

Katalóg Požiadaviek

Automatické Vyhľadávanie Ohlasov

Norbert Jurík
Tomáš Maťko
Adam Valentovič
Andrej Zemko

TIS 2019/20

Obsah

1	Úvod	3
1.1	Cieľ Katalógu Požiadaviek	3
1.2	Rozsah Systému	3
1.3	Slovník Pojmov	3
1.4	Referencie	3
1.5	Prehľad Nasledujúcich Kapitol	4
2	Všeobecný Popis	4
2.1	Perspektíva Systému	4
2.2	Funkcie Systému	4
2.3	Charakteristika Používateľa	4
2.4	Všeobecné Obmedzenia	4
2.5	Predpoklady Závislostí	5
3	Funkčné Požiadavky	5
3.1	Databázy	5
3.1.1	Využívanie Databáz	5
3.1.2	Nedostupné Databázy	5
3.1.3	Informácia o Nedostupných Databázach	5
3.2	Autorizačný Kľúč	5
3.3	Vstupné Prvky Úvodného Formulára	5
3.3.1	v Prípade, Nevyplneného DOI	5
3.3.1.1	Viaceré Roky/Rozsahy rokov Naraz	6
3.3.1.2	Viaceré Afiliácie Naraz	6
3.3.2	v Prípade Vyplneného DOI	6
3.3.2.1	Viaceré DOI Naraz	6
3.4	Zoznam Prác	6

3.5	Vyhľadávanie	6
3.5.1	Vyhľadávanie pri Vyplnenom DOI	6
3.5.2	Vyhľadávanie pri Nevypĺnenom DOI	6
3.5.2.1	pri Vyplnenej Afiliácii	7
3.5.2.2	pri Zadanom Roku/Rozsahu Rokov	7
3.6	Zoradenie	7
3.7	Informácia o Databáze v Ktorej Bola Položka Nájdená	7
3.8	Zaškrtávacie Políčka	7
3.8.1	Odškrtnutie	7
3.9	Tlačidlo „Vyhľadaj Citácie“	7
3.10	Zoznamy Citácií na Prácu/Práce	7
3.10.1	Postupné Zobrazovanie Zoznamov Citácií	7
3.10.2	Zobrazenie Zoznamov Citácií na Novej Karte	8
3.10.3	Vylúčenie Autocitácií	8
3.10.3.1	Možnosť Zobrazenia Vylúčených Autocitácií	8
3.10.4	Totožné Citácie	8
3.10.5	Zoradenie	8
3.10.5.1	Zoradenie Citácií v Zoznamoch Citácií	8
3.10.5.2	Zoradenie Zoznamov Citácií	8
3.10.6	Zoznamy Citácií	8
3.10.6.1	Citovaná Práca pri Zozname Citácií	8
3.10.6.2	Citujúca Práca	8
3.10.7	Počty Nájdenných Výsledkov pri Každom Zozname Citácií	9
3.10.7.1	Počet Nájdenných Citácií v Zozname s Vylúčením Auto- citácií	9
3.10.7.2	Počet Všetkých Citácií	9
3.10.8	Chýbajúce Výstupné Údaje z Databáz	9
3.10.9	Tlačidlo Kopírovania Záznamu v Internom Formáte do Podpoľa 591 v MARC 21	9
3.10.9.1	Totožné Citácie	9
3.10.9.2	Informácia o Skopírovaní	9
3.11	Formát Citácií	9
3.11.1	Uprednostňovanie Informácií z WoS	9
3.11.2	Konverzia Kapitalizácie Písmen	9
3.11.2.1	Názvy Článkov	9
3.11.2.2	Názvy Časopisov a Iných Zdrojových Dokumentov	10
4	Kvalitatívne Požiadavky	10
4.1	Rozšíriteľnosť	10
4.2	Dlhodobá Udržateľnosť	10
4.3	Dostupnosť a Využitelnosť	10

1 Úvod

1.1 Cieľ Katalógu Požiadaviek

Tento katalóg požiadaviek popisuje požiadavky zadávateľa na funkcionality projektu „Automatické vyhľadávanie ohlasov“, ktorý zadala PhDr. Adriana Gersová ako projekt na predmet Tvorba informačných systémov odboru Aplikovaná informatika na Fakulte matematiky, fyziky a informatiky, Univerzity Komenského v Bratislave.

Katalóg požiadaviek vznikol po dohode so zadávateľom projektu a ako taký je záväzný pre obe zainteresované strany.

Špecifikované požiadavky budú počas vývoja, ako aj po ukončení vývoja softvérového produktu, slúžiť na vyhodnocovanie správnosti softvéru zadávateľom aj realizátorom.

1.2 Rozsah Systému

Aplikácia, ktorej funkcionality bude realizovaná pomocou webovej aplikácie, API a skriptov, umožní používateľovi automaticky prehľadávať citácie k publikáciám v medzinárodných databázach.

Aplikácia cez vstupný formulár dostane potrebné údaje o citovanom dokumente, na základe ktorých pomocou API rozhrania vyhledá relevantné záznamy v daných databázach a spracuje spätné dáta, ktoré zobrazí v naformátovanom tvare. Vyhľadané citácie bude jasne rozlišovať podľa zdroja, informácií, autocitácií, či duplikátov.

Výsledky bude možné jednoducho kopírovať pomocou tlačidla do iných aplikácií a systémov.

1.3 Slovník Pojmov

API: Rozhranie pre programovanie aplikácií.

Back end: Vrstva softvéru na prístup a spracovanie dát.

DOI: jednoznačný autetifikátor záznamu.

Front end: Vrstva softvéru na prezentáciu dát.

Layout: Rozloženie prvkov na obrazovke.

Užívateľské rozhranie (GUI): Používateľské rozhranie, ktoré umožňuje ovládať elektronické zariadenie pomocou súboru interaktívnych obrazových prvkov.

Webová aplikácia: Aplikácia typu klient-server vytvorená pre prostredie internetu alebo intranetu.

Skript: Program v skriptovacom jazyku pôvodne navrhnutý na uľahčenie (zautomatizovanie) operácií v počítači.

Citovaná publikácia: Publikácia, ktorá je citovaná (odkazovaná) inou publikáciou.

Citujúca publikácia: Publikácia, ktorá cituje (odkazuje sa na) inú publikáciu.

1.4 Referencie

Scopus API dokumentácia: https://dev.elsevier.com/sc_apis.html

WoS API dokumentácia: <https://developer.clarivate.com/apis/wos>

1.5 Prehľad Nasledujúcich Kapitol

V druhej kapitole sa nachádza všeobecný popis plánovaného systému prirodzeným jazykom, plynulými vetami bez toho, aby sme išli do veľkých podrobností. Popísané budú funkcionality systému, jeho charakteristiky a obmedzenia.

Tretia kapitola obsahuje zoznam funkčných požiadaviek.

Štvrtá kapitola obsahuje kvalitatívne požiadavky.

2 Všeobecný Popis

2.1 Perspektíva Systému

Systém bude vyvinutý pre účely vyhľadania publikácií a ich počtov citácií (ohlasov o nich). Z výsledkov vyhľadávania bude možné dohľadané informácie ľahko preniesť do iných systémov v textovej podobe (napr. do databázy evidencie publikačnej činnosti).

Systém je prednostne navrhnutý pre publikačnú činnosť Fakulty matematiky, fyziky a informatiky Univerzity Komenského. Systém bude rozšíriteľný aj pre aplikovanie na iných fakultách univerzity.

2.2 Funkcie Systému

Systém po zadaní mena a priezviska autora (povinné údaje), popřípade afiliácie alebo rozsahu rokov vydania publikácie, vyhledá práce v daných databázach a zobrazí ich zoznam. Taktiež bude možné si tieto práce dohľadať podľa DOI. Tento zoradený zoznam bude obsahovať názvy publikácií, autora a rok vydania (3.6).

Následne sa budú dať označiť práce pre ktoré sa majú vyhľadať ohlasy (citácie). Po označení a vyhľadaní sa objavia ďalšie informácie o práci v novom okne a informácie bude možné skopírovať do schránky v textovom formáte pre jednoduchý prenos do iných aplikácií a systémov.

2.3 Charakteristika Používateľa

Systém bude dostupný len pre používateľov, ktorí sa autorizujú kľúčom, keďže niektoré databázy sú licencované.

Osoby pracujúce so systémom budú mať znalosti so spracovaním podobných dát, nakoľko je systém určený pre účely knižníc, evidencie publikačných činností alebo iných odborne zameraných inštitúcií v ktorých je potrebné vyhľadávať citácie pre publikácie.

Softvér bude používaný iba jedným typom používateľa ktorý bude mať prístup ku všetkým funkciám softvéru.

2.4 Všeobecné Obmedzenia

Niektoré z databáz, ktoré systém využíva, sú licencované zdroje, ktoré nie sú voľne prístupné ale sú prístupné len z univerzitných IP adries, alebo cez proxy s autorizáciou menom a heslom. Keďže systém bude fungovať na univerzitnom serveri, na ktorý majú

prístup všetci a z ktorého je prístup do licencovaných databáz, bude systém využívať autorizáciu autorizačným kľúčom (bude len jeden kľúč, uložený v konfiguračných súboroch na serveri).

Obmedzenia pre tento systém bude klásť hlavne dostupnosť kľúču a rýchlosť internetového pripojenia. Rýchlosť dohľadania dát obmedzí aj odozva príslušných databáz a objem dát, ktorý bude závisieť nad rozsahom hľadaných údajov.

2.5 Predpoklady Závislostí

Pre správny chod bude systém vyžadovať prístup k internetu, povolený prístup k databázam citačných indexov a korektne zadané údaje vstupného formulára od používateľa.

Pre optimalizáciu práce s danými údajmi musí byť systém dizajnovaný jednoduchý a prehľadný.

3 Funkčné Požiadavky

3.1 Databázy

3.1.1 Využívanie Databáz

Systém využíva predvolene všetky databázy, ktoré majú zadefinované implementované rozhranie API v zdrojovom kóde.

3.1.2 Nedostupné Databázy

V prípade nedostupnej databázy ju systém vynechá a vyhľadáva len v dostupných databázach.

3.1.3 Informácia o Nedostupných Databázach

Táto požiadavka je voliteľná.

V prípade nedostupných databáz na to systém zatiaľ neurčeným spôsobom upozorní používateľa (napr. formou upozornenia vo webovom prehliadači).

Spôsob závisí od náročnosti jeho implementácie.

3.2 Autorizačný Kľúč

V konfiguračných súboroch na serveri sa nachádza autorizačný kľúč. Pre sprístupnenie údajov (vyhľadávanie prác, zoznamov citácií) je potrebná autorizácia týmto kľúčom.

3.3 Vstupné Prvky Úvodného Formulára

3.3.1 v Prípade, Nevyplneného DOI

povinné: priezvisko autora (povinný údaj, bez diakritiky), meno autora (povinné aspoň jedno začiatkové písmeno, bez diakritiky).

nepovinné: rok/rozsah rokov, afiliácia.

3.3.1.1 Viaceré Roky/Rozsahy rokov Naraz

Možnosť vyplniť viaceré roky a rozsahy rokov oddelené oddeľovačom, ktorý od seba jednoznačne rozlíši roky a rozsahy rokov a bude špecifikovaný v dokumentácii.

3.3.1.2 Viaceré Afiliácie Naraz

Vzhľadom k nekonzistentnosti afiliácie v používaných databázach (napr.: UKOM FKAI, Katedra aplikovanej informatiky FMFI UK atď.) bude možné vyplniť viaceré formy jednej afiliácie oddelené oddeľovačom, ktorý jednotlivé afiliácie jednoznačne odlíši a bude špecifikovaný v dokumentácii.

Pre nájdenie všetkých výsledkov jednej afiliácie bude nutné zadať všetky jej formy, ktoré sú naprieč všetkými databázami ktoré systém využíva použité.

3.3.2 v Prípade Vyplneného DOI

V tomto prípade sa nebudú zohľadňovať žiadne iné údaje okrem údaje DOI, ktorý je v tomto prípade povinný.

3.3.2.1 Viaceré DOI Naraz

Možnosť zadať do formulára viaceré DOI naraz, ktoré budú oddelené oddeľovačom, ktorý ich medzi sebou jednoznačne odlíši a bude špecifikovaný v dokumentácii.

3.4 Zoznam Prác

Po kliknutí na tlačidlo „vyhľadaj záznamy“ sa vyhľadá zoznam prác podľa uvedených kritérií (resp. podľa údajov v úvodnom formulári (3.3)).

3.5 Vyhľadávanie

3.5.1 Vyhľadávanie pri Vyplnenom DOI

Ak je vyplnené pole DOI, zoznam prác je vyhľadaný podľa DOI nezávisle od vyplnenia ostatných (vrátane povinných) údajov.

3.5.2 Vyhľadávanie pri Nevyplnenom DOI

Ak sú vyplnené všetky povinné údaje podľa (3.3.1) a žiadne iné polia, vyhľadáva sa podľa nich.

3.5.2.1 pri Vyplnenej Afiliácii

Ak je okrem všetkých povinných údajov vyplnené aj pole afiliácia, vyhľadávajú sa len výsledky pre príslušnú afiliáciu.

3.5.2.2 pri Zadanom Roku/Rozsahu Rokov

Ak je okrem všetkých povinných údajov vyplnený rok/rozsah rokov (prípadne viaceré roky a rozsahy rokov), vyhľadávajú sa len práce z uvedeného (uvedených) roku/rozsahu rokov (rokov/rozsahov rokov).

3.6 Zoradenie

Zoznam prác bude zoradený najprv podľa roku, potom názvu práce abecedne.

3.7 Informácia o Databáze v Ktoej Bola Položka Nájdená

Vždy je nutné uviesť údaj, v ktorej databáze bola položka nájdená.

3.8 Zaškrťavacie Políčka

Pri každej položke zoznamu je zaškrťavacie políčko, ktorým vieme označiť práce, pre ktoré chceme vyhľadávať citácie.

3.8.1 Odškrtnutie

Po kliknutí na tlačidlo „vyhľadaj citácie“ sa všetky zaškrťavacie políčka odškrtnú, pretože je nepraktické odškrťávať ich pri ďalšom hľadaní manuálne, keďže vyhľadané citácie sa otvoria na novej karte (3.10.2).

3.9 Tlačidlo „Vyhľadaj Citácie“

Mimo zoznamu prác, sa bude nachádzať tlačidlo „vyhľadaj citácie“, ktoré vyhľadá zoznam (zoznamy) citácií na prácu (práce), ktorá (ktoré) mala (mali) zaškrtnuté zaškrťavacie políčko.

3.10 Zoznamy Citácií na Prácu/Práce

Po kliknutí na tlačidlo „vyhľadaj citácie“ sa zobrazí zoznam citácií pre každú zaškrtnutú položku zo zoznamu prác.

3.10.1 Postupné Zobrazovanie Zoznamov Citácií

V rámci technických možností a obmedzení API sa výsledky vyhľadávania (zoznamy citácií) budú zobrazovať postupne, ale zobrazené výsledky už budú finálne a nemenné.

3.10.2 Zobrazenie Zoznamov Citácií na Novej Karte

Zoznamy citácií sa zobrazia na novej karte (alebo v novom okne) vo webovom prehliadači.

3.10.3 Vylúčenie Autocitácií

V zoznamoch citácií na prácu/práce sú vylúčené všetky autocitácie (všetky tie, ktoré v zozname autorov obsahujú meno autora, ktorý a nachádza v zozname autorov citovaného textu).

3.10.3.1 Možnosť Zobrazenia Vylúčených Autocitácií

Zoznam vylúčených autocitácií bude zobraziteľný po kliknutí na tlačidlo „zobraziť vylúčené autocitácie“.

3.10.4 Totožné Citácie

V prípade súbežného vyhľadávania práce vo viacerých databázach (v Scopuse aj vo WoSe), pri nájdení totožných citácií vo viacerých databázach, je potrebné uviesť totožné citácie iba raz (pokiaľ je to možné).

Uprednostňujeme údaje z WoS (3.11.1).

3.10.5 Zoradenie

3.10.5.1 Zoradenie Citácií v Zoznamoch Citácií

Zoznam citácií na citovanú prácu (resp. práce) bude zoradený podľa roku, priezviska a mena prvého uvedeného autora citujúcej práce (citácie).

3.10.5.2 Zoradenie Zoznamov Citácií

Zoznamy citácií, budú zoradené na základe práce naktorú sa vzťahujú podľa pravidiel zoradenia zoznamu prác (3.6).

3.10.6 Zoznamy Citácií

Výsledkom sú zoznamy citácií pre príslušné práce, pre ktorú boli citácie vyhľadávané (3.4).

3.10.6.1 Citovaná Práca pri Zozname Citácií

Pri každom zozname citácií na prácu bude jasne uvedené, ktorá práca bola citovaná.

3.10.6.2 Citujúca Práca

Pri každej citácii na prácu (to sú položky zoznamov citácií) bude jasne uvedené, ktorá práca bola citujúca. Taktiež bude uvedené v ktorej databáze (resp. databázach) bola citácia nájdená.

3.10.7 Počty Nájdenných Výsledkov pri Každom Zozname Citácií

Pri každom zozname citácií, bude uvedené:

3.10.7.1 Počet Nájdenných Citácií v Zozname s Vylúčením Autocitácií

3.10.7.2 Počet Všetkých Citácií

Počet všetkých nájdenných citácií, vrátane autocitácií, pre každý zoznam (3.10.6).

3.10.8 Chýbajúce Výstupné Údaje z Databáz

Ak bude chýbať požadovaný údaj, zvýrazní sa vo výstupe jeho absencia. Požadované údaje sú všetky tie, ktoré sú popísané v dokumente Interný Formát do 591.

3.10.9 Tlačidlo Kopírovania Záznamu v Internom Formáte do Podpoľa 591 v MARC 21

Pri každej citácii bude tlačidlo „Kopírovať“, ktoré skonvertuje údaje o citácii do interného formátu, ktorého pravidlá sú v dokumente Interný Formát do 591, ktorý sa nachádza na githube: <https://github.com/TIS2019-FMFI/hladanie-ohlasov>.

3.10.9.1 Totožné Citácie

V prípade totožných citácií treba prítomnosť citácie vo viacerých databázach vyznačiť podľa pravidiel z dokumentu Interný Formát do 591.

3.10.9.2 Informácia o Skopírovaní

Pri tlačidle sa po skopírovaní zobrazí informácia „Skopírované“.

3.11 Formát Citácií

3.11.1 Uprednostňovanie Informácií z WoS

Pri vyhľadávaní citácií, ktoré sú vo viacerých databázach, je potrebné v prípade totožných citácií (3.10.9.1) uprednostňovať všetky názvy (aj názvy článkov, aj názvy časopisov a zdrojových dokumentov) z databázy WoS.

3.11.2 Konverzia Kapitalizácie Písmen

Aplikácia má za úlohu konvertovať kapitalizáciu písmen podľa nasledujúcich pravidiel:

3.11.2.1 Názvy Článkov

Názvy článkov budú obsahovať veľké písmená iba na začiatku vety, prípadne podľa vlastných mien.

3.11.2.2 Názvy Časopisov a Iných Zdrojových Dokumentov

V prípade názvov časopisov a iných zdrojových dokumentov bude každé začiatkové písmeno slova veľké okrem predpôň a spojok.

Neznáme krátke slovo sa celé vypíše UPPERCASE.

4 Kvalitatívne Požiadavky

4.1 Rozšíriteľnosť

Program bude implementovaný tak, aby ho bolo v budúcnosti možné rozšíriť aj pre prácu s inými databázami okrem WoS a Scopus.

4.2 Dlhodobá Udržateľnosť

Program bude dlhodobo udržateľný a správne funkčný.

4.3 Dostupnosť a Využiteľnosť

Program má byť využiteľný aj pre ďalšie knižnice UK.