

ELVIRA 2.0

Popis produktu

Ing. Jakub Dubec
jakub.dubec@stuba.sk
sik@fiit.stuba.sk

10. januára 2024

Názov: ELVIRA 2.0

Verzia dokumentu: 1.1

Organizácia: FIIT STU

Dátum vytvorenia: 10. januára 2024

História:

Dátum	Autor	Popis
5. november 2023	Ing. Jakub Dubec	Inicializácia
10. januára 2024	Ing. Jakub Dubec	Adaptcia pre širokú verejnosť

Obsah

1	Úvod	5
2	Popis funkcionality	6
2.1	Organizácia publikácií	6
2.2	Generovanie citácií	7
2.3	Zdieľanie dokumentov	9
2.4	Autentifikácia	9
2.5	Responzivita	9
2.6	Polička	10
2.7	Vyhľadávanie	11
2.8	Stahovanie meta-informácií o publikáciách	11
2.9	Hromadné vkladanie publikácií	12
2.10	Ochrana dokumentov a vodoznak	12
2.11	Sledovanie návštevnosti	12
2.12	Dátová analytika	13
3	Prevádzka	14
3.1	Režimy prevádzky	14
3.2	Vkladanie dokumentov	15
4	Technická realizácia	16
4.1	Komponenty	16
4.1.1	Webové rozhranie	16
4.1.2	Prehliadač dokumentov	17
4.1.3	Dokumentový server	18
4.2	Aplikačné rozhrania	19
4.3	Ukladanie dokumentov	19
5	Smerovanie projektu	20
5.1	Digitálne výpožičky	20
5.2	OCR	20
5.3	Širšia podpora rôznych digitálnych formátov	20
5.4	Prístup pre čítačky	20
5.5	Pokročilé vyhľadávanie	21
5.6	Anotačné vrstvy	21
5.7	Záložky	21

Zoznam obrázkov

1	Schéma grarovej organizácie pre FIIT STU	6
2	Základné navigačné feedy	7
3	Generovanie citácií	8
4	Detail publikácie v administrácii	8
5	Vytvorenie kódu	9
6	Úvodná stránka (vľavo) a otvorený dokument (vpravo) na mobilnom zariadení	10
7	Pokročilé vyhľadávanie	11
8	Sťahovanie dát pomocou ISBN v administrácii	12
9	Sledovanie návštevnosti	13
10	Základný prehľad o používaní dokumentového servera v prostredí Metabase	13
11	Režimy prevádzky: distribuovaný (naľavo) a centralizovaný (napravo)	14
12	Diagram komponentov	16
13	ELVIRA portál	17
14	Prehliadač PDF dokumentov	18

Zoznam tabuľiek

2	Definície pojmov	5
3	Aktívne Elvira Inštancie	6

1 Úvod

Tento dokument opisuje funkcionality, technickú realizáciu a spôsob prevádzky pre digitálnu knižnicu ELVIRA.

Dokument je rozdelený na nasledovné kapitoly (v závorke sa nachádzajú aj role, ktorým sú dané kapitoly venované):

- Popis funkcionality: demonštrácia klúčovej funkcionality na obrázkoch s textovým opisom (pre zamestnancou knižnice).
- Prevádzka: opisuje možnosti inštalácie v inštitúcii (pre správcov systému).
- Technická realizácia: opis komponentov, aplikačných rozhraní spolu so spôsobom ukladania obsahu na fyzické médium (pre správcov systému).
- Smerovanie projektu: Popis plánov na rok 2024.

Prípadné cudzie pojmy alebo pojmy so špeciálnym významom sú na tabuľke 2.

Pojem	Popis
Feed	Neusporiadany zoznam publikácií alebo iných feedov

Tabuľka 2: Definície pojmov

2 Popis funkcionality

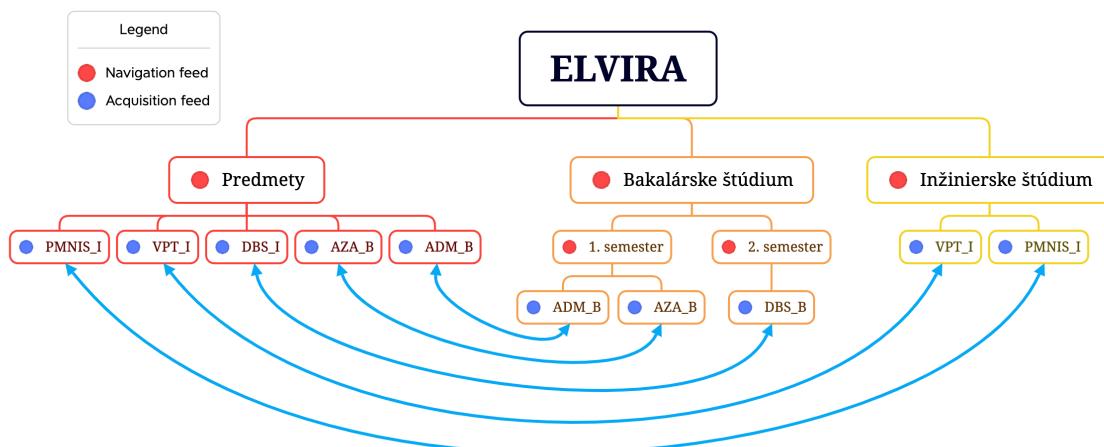
Táto kapitola popisuje klúčovú funkciu digitálnej knižnice. Opísanú funkciu je možné si aj prakticky vyskúšať na inštanciach opísaných v tabuľke 3.

Názov	URL	Prihlasovacie meno	Heslo
DEMO	elvira.vogosphere.sk	arthur.dent	Andromeda246

Tabuľka 3: Aktívne Elvira Inštancie

2.1 Organizácia publikácií

Publikácie na dokumentovom serveri môžu byť organizované formou orientovaného grafu. Uzel takého grafu nazývame **Feed** - zoznam publikácií alebo iných **Feedov**. Príklad takejto organizácie je vizualizovaný na obrázku 1.



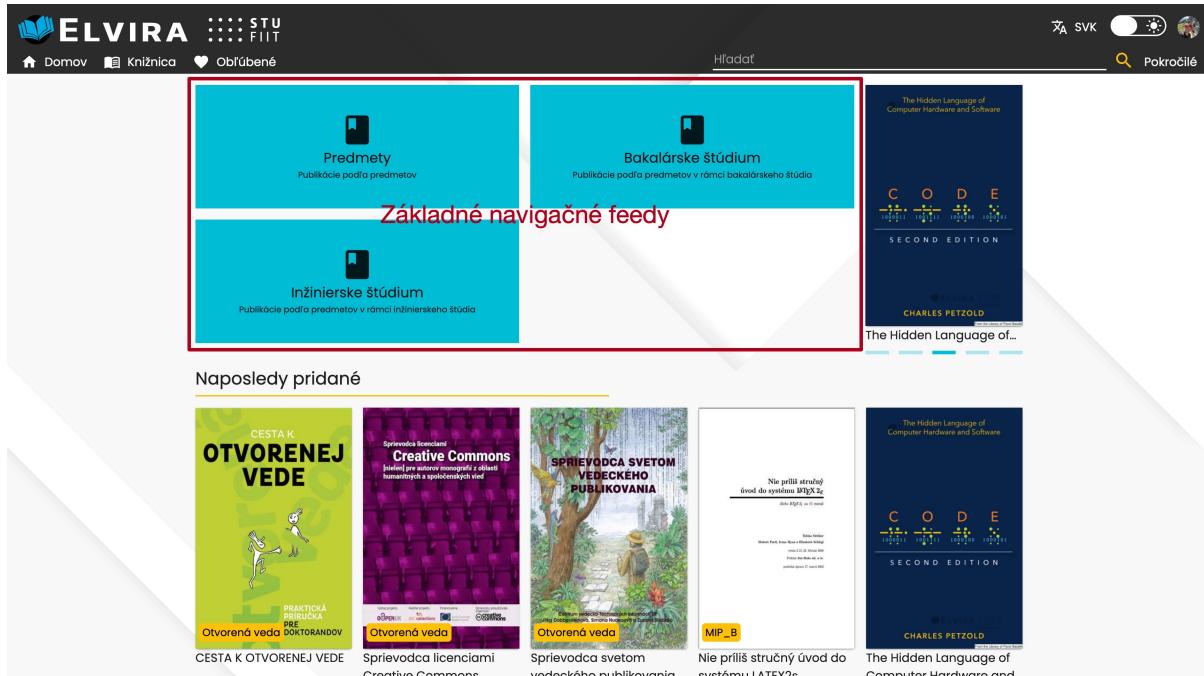
Obr. 1: Schéma grafovej organizácie pre FIIT STU

Podľa obsahu **Feedy** delíme na dve kategórie:

- **Navigačný (navigation) feed** - obsahuje ďalšie feedy. Navigačné alebo akvizičné (acquisition).
- **Akvizičný (aquisition) feed** - obsahuje publikácie.

Akvizičný feed môže byť potomkom viacerých navigačných feedov. V rámci jedného katalógu nemôžu existovať dva feedy s rovnakým názvom.

Pomocou navigačných feedov je možné študenta jednoducho navigovať knižničným katalógom vzhľadom na jeho potreby. Príklad vstupného bodu je na obrázku 2.



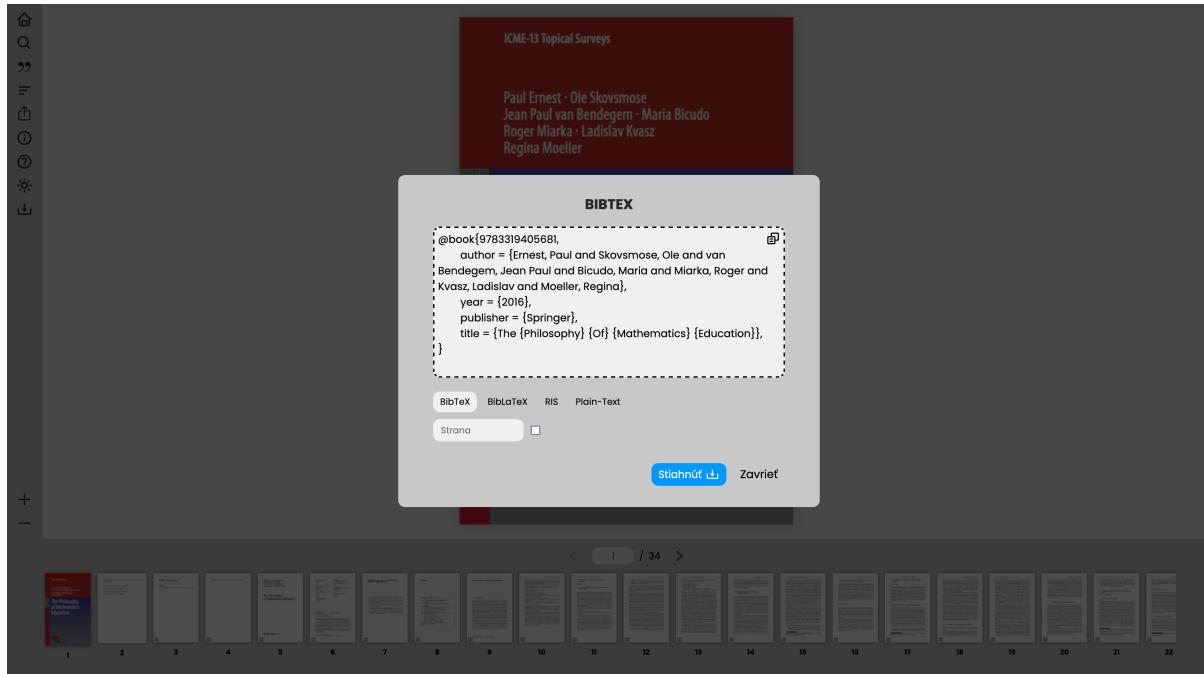
Obr. 2: Základné navigačné feedy

2.2 Generovanie citácií

V prípade, že systém obsahuje všetky potrebné informácie (BibTeX vstup - dá sa automaticky stiahnuť na základe DOI alebo ISBN) používateľ môže generovať citácie pre nasledovné formáty:

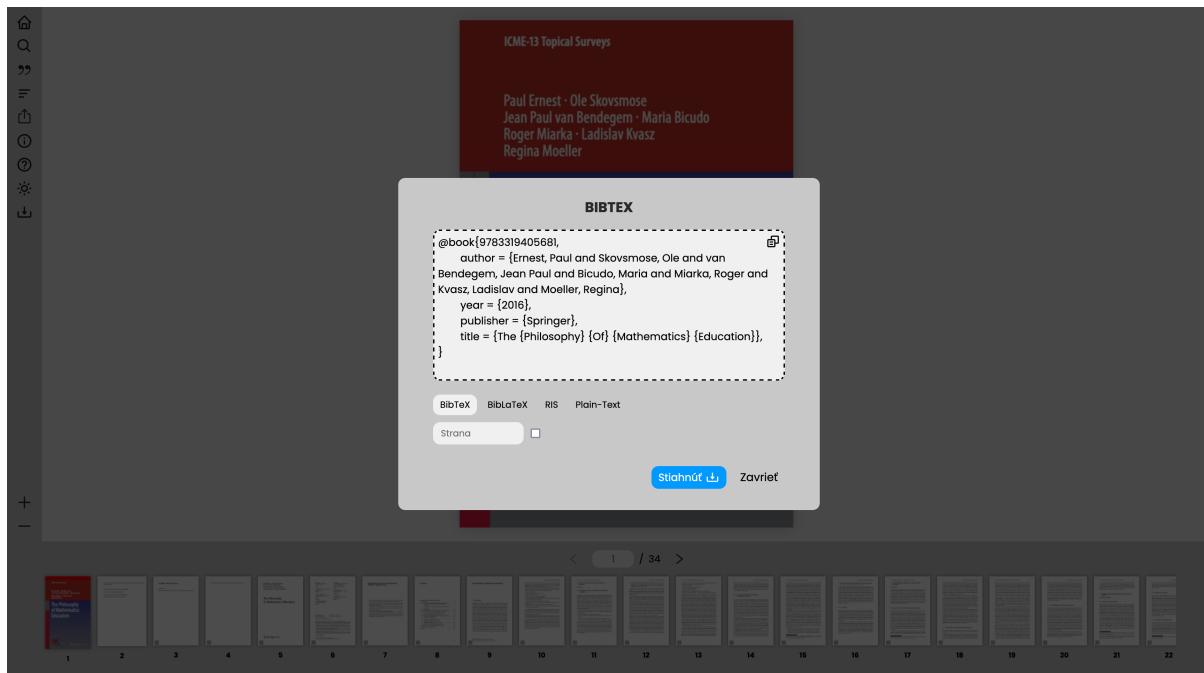
- BibTeX,
- BiBLaTeX,
- RIS,
- Plain-Text (IEEE).

Citácie sú aktuálne podporované na úrovni strany v dokumente. Príklad používateľského rozhrania pre generovanie citácie je na obrázku 3.



Obr. 3: Generovanie citácií

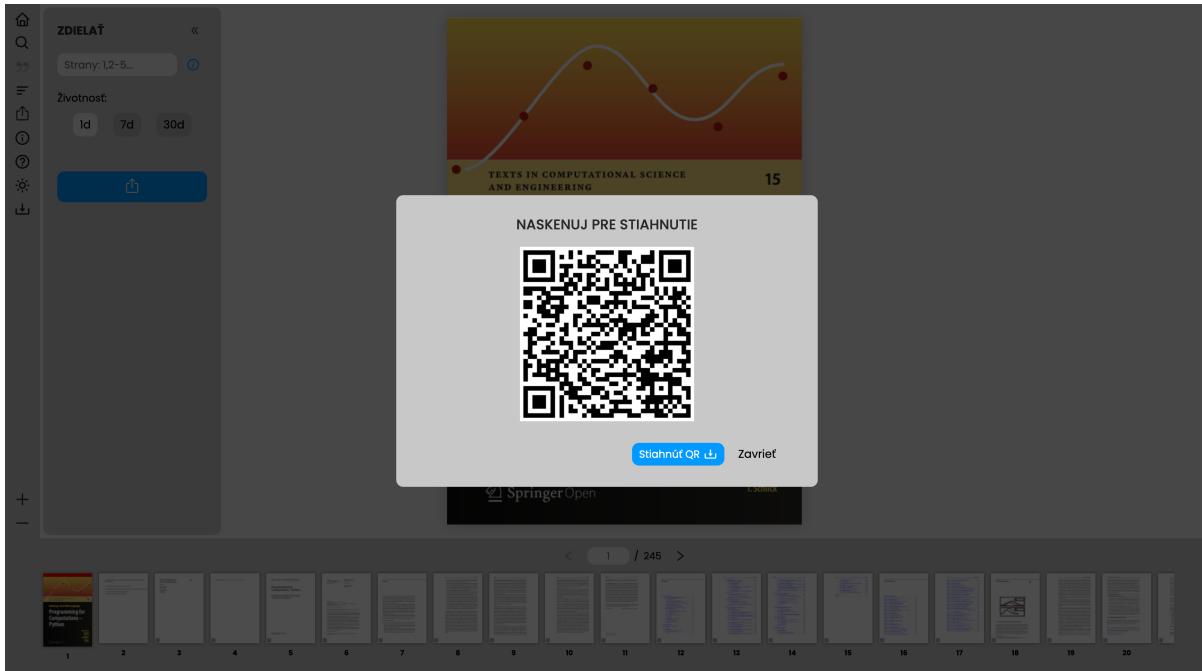
Príklad záznamu, ktorý obsahuje všetky potrebné informácie na poskytovanie generovania citácie je na obrázku nižšie (príklad z administračného rozhrania).



Obr. 4: Detail publikácie v administrácii

2.3 Zdieľanie dokumentov

Systém umožňuje lektorom a študentom zdieľanie dokumentov alebo jeho častí pomocou pohodlných QR kódov (obrázok 5). Taktô vytvorený kód môže odkazovať na konkrétné strany a môže mať obmedzenú platnosť. Pre každý kód systém sleduje návštěvnosť. Vhodné pre zdieľanie učebných materiálov pre konkrétné cvičenia alebo domáce úlohy.



Obr. 5: Vytvorenie kódu

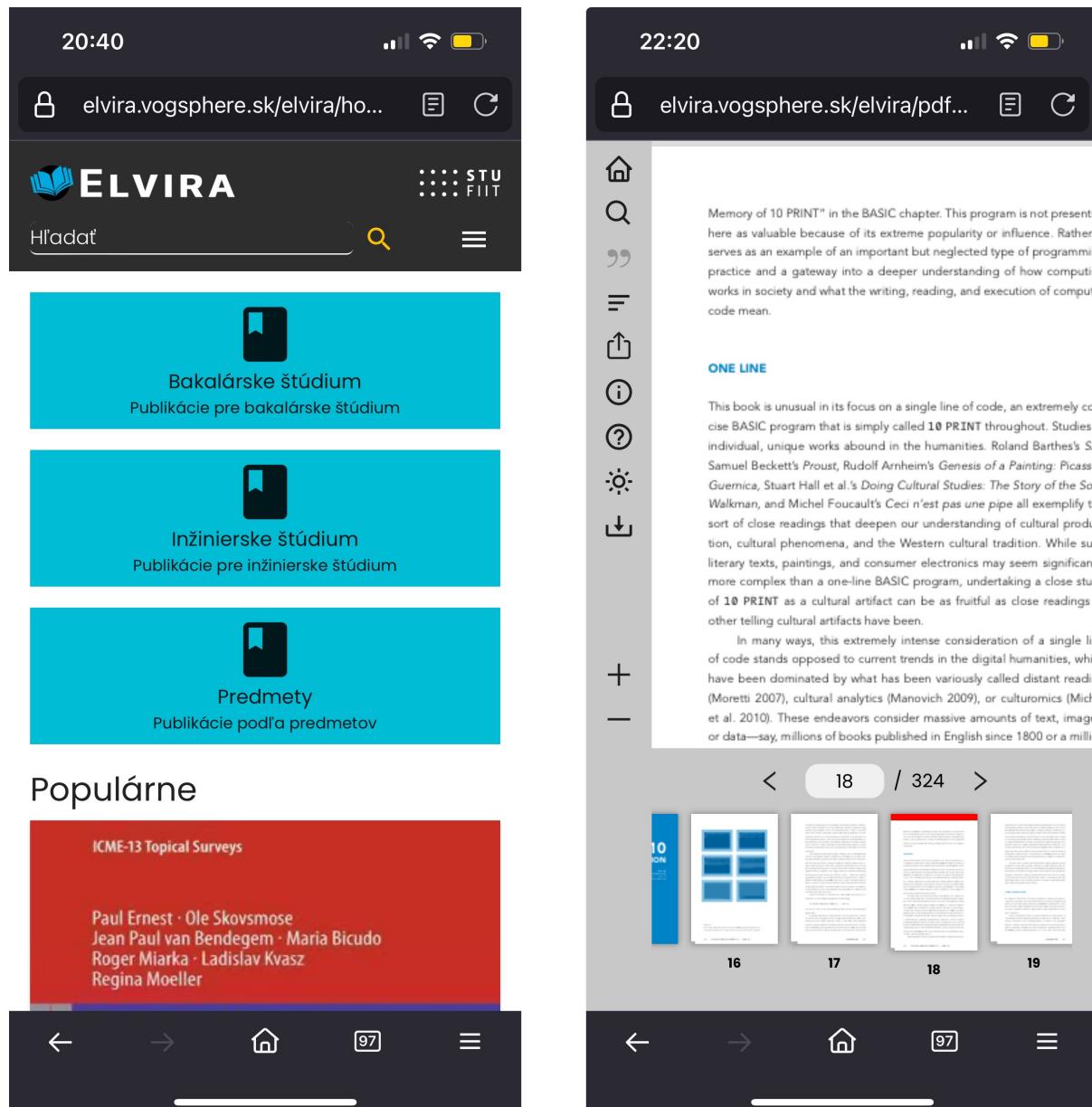
2.4 Autentifikácia

Systém umožňuje čítať informácie o používateľoch z viacerých zdrojov. Aktuálne podporuje lokálnych databázových používateľov a používateľov z LDAP alebo Active Directory serverov. Takýchto zdrojov môže byť definovaných viac (v jednom čase dokáže systém overovať používateľov z viacerých LDAP serverov).

Na základe atribútov z LDAP profilu systém vie upraviť prístup k jednotlivým katalógom prípadne nastaviť širšie práva pre daného používateľa. Aktuálne otestované na STU LDAP serveri.

2.5 Responzivita

Webové rozhranie je pohodlne dostupné z mobilných telefónov a tabletov.



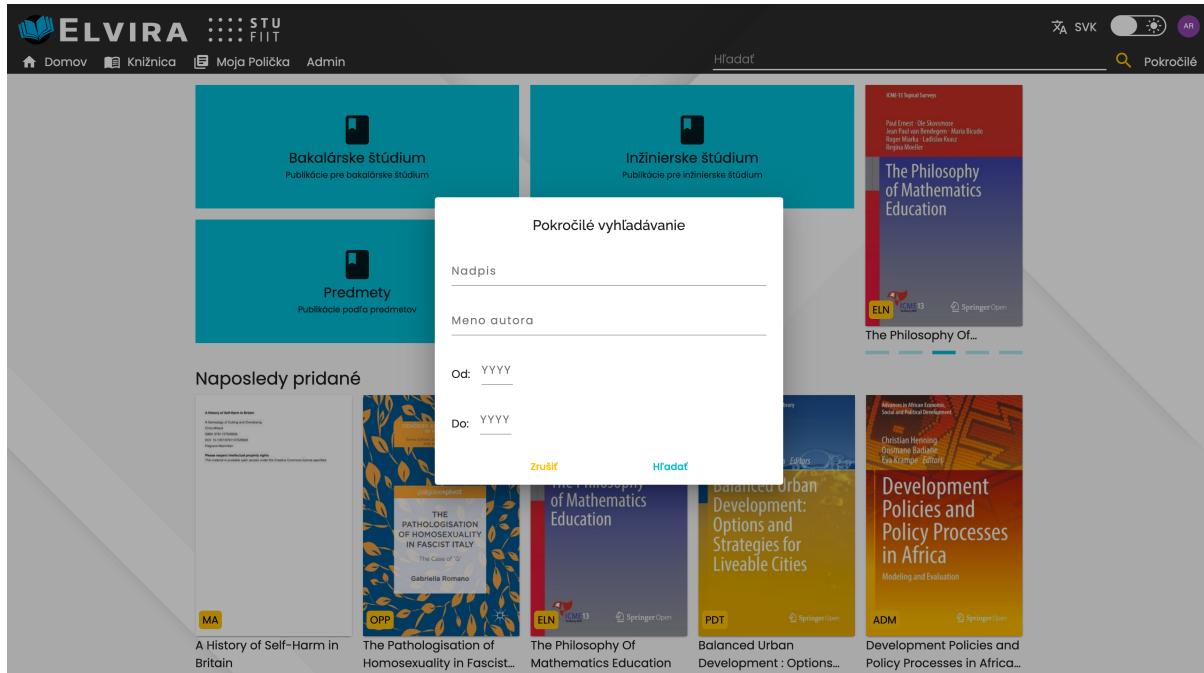
Obr. 6: Úvodná stránka (vľavo) a otvorený dokument (vpravo) na mobilnom zariadení

2.6 Polička

Používatelia si môžu organizovať publikácie v knižnici do svojej poličky pre jednoduchší prístup.

2.7 Vyhladávanie

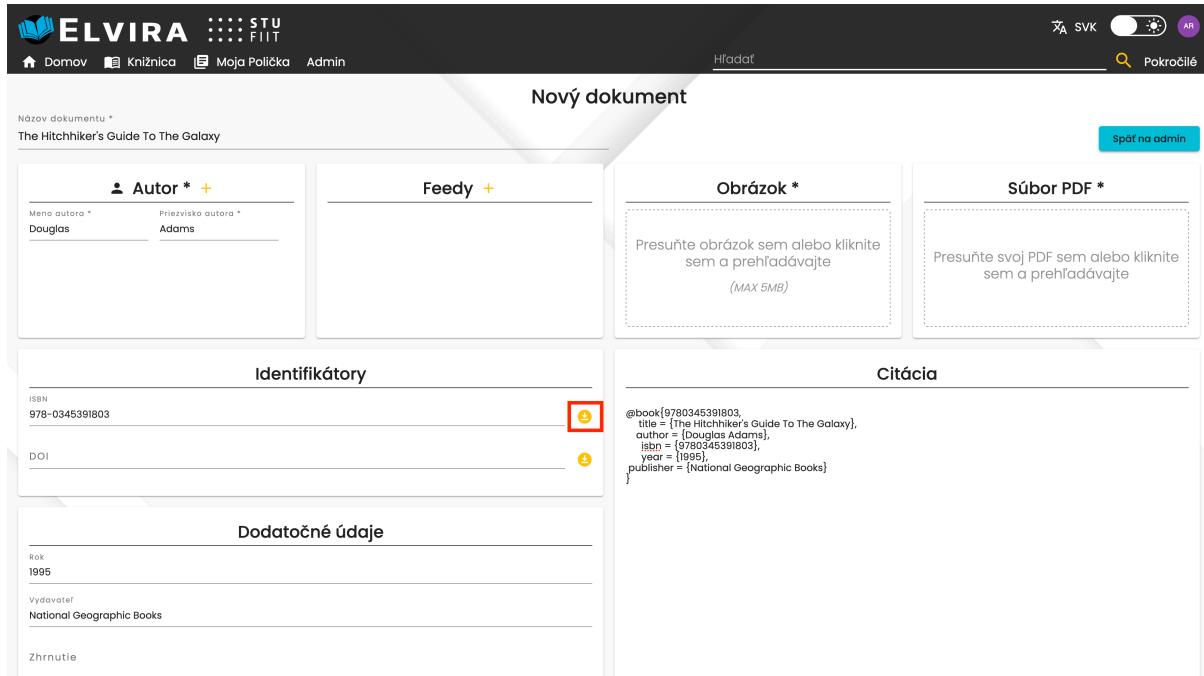
Webové rozhranie poskytuje základné môžnosti vyhladávania medzi publikáciami nezávisle na diakritike a veľkosti písmen. Zároveň sú umožnené aj pokročilé vyhladávania na základe mena autora alebo roku publikovania. Táto funkcia ešte prebieha aktívnym vývojom s cieľom umožniť komplikovanejšie dotazy. Príklad pokročilého vyhladávania na obrázku 8.



Obr. 7: Pokročilé vyhľadávanie

2.8 Sťahovanie meta-informácií o publikáciách

Systém umožňuje automatické vyplňanie detailu o publikácii na základe DOI alebo ISBN. Stačí vyplniť identifikátor a v prípade, že existuje záznam vo verejných databázach automaticky sa vyplnia informácie o autoroch, názve titulu a podklady pre generovanie citácií.



Názov dokumentu *
The Hitchhiker's Guide To The Galaxy

Autor * +
Meno autora * Douglas
Priezvisko autora * Adams

Feedy +

Obrázok *
Presuňte obrázok sem alebo kliknite sem a prehľadávajte (MAX 5MB)

Súbor PDF *
Presuňte svoj PDF sem alebo kliknite sem a prehľadávajte

Identifikátory

Citácia

Dodatačné údaje

ISBN
978-0345391803

DOI

Rok
1995

Vydavateľ
National Geographic Books

Zhrnutie

```
@book{9780345391803,
  title = {The Hitchhiker's Guide To The Galaxy},
  author = [Douglas Adams],
  isbn = {9780345391803},
  year = {1995},
  publisher = {National Geographic Books}
}
```

Obr. 8: Sťahovanie dát pomocou ISBN v administrácii

2.9 Hromadné vkladanie publikácií

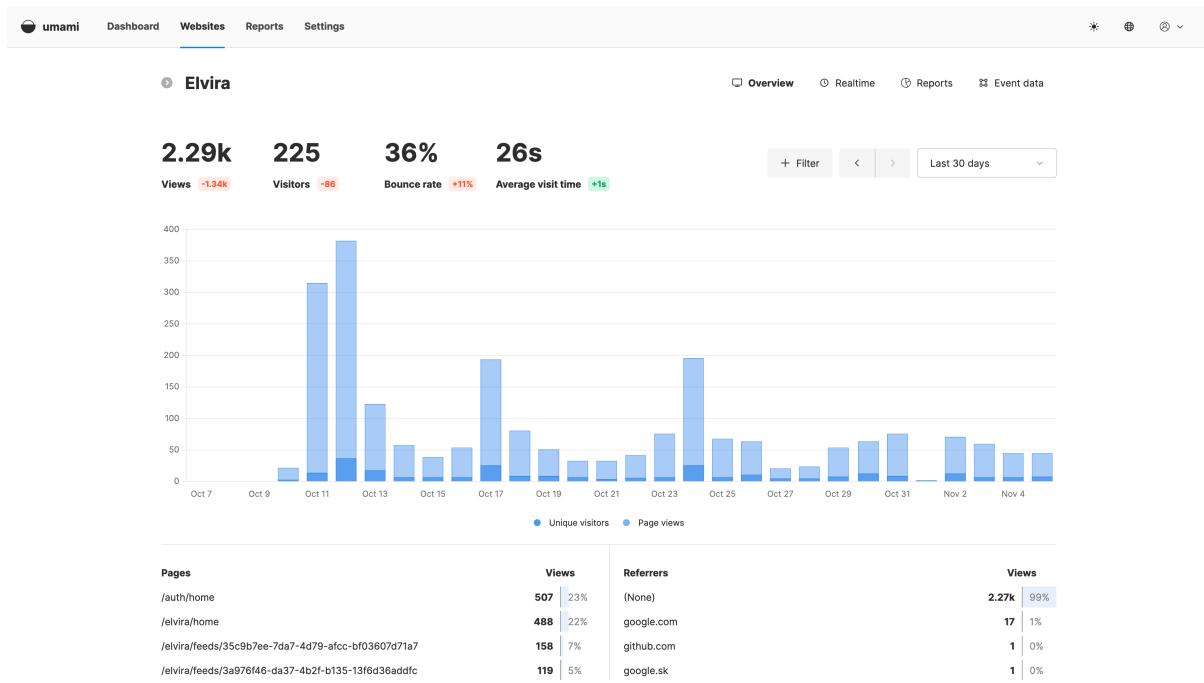
V prípade potreby vie byť jednoducho implementovaná migrácia dát (v prípade zabezpečenia všetkých podmienok z kapitoly Prevádzka).

2.10 Ochrana dokumentov a vodoznak

Každý dokument v systéme ELVIRA je špeciálne generovaný pre používateľa, ktorý ho preberá. Prebratý dokument obsahuje informácie komu bol vystavený a s akou licenciou. Každá strana obsahuje QR kód, ktorý nesie informáciu o tom kde, kedy a komu bol tento dokument vystavený. Systém podporuje aj vkladanie vodoznakov vo forme obrázka.

2.11 Sledovanie návštevnosti

Na zbieranie informácií o návštevnosti používame open-source nástroj umami, ktorý kladie dôraz na zachovanie súkromia. Príklad výstupu na obrázku 9.



Obr. 9: Sledovanie návštěvnosti

2.12 Dátová analytika

Na pokročilejšiu dátovú analytiku nad dokumentovým serverom používame open-source nástroj (Metabase). Vďaka tomuto komponentu je možné, aby aj ľudia mimo technického prostredia vedeli získavať základné analytyky o používaní systému formou kladenia otázok v grafickom rozhraní.



Obr. 10: Základný prehľad o používaní dokumentového servera v prostredí Metabase

3 Prevádzka

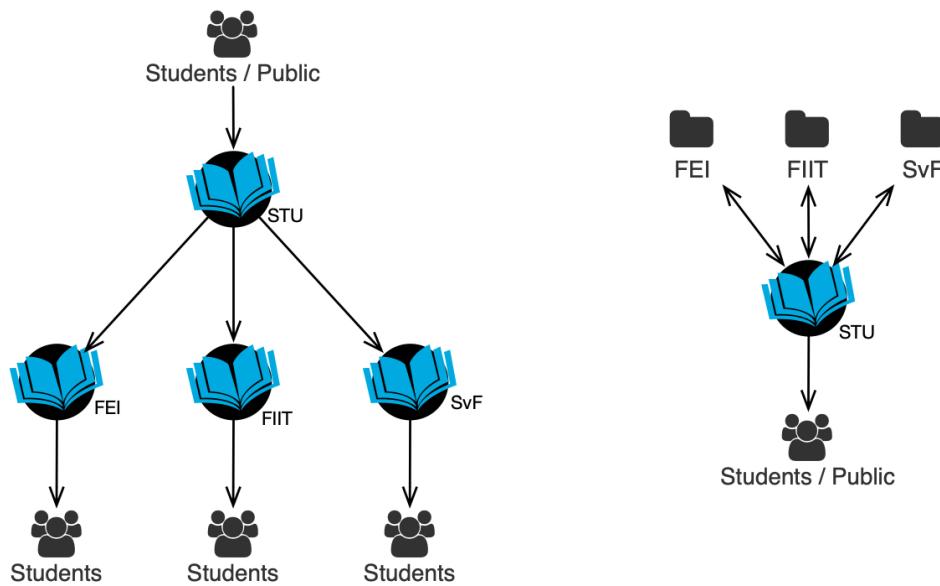
Dokumentový server je distribuovaný ako Docker image, ktorý má závislosť na PostgreSQL databáze. Dokumenty môže ukladať na fyzický disk alebo sieťový pomocou S3 protokolu (napríklad na CephFS cluster alebo kompatibilného cloud úložiska).

V rámci univerzity je možnosť centrálnej prevádzky, kde každá fakulta disponuje samostatným katalógom. Prípadne, pre zvýšenie kontroly každá fakulta môže disponovať vlastnou inštanciou. Tieto inštancie môžu bežať aj na serveroch jednotlivých fakúlt. Inštancie dokumentového serveru disponujú proxy funkciaľitou a vedia distribuovať odkazovanie obsahu na ostatné inštancie.

3.1 Režimy prevádzky

Systém podporuje dva režimy prevádzky:

- centrálny: existuje centrálna inštancia pre všetky fakulty, kde každá fakulta má vlastný katalóg.
- distribuovaný: každá fakulta má vlastnú inštanciu, ktorá môže byť prevádzkovaná aj na ich vlastnej infraštukture. Časom môže vzniknuť aj inštancia v strede, ktorá bude distribuovať verejne dostupný obsah širokej verejnosti (dáta ostávajú nadálej v jednotlivých inštanciách).



Obr. 11: Režimy prevádzky: distribuovaný (naľavo) a centralizovaný (napravo)

V oboch režimoch si fakulty zodpovedajú samé za distribuovaný obsah (administrácia katalógu je priamo v rukách každej fakulty). Režimy sú kombinovateľné.

3.2 Vkladanie dokumentov

Systém aktuálne podporuje vkladanie dokumentov po jednom pomocou administračného rozhrania. V prípade, že je potrebné vykonať migráciu z existujúceho systému, tento proces je realizovateľný vďaka dostupnému RESTful aplikáčnému rozhraniu na strane dokumentového servera. Aktuálne disponujeme migračnými skriptami na migráciu z OPDS kompatibilných dokumentových serverov.

Prípady, ktoré vedia byť strojovo vyriešené:

- Migrácia z lokálneho disku (v prípade, že disponujeme zoznamom publikácií spoločne s ich meta-informáciami a cestou k PDF súboru).
- Migrácia z existujúceho systému (za predpokladu, že dostaneme popis vstupu a testovacie dát).

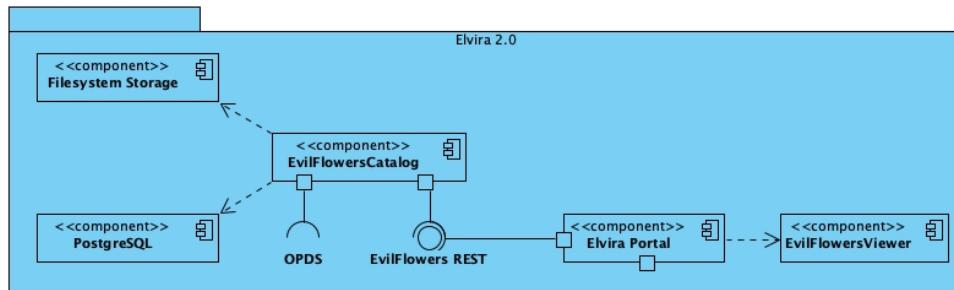
4 Technická realizácia

Nasledovná kapitola stručne opisuje technickú realizáciu ELVIRA projektu a ich konkrétnu funkcionality.

4.1 Komponenty

Projekt ELVIRA môže byť rozdelený na komponenty z obrázku 12:

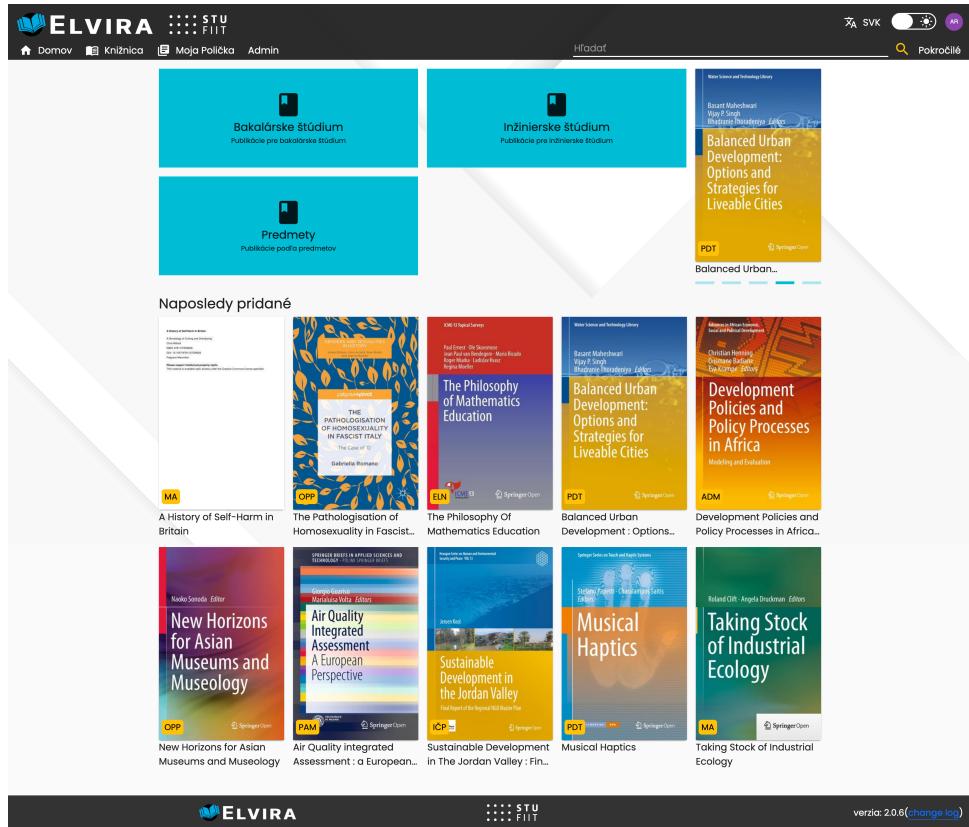
- **ElvFlowersCatalog**: reprezentuje dokumentový server, ktorý organizuje a distribuuje uložené dokumenty.
- **Elvira Portal**: webové rozhranie, ktoré používa čítač a správca knižnice na prístup k dokumentom.
- **ElvFlowersViewer**: webový prehliadač na PDF dokumenty.



Obr. 12: Diagram komponentov

4.1.1 Webové rozhranie

Webová aplikácia napísaná v Next.js frameworku, ktorá slúži na prehliadanie dokumentového servera. Jej súčasťou je administrácia pre správcu knižnice. Príklad stránky, ktorá sa ukáže hned po príhlásení je na obrázku 13.



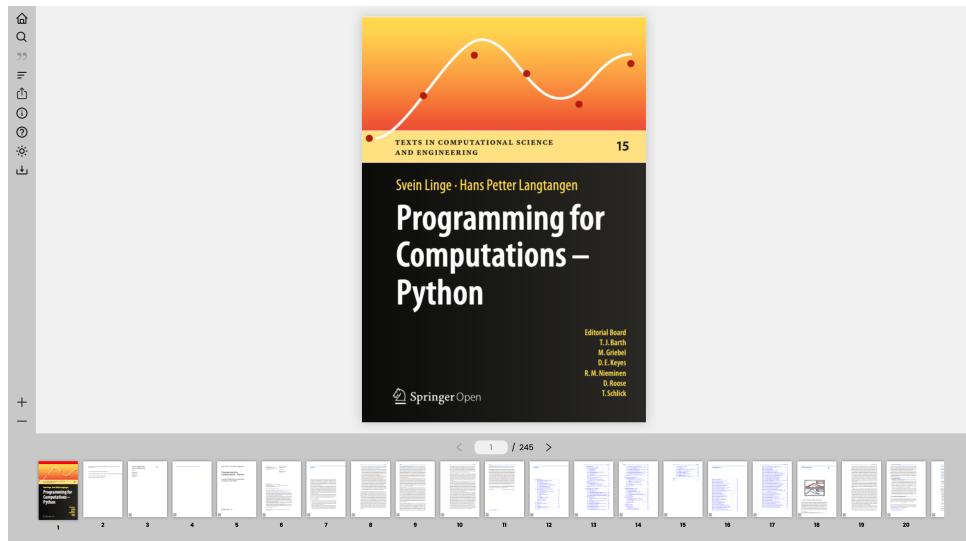
Obr. 13: ELVIRA portál

4.1.2 Prehliadač dokumentov

Webová aplikácia, ktorá slúži na prehliadanie dokumentov (obrázok 14) spolu s nasledovnými funkcionality:

- vyhľadávanie v dokumente,
- klávesové skratky na pohodlné čítanie,
- generovanie citácií,
- prehliadanie a navigácia v obsahu,
- zdieľanie dokumentu alebo jeho častí,
- stiahnutie dokumentu (ak je pre danú publikáciu dovolené).

Zdrojový kód projektu môžete nájsť na GitHube: EvilFlowersCatalog/EvilFlowersViewer.



Obr. 14: Prehliadač PDF dokumentov

4.1.3 Dokumentový server

Server postavený nad Django frameworkom, ktorý organizuje a distribuuje dokumenty. Zoznam hlavných funkcionalít:

- organizácia do feedov,
- autentifikácia,
- tvorba vodotlače,
- sledovanie premávky,
- zdieľanie dokumentov,
- sťahovanie meta-informácií na základe DOI alebo ISBN,
- pokročilé vyhľadávanie v dokumentoch,
- REST a OPDS rozhrania,
- rôzne dátové úložiská (lokálny systém alebo S3),
- úprava dokumentov (vkladanie strán),
- optimalizácia dokumentov (komprezia, značkovanie).

Zdrojový kód projektu môžete nájsť na GitHube: [EvilFlowersCatalog/EvilFlowersCatalog](https://github.com/EvilFlowersCatalog/EvilFlowersCatalog).

4.2 Aplikačné rozhrania

Dokumentový server implementuje niekoľko aplikačných rozhraní za účelom zabezpečenia strojového prístupu a prípadnej automatizácie:

- REST: strojový prístup ku všetkým funkciaľitám dokumentového servera pomocou HTTP protokolu. Používa webové rozhranie na komunikáciu s dokumentovým serverom.
- OPDS: používa sa na prístup k publikáciám pomocou čítačky.

4.3 Ukladanie dokumentov

Dokumenty vedia byť ukladané v dvoch režimoch:

- Priamo na **súborový systém**, priamo pripojený k dokumentovému serveru.
- Pomocou **S3 protokolu** na sieťový súborový systém ako napríklad CephFS.

5 Smerovanie projektu

Projekt ELVIRA podlieha aktívnemu vývoju, v ktorom sú sledované dlhodobé ciele, ale zároveň flexibilne reaguje na aktuálne požiadavky od používateľov. Táto kapitola obsahuje aktuálne stanovené ciele pre rok 2024, ktorých vývoj už prebieha.

Vývoj dokumentového servera môže byť sledovaný na verejnej stránke projektu: Roadmap. Postupne sa na danom odkaze zbierajú požiadavky pre nové funkcionality. Začiatkom roka 2024 na odkaze už budú sledované úlohy aj pre webové rozhranie aplikácie.

Vývojový cyklus projektu je stanovený na jeden rok, kedy väčšie funkcionality sú doručené vždy pred začiatkom zimného semestra (prelom augusta / septembra). Menšie funkcionality, úpravy zabezpečenia alebo kozmetické zmeny sú doručované aj počas semestra.

5.1 Digitálne výpožičky

Podpora digitálnych výpožičiek pomocou platformy Radium. Používatelia si budú vedieť v rámci ELVIRA vypožičať digitálne dokumenty, ktorých licencia vyžaduje sledovanie vypožičaných kusov.

5.2 OCR

Automatické pridávanie textovej anotačnej vrstvy pre PDF dokumenty, ktoré neobsahujú textovú vrstvu. Dokumentový server aktuálne disponuje OCR funkcionalitou. Nie je vyriešené riadenie z administrácie.

5.3 Širšia podpora rôznych digitálnych formátov

Aktuálne digitálna knižnica podporuje PDF dokumenty. Cieľom je rozšíriť systém tak, aby podporoval aj rôzne multimediálne formáty a dokumenty typu ePUB. Presný zoznam potrebných formátov by mal vzniknúť po diskusii so správcami knižníc podľa ich konkrétnych požiadaviek (výkresy, plagáty a podobne).

5.4 Prístup pre čítačky

Sprístupniť digitálnu knižnicu pre čítačky, ktoré podporujú OPDS protokol. Dokumentový server aktuálne disponuje OPDSv1.2. v rámci tohto cieľa plánujeme podporovať aj OPDSv2 a napísanie návodov pre populárne čítačky.

5.5 Pokročilé vyhľadávanie

Aktuálne vyhľadávanie prebieha na úrovni vybraných atribútov ako názov, popis alebo kategória alebo iba v obsahu konkrétneho dokumentu. Cieľom je vybudovať vyhľadávanie, ktoré rozumie jazyku a vyhľadáva nad korpusom dokumentu (spolu s jeho parametrami) v rámci celého systému.

5.6 Anotačné vrstvy

Umožniť čitateľovi zvýrazňovať text alebo kresliť na dokumenty v rámci webového rozhrania. Tieto anotačné vrstvy budú držané v rámci systému a používatelia si ich budú môcť navzájom zdieľať.

5.7 Záložky

Umožniť čitateľovi vytvárať záložky v dokumente podľa jeho vlastnej potreby.