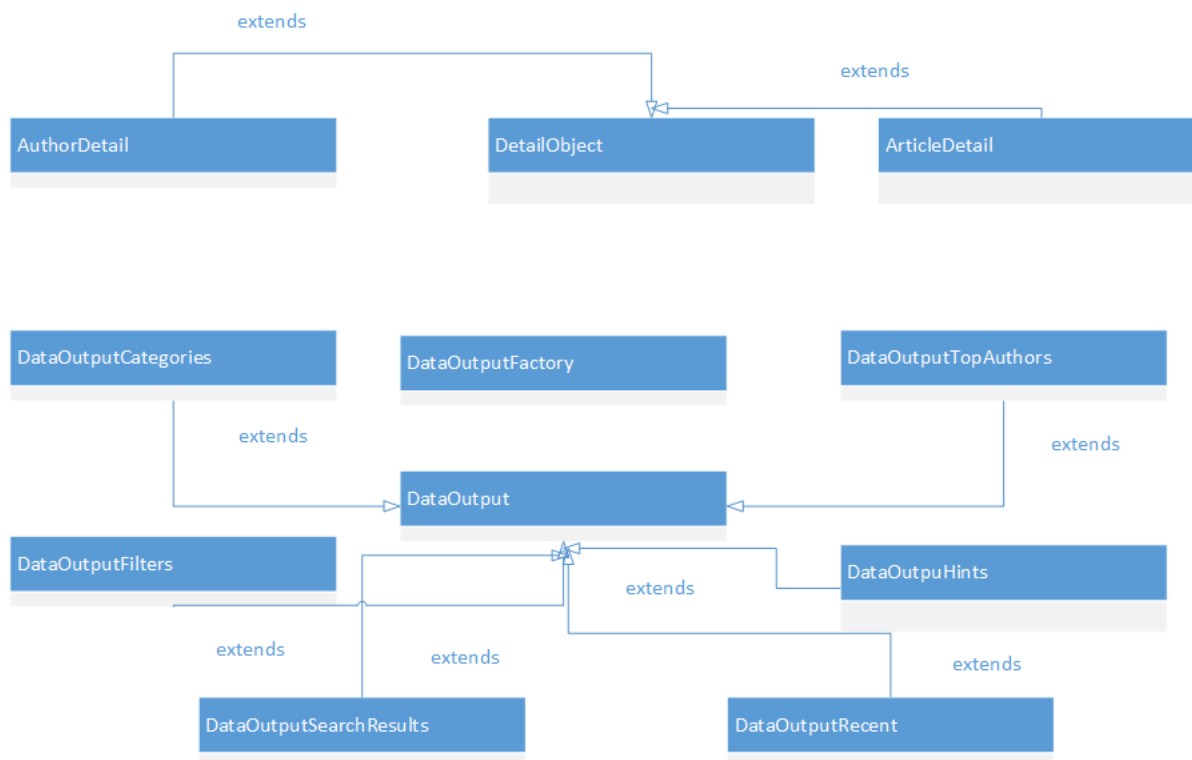
Klientsky systém zobrazuje užívateľom publikácie z lokálnej databázy. Pri vstupe na portál sú užívateľovi zobrazené aktuálne pridané publikácie a najpopulárnejší autori publikácií. O túto funkcionálnosť sa stará trieda *ClientPortal*.

Pri vstupe na portál je vytvorená inštancia tejto triedy a sú volané jej metódy v poradí:

1. *public function Init()*
 - portál sa napojí na lokálnu databázu a získa najnovšie publikácie, najpopulárnejších autorov a kategórie publikácií
2. *private function MakeHTML()*
 - vytvorí zo získaných dát HTML kód a zobrazí ho užívateľovi portálu

5.4.3.1 Objektový návrh

Na zobrazovanie informácií z databázy využíva systém viacero tried. Diagram tried je uvedený nižšie.



Obr. 5.16 Objektový návrh klientskeho systému

Trieda *DataOutput* a triedy vzniknuté jej špecializáciou sa používajú na zobrazovanie základných informácií o záznamoch v tabuľkách databázy. Trieda *DetailObject* a triedy vzniknuté jej špecializáciou sa používajú na zobrazovanie detailov o jednotlivých záznamoch.

Trieda *DataOutputFactory* implementuje návrhový vzor výrobná [27] a vytvára triedy na zobrazovanie základných informácií záznamov podľa zvoleného typu.

5.4.3.2 Databázové pohľady

Na zobrazovanie informácií z databázy, ako sú napríklad publikácie, najpopulárnejší autori a podobne, je potrebné spájať viacero tabuliek na základe určitých podmienok a vykonávať určité agregované funkcie nad záznamami tabuliek.

Z tohto dôvodu boli vytvorené databázové pohľady, ktorý túto réžiu zapuzdrujú a portál čerpá dáta iba z týchto pohľadov. Implementované sú tieto pohľady:

- *vw_articles* – informácie o publikáciách
- *vw_article_authors* – každej publikácii priradení jej všetci autori
- *vw_article_categories* – zoznam kategórií priradený k publikáciám
- *vw_article_keywords* – kľúčové slová publikácií
- *vw_authors* – informácie o autoroch
- *vw_author_articles* – každému autorovi priradené všetky publikácie, ktoré publikoval
- *vw_author_categories* – oblasti, v ktorých autor publikuje

- *vw_categories* – informácie o kategóriách
- *vw_stem_categories* – každému základnému tvaru kľúčového slova priradená kategória a počet výskytov tohto slova v danej kategórii

5.4.3.3 Zobrazovanie informácií v prehľade

Základné informácie o záznamoch uložených v databáze sú na portáli zobrazované v prehľadoch, a to formou tabuliek alebo zoznamov. Takýmito informáciami sú:

- kategórie publikácií
- najpopulárnejší autori
- ponúkané kategórie pri vyhľadávaní
- aktuálne pridané články
- filtre
- výsledky vyhľadávania

Na získanie týchto informácií sa využívajú triedy odvodené od triedy *DataOutput*. Tieto triedy implementujú abstraktnú metódu:

- *abstract public function ToHTML()*
 - metóda zodpovedná za HTML štruktúru získaných dát

Dáta sú získavané metódou *Query*, ktorej vstupný parameter je dotaz v jazyku SQL. Dotaz je vytváraný podľa informácií, ktoré potrebuje systém získať a sú získavané z databázových pohľadov (kap. [5.4.3.2](#)).

5.4.3.4 Zobrazovanie detailov informácií

Na zobrazovanie detailných informácií sa využívajú triedy odvodené od triedy *DetailObject*. Portál zobrazuje detailné informácie o autoroch a publikáciách podľa návrhu. (kap. [4.2.4.2](#) a kap. [4.2.4.3](#))

Detail je zobrazovaný vo forme dialógového okna a obsah detailu je zobrazený do dvoch. Vľavo sú základné informácie, vpravo detailnejšie informácie.

Každá z odvodených tried implementuje metódy:

- *abstract public function HTMLDetail()*
 - metóda zodpovedná za HTML štruktúru detailu
- *abstract protected function HTMLLeftFields()*
 - HTML štruktúra základných informácií o zázname
- *abstract protected function HTMLRightFields()*
 - HTML štruktúra detailnejších informácií o zázname

Na získanie dát je použitá metóda *Query*, ktorá zodpovedá za získanie dát z databáze využitím databázových pohľadov.

5.4.4 Vyhľadávanie publikácií

Vyhľadávanie publikácií je implementované podľa návrhu (kap. [4.2.3](#)).

5.4.4.1 Dopĺňovanie vstupu

Po zadávaní vstupu do formulárového poľa je kontaktovaný server, ktorý vykonáva skript `Portal/client_system/engine/get_hints.php`.

Skript získava z databáze kľúčové slová, ktoré navrhuje užívateľovi a odporúča iné kategórie užívateľovi. (kap. [4.2.3.4](#)) Dáta sú získavané z databázového pohľadu `vw_keyword_hints`.

Interakciu s užívateľom a kontaktovanie serveru má na starosti skript `Portal/client_system/js/get_hints.js`.

5.4.4.2 Získanie výsledkov vyhľadávania

Po odoslaní formuláru na vyhľadávanie, klient kontaktuje server, ktorý vykoná skript `Portal/client_system/engine/search_query.php`.

Skript obdrží zoznam kľúčových slov v základnom tvare, kontaktuje databázu a vráti záznamy tak, ako je to definované v návrhu (kap. [4.2.3.1](#)). Systém vracia publikácie na základe zvolenej kategórie. (kap. [4.2.3.2](#))

Po obdržaní výsledkov sú užívateľovi zobrazené filtre (kap. [4.2.4.1](#)), ktoré umožňujú filtrovať na základe týchto atribútov:

- Years – rok vydania publikácie
- Authors – autori publikácie
- Sources – zdroj publikácie, implementácia tento atribút nevyužíva a je to vhodné rozšírenie

Filtre, zvolená kategória a zoznam publikácií sú odosielané serveru HTTP metódou POST, ktoré spracováva skript `Portal/client_system/engine/search_query.php`.

O kontaktovanie serveru, interakciu s užívateľom a zobrazenie výsledku sa starajú skripty `Portal/client_system/js/search_query.js` a `Portal/client_system/js/search_query_results.js`.

O každej zobrazenej publikácii vo výsledku je možné zobrazit' detail. Detail je implementovaný podľa návrhu a spôsob získania detailu je popísaný v predchádzajúcej kapitole (kap. [4.2.4.2](#)).

6 Testovanie

Aplikácia bola testovaná na lokálnom webovom serveri (localhost) na mojom osobnom počítači, na ktorom boli nainštalované všetky potrebné aplikácie a rozšírenia popísané v manuáli.

Testoval som nasledujúce scenáre:

- Administrátorský systém
 - Získanie upozornení o publikáciách v prípade, že neboli zadané žiadne zdroje
 - Získanie upozornení o publikáciách v prípadoch, že server nepodporoval používanie protokolu IMAP alebo zadaná e-mailová schránka neexistuje
 - Administrátor nevyplní všetky povinné polia pri ukladaní publikácie
 - Systém správne reaguje na odmietnutie publikácie a na prijatie publikácie
 - Systém nastavuje synchronizačný balík do stavu PROCESSED po úspešnej synchronizácii
 - Systém umožňuje v prípade neúspešnej synchronizácie znovu synchronizovať balík
- Klientsky systém
 - Zobrazenie všetkých kategórií, ktoré existujú v databáze
 - Systém správne reaguje na pokus o SQL injection vo vyhľadávacom formulári
 - Systém vracia správne výsledky po vyhľadávaní
 - Detail publikácie obsahuje správne dáta

Počas testovania som odhalil viacero chýb, medzi najzávažnejšie patrili:

- Spadnutie systému v prípade, ak server nepodporoval používanie protokolu IMAP
- Nedostatočné ošetrovanie formulárov

Implementovaný portál je robustný a vyžaduje neporovnateľne viac testov. Návrh na rozšírenie testov sa nachádza v ďalšej kapitole (kap. [7.2](#)).

7 Záver

7.1 Zhodnotenie práce

Cieľom tejto práce bolo implementovať vedecký webový portál s automatickým generovaním obsahu. Bolo potrebné sa zoznámiť s existujúcimi webovými portálmi a zanalyzovať možnosť použitia upozornovacích služieb na vytváranie obsahu portálu.

Bol implementovaný administrátorský systém, ktorý umožňuje definovať zdroje, z ktorých portál získava publikácie. Systém dokáže pracovať s publikáciami, ktoré sú získané pomocou služby Google Scholar a s publikáciami získanými z ľubovoľných RSS kanálov. Získané publikácie systém spracováva a ponúka ich administrátorovi.

Administrátorský systém dokáže získavať o publikáciách základné informácie, dokáže ich kategorizovať a navrhovať kľúčové slová, ktoré publikácie charakterizujú. Je implementovaná podpora na synchronizáciu publikácií do databáze portálu.

Taktiež bol implementovaný klientsky systém, ktorý obsah portálu zobrazuje návštevníkom portálu. Užívateľ môže publikácie vyhľadávať a filtrovať podľa svojich potrieb. Bola implementovaná podpora na automatické dopĺňovanie zadávaných kľúčových slov pri vyhľadávaní. Systém dokáže na základe zadaných kľúčových slov odporučiť užívateľovi publikácie z iných kategórií.

Po vykonaní testov funkčnosti systémov sa ukázalo, že portál spĺňa požiadavky v návrhu.

7.2 Možné rozšírenia

Portál ako taký by mal ponúkať rozsiahlejší balík služieb a mal by byť ľahko udržiavaný administrátormi portálu. Tieto podmienky aktuálny stav pokrýva iba čiastočne.

Počas implementácie portálu ma napadalo viacero rozšírení, ktoré by mohli byť implementované a zaintegrované do systému v budúcnosti.

7.2.1 Kvalitnejší PDF parser

V súčasnosti využívaný konvertor PDF súborov na reťazec a nástroj na získavanie metainformácií PDF nie sú schopné spracovať každý súbor. Využívaná je voľne dostupná knižnica od Sebastiena Malota, ktorá je stále vo vývoji. Bolo by vhodné sa poobzerať po inej dokonalejšej knižnici alebo kombinovať viacero knižníc, aby sa zminimalizoval počet nespracovaných PDF súborov.

7.2.2 Efektívnejšie získavanie kľúčových slov

Systém nie je schopný získavať všetky kľúčové slová vymenované v úvode publikácie, keďže každý autor publikácie poskytuje tieto kľúčové slová v inom formáte, na inom mieste publikácie a takisto autori používajú iné oddeľovače kľúčových slov. Bolo by vhodné zanalyzovať množstvo publikácií a formy zápisu kľúčových slov a vylepšiť implementáciu získavania kľúčových slov.

7.2.3 Podpora viacerých zdrojov

Systém dokáže spracovávať iba e-maily zasielané službou Google Scholar. Keďže väčšina poskytovateľov upozorňovacích služieb posiela upozornenia práve formou elektronickej pošty, je vhodné podporovať spracovávanie e-mailov získaných aj z iných externých zdrojov.

7.2.4 Automatizované testy

Implementácia jednotkových testov (angl. unit test) na všetky triedy a metódy, ktoré portál používa v súčasnej implementácii chýba. Takisto by bolo vhodné implementovať automatizované GUI testy, keďže portál disponuje množstvom interakcií s užívateľom.

7.2.5 Spracovanie chýb

V súčasnom stave sa portál pri výskyte neočakávaných chýb nezachová korektne. Nemalo by sa stať, že užívateľovi je vypísané systémové chybové hlásenie. Chybové hlásenia by bolo vhodné uchovávať na serveri vo forme logov.

Všetky chyby a varovania by mali byť prezentované užívateľovi jednotnou formou.

7.2.6 Vytváranie užívateľských účtov

Súčasťou portálov je možnosť vytvárania účtov užívateľom. Užívateľ by mal mať možnosť si prispôbovať zobrazený obsah na portáli a systém si tieto nastavenia zapamätať. Taktiež niektorý obsah by mohol byť ponúkaný len registrovaným užívateľom.

7.2.7 Implementovanie modulov Statistics, Help a About

V administrátorskom systéme chýba modul, ktorý popisuje, ako používať administrátorský portál a modul pre podporu administrátora v prípade, že nastane nejaký problém počas obsluhy administrátorského systému.

Ďalej by bolo vhodné implementovať modul, ktorý by poskytoval štatistické dáta o tom, koľko publikácií bolo prijatých alebo odmietnutých administrátorov, z akých boli zdrojov, ktorý kanál je najviac využívaný a podobne. Tieto štatistiky by mohli byť prezentované formou grafov a tabuliek.

Literatúra

- [1] *Yahoo!* [online]. 1994 [cit. 2014-05-20].
<https://www.yahoo.com>
- [2] *Automobil Revue* - úvodní stránka: Nejrozsáhlejší motoristický portál. *Automobil Revue* [online]. 2011 [cit. 2014-05-20].
<http://www.automobilrevue.cz/>
- [3] What is a Web Portal?: Portals, and what they are explained... WINKLER, Ramona. PRODUCT DESIGN CENTER, SAP AG. *Atlantic WEB FITTERS* [online]. 2012 [cit. 2014-05-20].
<http://www.atlanticwebfitters.ca/AboutCMS/WhatisaWebPortal/tabid/95/Default.aspx>
- [4] Definition portal. *Business Dictionary* [online]. [cit. 2014-05-20].
<http://www.businessdictionary.com/definition/portal.html>
- [5] Alert Services: A How-To Guide. LINDSTROM, Joyce a Tim Dolighan DOLIGHAN. *Iowa State University* [online]. 2004 [cit. 2014-05-20].
http://instr.iastate.libguides.com/content.php?pid=100826&search_terms=alerts
- [6] Alerts and Related Services: Guide To Guides. POWOROZNEK, Emily. *University of New Hampshire: Library* [online]. 2013 [cit. 2014-05-20].
<http://libraryguides.unh.edu/content.php?pid=244214>
- [7] Keeping Current with Research: Database Search Alerts. MONKS, Kathleen Monks a Erica Coe COE. *UW TACOMA LIBRARY. Asa H Gordon Library* [online]. 2013 [cit. 2014-05-20].
<http://savannahstate.libguides.com/content.php?pid=345964&sid=2830069>
- [8] Databases and electronic resources: Alerting services. *The University of Sydney: LIBRARY* [online]. 2002 [cit. 2014-05-20].
<http://www.library.usyd.edu.au/databases/alerts.html>
- [9] Alert services. KIRSI, Lepistö. *Tampere University Of Technology* [online]. 2013 [cit. 2014-05-20].
<http://www.tut.fi/en/library/research-support-services/ideas-and-research/alert-services/index.htm>
- [10] *Google Scholar* [online]. 2011 [cit. 2014-05-20].
<http://scholar.google.cz/>
- [11] *Bootstrap: front-end framework* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com>

- [12] Components: Navbar. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/components/#navbar>
- [13] CSS: Tables. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/css/#tables>
- [14] *DataTables: Table plug-in for jQuery* [online]. 2007 [cit. 2014-05-20].
<https://datatables.net/>
- [15] Components: Modals. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/javascript/#modals>
- [16] CSS: Buttons. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/css/#buttons>
- [17] MALOT, Sebastien. *PDF Parser: PHP library to parse PDF files and extract elements like text* [online]. 2014 [cit. 2014-05-20].
<http://www.pdfparser.org>
- [18] Client URL Library. *PHP.net* [online]. 2000 [cit. 2014-05-20].
<http://www.php.net/manual/en/book.curl.php>
- [19] Stripos. *PHP.net* [online]. 2004 [cit. 2014-05-20].
<http://www.php.net/manual/en/function.stripes.php>
- [20] Mb_split. *PHP.net* [online]. 2000 [cit. 2014-05-20].
<http://www.php.net/manual/en/function.mb-split.php>
- [21] PorterStemmer. HEYES, Richard. [HTTP://WWW.PHPGURU.ORG/](http://WWW.PHPGURU.ORG/). *Tartarus* [online]. 2005 [cit. 2014-05-20].
<http://tartarus.org/martin/PorterStemmer/php.txt>
- [22] Components: breadcrumbs. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/components/#breadcrumbs>
- [23] Components: panels. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/components/#panels>
- [24] Components: media. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/components/#media>
- [25] Components: input-groups. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/components/#input-groups>
- [26] Components: labels. *Bootstrap* [online]. 2010 [cit. 2014-05-20].
<http://getbootstrap.com/components/#labels>
- [27] GUTMANS, Andi, Stig Saether BAKKEN a Derick RETHANS. *Mistrovství v PHP 5*. 1. vyd. Brno: CP Books, a.s., 2005, s. 3. ISBN 80-251-0799-X.

A **Obsah CD**

Priložený CD nosič má nasledujúcu adresárovú štruktúru:

- `/bp_documentation` – dokumentácia systémov
- `/bp_source_codes/Portal` – zdrojové kódy aplikácie
- `/bp_source_codes/Portal/admin_system` – zdrojové kódy administrátorského systému
- `/bp_source_codes/Portal/client_system` – zdrojové kódy klientskeho systému
- `/bp_source_codes/Portal/shared` – zdrojové kódy využívané oboma systémami
- `/bp_source_codes/Portal/dependancies` – konfigurácia utility *composer*, ktorá sťahuje najnovšie verzie voľne dostupných knižníc využívaných v portáli
- `/bp_TEC` – zložka s technickou prácou

B Manuál

B.1 Príprava databázy

Na fungovanie portálu je nutné vytvoriť databázu portálu. Na priloženom CD sa v adresári `Portal/shared/sql` nachádzajú SQL skripty, ktoré je nutné spustiť na databázovom serveri v tomto poradí:

1. `createDB.sql`
2. `createTables.sql`
3. `createViews.sql`
4. `insertInit.sql`
5. `createFunctions.sql`
6. `createProcedures.sql`

Skripty vytvoria tabuľky, pohľady, funkcie, procedúry používané v portáli a niektoré tabuľky naplnia potrebnými dátami.

B.2 Príprava webového serveru

Na webový server je potrebné skopírovať všetky zdrojové súbory webového portálu, čiže celý adresár `Portal`. Na serveri má byť nainštalovaný server PHP 5 s rozšíreniami:

- *php IMAP*
- *composer*
- *mcrypt_create_iv*

Pomocou programu *composer* server získava aktuálne verzie knižníc a rozšírení.

Zatiaľ čo rozšírenie *php IMAP* je nevyhnutnou súčasťou správneho fungovania aplikácie, ostatné rozšírenia sú doporučené.

V súbore `Portal/config.php` je potrebné nastaviť údaje na prihlásenie sa do databázy a cestu ku koreňovému adresáři na webovom serveri.

V súbore `Portal/shared/engine/classes/InitConfig.php` je nutné nastaviť v metóde *ReadConfig* cestu ku konfiguračnému súboru.

V nasledujúcich súboroch je nutné nastaviť cestu k súboru `Portal/general.inc`, kde sa nachádza funkcia *my_autoloader(\$file_name)*, ktorá automaticky načítava závislosti medzi jednotlivými súbormi.

- `Portal\admin_system\engine\channels_edit.php`
- `Portal\admin_system\engine\classes\PDFParser.php`
- `Portal\admin_system\engine\decline_alert.php`

- Portal\admin_system\engine\edit_mail.php
- Portal\admin_system\engine\get_channels.php
- Portal\admin_system\engine\get_email.php
- Portal\admin_system\engine\get_keywords.php
- Portal\admin_system\engine\get_recommended_categories.php
- Portal\admin_system\engine\get_synchro_detail.php
- Portal\admin_system\engine\new_alerts.php
- Portal\admin_system\engine\save_alert.php
- Portal\admin_system\engine\synchro_to_db.php
- Portal\admin_system\tests\AlertSystemTest.php
- Portal\client_system\engine\get_article_detail.php
- Portal\client_system\engine\get_author_detail.php
- Portal\client_system\engine\get_filters.php
- Portal\client_system\engine\get_hints.php
- Portal\client_system\engine\get_recent.php
- Portal\client_system\engine\load_client.php
- Portal\client_system\engine\load_top_authors.php
- Portal\client_system\engine\search_query.php
- Portal\shared\engine\classes\PObject.php
- Portal\shared\tests\PObjectTest.php

V súbore Portal\general.inc je nutné nastaviť v konštante PORTAL_DIR cestu ku koreňovému adresáru.

Administrátorský systém spustíme interpretovaním skriptu Portal\admin_system\index.php a klientsky systém spustíme interpretovaním skriptu Portal\client_system\index.php.

B.3 Demonštračné dáta

Ukážku portálu je možné vidieť na webovom serveri FIT VUT. Na URL adrese http://www.stud.fit.vutbr.cz/~xpekar09/Portal/admin_system sa nachádza administrátorský systém a na URL adrese http://www.stud.fit.vutbr.cz/~xpekar09/Portal/client_system sa nachádza klientsky systém.