

АКАДЕМИЧНА СИСТЕМА ЗА ПУБЛИКАЦИИ

Георги Георгиев, Димитър Христов, Ася Стоянова-Дойчева

Резюме. В настоящата статия се описват и анализират проблемите и задачите, свързани със съхранението и обработката на данни във ViPS (Virtual Physical Space) [1]. Изследват се съществуващите платформи за съхранение на данни, в частност DSpace [2] и възможностите за тяхното приложение не само като средство за публикуване, даващо по-голяма видимост и достъпност на научните материали, но и по начин, позволяващ интеграцията на дигитална библиотека във ViPS и взаимодействието ѝ с останалите компоненти на пространството.

Представя се функционална уеб-базирана академична система за публикации, създадена посредством имплементиране на DSpace REST API [3], която администрира факултети и катедри, позволява изграждането на потребителски колекции в тях, както и създаване, промяна и изтриване на публикации в дадена колекция. Приложението е база за надграждане с бъдещи интелигентни функционалности чрез използване на BDI (belief-desire-intention) [4] модела и стъпка към създаването на интелигентно хетерогенно хранилище за данни.

Ключови думи: *Virtual Physical Space, дигитална библиотека и институционално хранилище (institutional repository), система за публикации*

1. Въведение

Дигиталната библиотека [5] представлява фокусирана онлайн колекция от разнообразни дигитални обекти, включващи текст, видео и аудио, заедно с методите за достъп и извличане, за селекция, организация и поддръжка на колекцията. Понятието „дигитални обекти“ е с широко тълкувание. Освен аудио, текст и видео се включват 3D обекти, симулации, динамични визуализации и виртуална реалност. Терминът фокусирана колекция, съсредоточена върху дадена област на знанието, заменя Александрийския модел на идеална библиотека, която е огромна и расте вечно – това неминуемо води до проблем с огромните колекции книги.

Цифровата библиотека не е просто „дигитализирана библиотека“, а нов начин за боравене със знанието – съхраняване, събиране, разпространение и достъп. Дигиталните библиотеки могат да се различават значително по размер и обхват и да бъдат поддържани от отделни лица или организации.

Предимствата на цифровите библиотеки като средство за лесно и бързо достъпване на материали от различен тип вече са широко признати. Те нямат физическа граница и са налични денонощно онлайн, позволявайки едновременно достъп. Извличането на информация е лесно при подходящо изграден интерфейс - потребителят може да използва всяка фраза за търсене в цялата колекция. Докато традиционните библиотеки са ограничени от пространството за съхранение, цифровите библиотеки имат потенциала да съхраняват много повече информация, изискват много малко физическо пространство. Има възможност за добавяне на стойност – определени характеристики на обектите, примерно качеството на изображенията може да бъде подобро, да се подобри четливостта и да премахне видими недостатъци като петна и обезцветяване.

Цифровите библиотеки имат и своите предизвикателства в области като: Потребителска идентификация за достъп до колекции, авторско право, цифрово съхранение и организация на информацията, дизайн на интерфейса, оперативна съвместимост между системите и софтуера, качество на метаданните и други.

2. Основни технологии в дигиталните библиотеки

Основните и най-разпространените решения в сферата на дигиталните хранилища са няколко на брой: EPrints, Digital Commons – Bepress, Fedora Commons – DuraSpace, DSpace – DuraSpace.

2.1. EPrints

EPrints [6] е първата професионална софтуерна платформа за изграждане на висококачествени хранилища, съвместими с Open Archive Initiative (OAI) [11] и е утвърдена като един от най-лесните и бързи начини за създаване на хранилища с научна литература, научни данни, теми, доклади и мултимедия. Платформата е написана на езика Perl.

EPrints е система за управление на документи, насочена към висшите учебни заведения и личности. Тя се използва за предоставяне на отворен достъп до резултатите от изследвания, споделяне на образователни ресурси и изграждане на портфолиа. Той има гъвкав модел на метаданни и работен процес, който поддържа разнообразните бизнес нужди, както и предлага взаимодействие с други веб и институционални системи.

EPrints се състои от три ключови компонента: Core API – предоставя основно въведение към използването на EPrints API, StyleGuide – дава указания как се кодира в съвместим стил, Modules.

2.2. Digital Commons

Digital Commons [7] е хоствана IR (Institutional Repository) платформа, лицензирана от Berkeley Electronic Press (или Bepress) и официално стартира през 2004 г. Външният хостинг отнема голяма част от техническата работа, както и необходимостта от програмисти участващите в поддържане на собствена инфраструктура за хранилища. Абонатите подписват годишен лиценз, разходите от които варират според размера на институцията.

Digital Commons поддържа Qualified Dublin Core. Meta Encoding and Transmission Standard (METS) [16], Machine-Readable Cataloging (MARC) [17], въпреки че в интерфейса има елементи, които не са Dublin Core специфични. Това дава възможност, все пак, да се добавят и допълните метаданни.

Digital Commons е регистриран доставчик на OAI данни [11] (т.е. „излага метаданните на света чрез използване на протоколът OAI-PMH“), но не предоставя услуги („използва метаданните, събрани чрез OAI-PMH като основа за изграждане на допълнителни услуги“). В случай на необходимост от миграция, съдържанието и метаданните могат да бъдат експортирани чрез OAI търсене(OAI harvesting), но това изисква усъвършенствани програмни умения и има доста институции, които срещат проблеми в това отношение.

Потребителският интерфейс е чист и удобен за потребителя, но не толкова гъвкав, колкото Fedora или EPrints. За потребителите е лесно да представят и управляват файлове, да получават известия по имейл и RSS емисии, както и месечни имейл съобщения от активност / изтегляне на изявления.

2.3. Fedora Commons

Fedora Commons [8] е модулна архитектура за управление на цифрови активи, първоначално разработена от изследователи в Университета Корнел и Библиотеката на Университета във Вирджиния през 2003 г. Това е рамка без вградена функционалност за управление, индексване, откриване и доставка на елементи. Вместо това, тя е предназначена да позволи висока степен на гъвкавост, което дава възможност на разработчиците да реализират практически всяка функция, както и да интегрират софтуера на трети страни в рамката. Това ниво на гъвкавост идва с цената на леснота на изпълнение.

Islandora, първоначално разработена от University of Prince Edward Island през 2009 г., е реализирала „най-добрите практики“ на базиран на Fedora софтуер, който използва Drupal [18] като клиентска част. Основното предимство на Islandora е, че премахва много от най-големите бариери, срещани често от институциите, заинтересовани от създаването на Fedora хранилища свързани с нуждата от цял екип програмисти за получаването на краен продукт. Проектът също така се издава с редица силно конфигурируеми модули, инструменти, които улесняват изпълнението на широк спектър от действия.

Основните метаданни са Dublin Core XML, въпреки че е възможен всеки формат (такъв като MODS или Qualified Dublin Core). Всеки пакет позволява създаването на персонализирани метаданни, формуляри, използващи пакета от модули XML форми.

Друго решение, основаващо се на Fedora е Hydra. Вместо да използвате PHP базиран интерфейс като Drupal, Hydra използва Ruby on Rails. Понастоящем са ангажирани четиринадесет институционални партньори за подпомагане на проекта.

2.4. DSpace

DSpace [2] е цифрова система за управление на активи с отворен код, създадена първоначално от разработчиците от MIT и HP Labs през 2002 г. Най-често се използва като институционално хранилище (Tansley, Smith, & Walker, 2005). Това е най-популярното и изпитано решение за хранилища, което е на разположение. Докато DSpace споделя известно припокриване на функции със системи за управление на съдържанието и системи за управление на документи, софтуерът на DSpace обслужва специфична нужда като цифрова архивна система, фокусирана върху дългосрочното съхранение, достъп и запазване на цифровото съдържание.

DSpace поддържа по подразбиране метаданни с Qualified Dublin Core и може да експортира в 11 други формати: OAI_DC, METS, MODS, OAI-ORE, RDF, MARC и XOAI. Освен това е възможно да се създаде схема за персонализирани метаданни с помощта на XML.

Потребителският интерфейс е функционален, но донякъде датиран и не отговаря на съвременните изисквания за дизайн, където имаме широк набор от различни устройства с различни размери. Към 2018 година DSpace се разпространява с две клиентски имплементации – JSP UI, XML UI. В началото на 2019 се очаква да излезе предварителна версия на DSpace 7. Според Dspace Working Group [12] две са основните задачи на седмата версия: Да унифицира двата различни потребителски интерфейса в един написан на Angular 2+, Да пренапише цялостно REST интерфейса, така че

да може да се използва пълната функционалност на DSpace (включително потребители и техните групи, което не е възможно във версия 6.x)

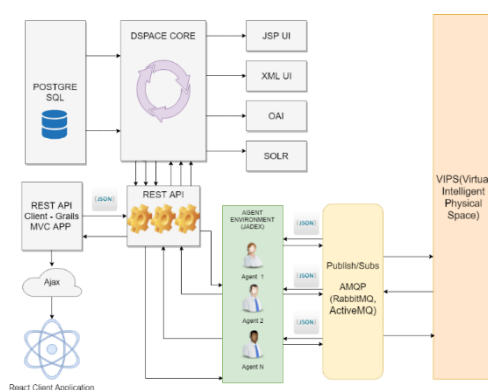
DSpace се е доказала като солидна платформа за хранилища от стартирането ѝ през 2002 г. С пускането на 7-ма версия през януари / февруари 2019 г. – DSpace остава обещаващ и достоен избор сред своите конкуренти, имайки предвид нарастващата нужда от по-голяма подкрепа за данните от изследванията.

За целта на проекта „Академична система за публикации“ сме се спрели на DSpace. Бидейки една от най-стабилните и широко разпространени системи в своята категория, DSpace предлага най-голям набор от инструменти за постигането на целите на проекта. Ползва стандартно семантично описание на публикации – Dublin Core стандарт за метаданни [9], разработен, за да се използва при описанието на уеб базирани документи. Dublin Core е широко разпространен и се използва от много общности и проекти за каталогизиране на данни, което би улеснило евентуално взаимодействие на системата за публикации с тези приложения и обмена на данни с тях. Една ключова характеристика, обаче, имаше най-голяма тежест при нашия избор. DSpace предлага RESTful API, което позволява надграждането на вече текущите вградени функционалности. Фундаментално за този проект е вграждането на интелигентни персонални агенти в основната на всяко действие предприето от потребителя на системата.

3. Реализация на система за публикации

3.1. Архитектура

Общата диаграма на архитектурата на приложението за Академични публикации изглежда така:



Фигура 1. Архитектура на системата за публикации

Dspace Core, Postgre SQL базата данни и XML UI са част от основния пакет от проекти на DSpace. Същинската част на имплементацията на проекта за Академични публикации започва от Rest API Client/ MVC App.

За реализацията на това приложение е използвана рамката за създаване на уеб приложения Grails. Версията е 3.2.5. След като клиента консумира данните от RESTful API на DSpace, данните се трансформират в локални модели и се подават към следващия клиент. Като входни и изходни точни са предоставени контролери, които предоставят данните в JSON формат. За създаване на потребителския интерфейс е използвана javascript рамката Reactjs. Посредством асинхронни заявки се консумират данните от контролерите и се излагат в отделни изгледи. Общия модел на такъв тип приложения се наричат SPA приложения (single-page application). За управлението на състоянието на клиентското приложение се използва React-Redux.

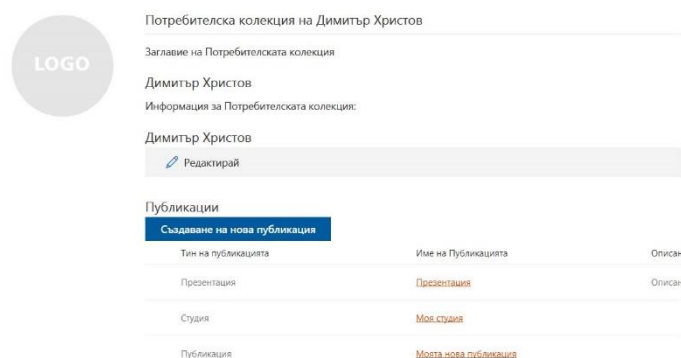
Създавайки такъв междинен слой между RESTful API и крайния клиент, е най-подходящото място, където интелигентната функционалност на архитектурата (фигура 1) може да бъде реализирана. Както е описано по-горе, това дава възможност за наследяване и разширяване на вече съществуващата функционалност на DSpace. Подробно описание на агентската част (фигура 1) ще бъде дадено в последващи публикации.

3.2. Описание на потребителския интерфейс

Академичната система за публикации е изградена на база имплементирането на DSpace REST API. Програмирани са методи за автентикация, както и за обработване на съставните елементи на дигиталната библиотека – общности, колекции, публикации, метаданни и потоци от битове.

Удостоверяването се осъществява чрез LDAP посредством конфигуриране на DSpace инсталацията.

Изградена е йерархия от факултети и катедри, в които се намират потребителските колекции. Когато потребител се удостовери за първи път, има възможност да посочи в кой факултет и катедра ще се намира неговата колекция. Веднъж направил своя избор, потребителят се сдобива и с опция на интерфейса „Мои публикации“ (фигура 2) за бърз достъп до собствените научни материали.



Фигура 2. Моите публикации

Факултетът в системата за публикации се представя със своето описание и катедрите, които включва. Катедрата в системата за публикации се представя също със своето описание и потребителските колекции, които включва.

Потребителската колекция включва информация за своя притежател и неговите публикации.

Могат да се създават различни типове публикации (фигура 3): презентация, монография, студия, учебник, учебно помагало и други, а към всяка публикация може да се прикачат файлове в различен формат.

Публикация -

Детайли Допълнителна информация

Тип на публикацията:

Заглавие: *

Тема: *

Автор: * Издател:

Формат: Език:

Произход: Връзка с:

Обхват: Достъп до права:

Описание:

Фигура 3. Създаване на нова публикация

За описване на дигиталните ресурси се използва разширената схема на стандартът Dublin Core, Qualified Dublin Core. Освен имплементацията на стандартната схема, в системата за публикации са добавени специфични за катедра „Компютърни системи“ метаданни, обособени в отделна схема custom: custom.impact.factor, custom.impact.rank, custom.indexed.scopus, custom.indexed.webofscience, custom.publish.location, custom.report.held, custom.institution.intern

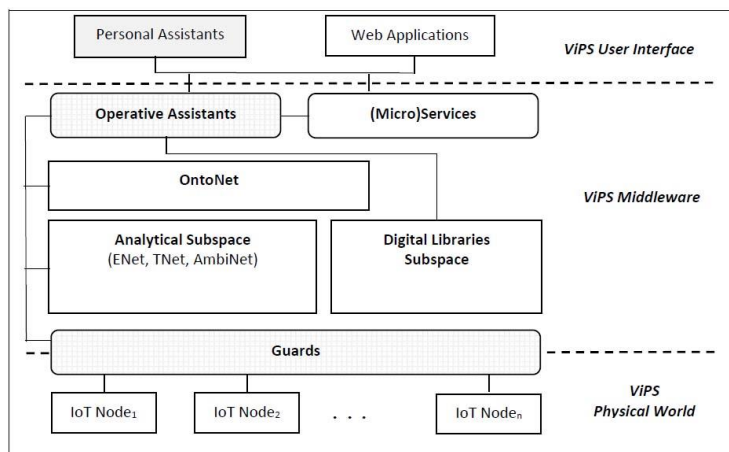
Системата дава възможност за генериране на обобщена справка за направените публикации през последните 5 години. Използва се стандартен формат, описващ броя и типа на публикациите, техните индексирания и други характеристики.

Бъдещо развитие на системата за академични публикации

Реализацията на системата за публикации е част от един по-голям проект – ViPS, който се разработва в катедра „Компютърни системи“ [13].

Референтната архитектура (фигура 4) показва структурата на едно Виртуално реално социално пространство (CPSS [10]) каквото е ViPS. Средата е изградена от няколко нива, като на най-горно ниво стоят персоналните асистенти (РА) [1] и няколко на брой публични уеб приложения. Междинното ниво е най-сложно от към комплексност на

компонентите и комуникация помежду си. Съставни части на това ниво са Оперативните агенти, микро услуги, OntoNet, Аналитичното пространство (ENet, TNet, AmbiNet) и дигитални библиотеки. Връзката между виртуалния и дигиталния свят се извършва посредством най-долното ниво и в частност агенти наречени “гардове”, които директно комуникират със сензорното ниво на физическия свят.



Фигура 4. Архитектура на ViPS

Моделирането на реални обекти спрямо тази архитектура взима под внимание фактори като събития, време, пространство и местоположение [1]. Много важна роля, за да се окаже тази архитектура успешна, играе персоналният асистент. Основният фокус на повечето подобни системи от тип CPSS пада върху потребителя. Именно тук задачата на персоналният асистент е да подпомага, да оперира от негово име към по-долните нива на системата, за да може интелигентността на отделните компоненти да достигне в най-приемлива форма до крайния потребител.

Системата за публикации е част от междинното ниво на референтната архитектура, като се използва за съхраняване на знание, учебно съдържание, публикации и документи. Идеята е системата да се разшири с интелигентни компоненти, които ще извършват интелигентни услуги и ще позволят плавната интеграция с ViPS. Предстои програмирането на интелигентни агенти посредством Jadex [15], който поддържа BDI (Belief Desire Intention) модел, които биха направили възможна интеграцията във ViPS. За комуникация между ViPS и системата за публикации ще се използват wrapper агенти и стандартни синхронни HTTP заявки с възможност за използването на асинхронна комуникация през AMQP.

Заклучение

Разработена е първа версия на системата за академични публикации. Налице са и прототипи на интелигентни компоненти за връзка със ViPS, които ще бъдат представени в следващи публикации. Реализацията ще

улесни отчитането на научната работа в катедра „Компютърни системи“. Дигиталната библиотека може да бъде използвана за съхранение и на друг вид дигитални ресурси, а не само на публикации. По тази причина може да се използва като хранилище от останалите компоненти във виртуалното пространство – например туристическия гид [14].

Благодарности

Това изследване е частично финансирано от проект ФП17-ФМИ-008 към НПД на Пловдивски университет „Паисий Хилендарски“.

Референции

- [1] S. Stoyanov, A. Stoyanova-Doycheva, T. Glushkova, E. Doychev, Virtual Physical Space – An Architecture Supporting Internet of Things Applications, *XX-th International Symposium on Electrical Apparatus and Technologies SIELA 2018* 3-6 June 2018, Bourgas, Bulgaria.
- [2] DSpace – A Turnkey Institutional Repository Application, <https://duraspace.org/dspace/>, 12.01.2019.
- [3] DSpace Rest API, A programmatic interface to DSpace Communities, Collections, Items, and Bitstreams, <https://wiki.duraspace.org/display/DSDOC6x/REST+API>, 12.01.2019.
- [4] Bratman, M., *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, 1987
- [5] Witten, I., D. Bainbridge, D. Nichols, *How to Build a Digital Library*, Department of Computer Science University of Waikato New Zealand.
- [6] Callan, P., and M. Gregson, Implementing an enhanced repository statistics system for QUT ePrints. Poster presented at the annual meeting of the *International Conference on Open Repositories*, July 2012, Edinburgh, UK
- [7] Witt, M., and M. Newton, *Preparing batch deposits for Digital Commons repositories*, 2008.
- [8] Shin, E., and A. Woods, *Building the future of Fedora*, 2013.
- [9] Dublin Core Metadata Initiative, <http://dublincore.org/>, 12.01.2019.
- [10] Sheth, A., P. Anantharam, and C. Henson, Physical-Cyber-Social Computing: An Early 21st Century Approach, *IEEE Intelligent Systems*, vol. 28, no. 1, Feb. 2013, 1541–1672.
- [11] The Open Archive Initiative Protocol for Meta Harvesting. Retrieved January, 2019, <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [12] DSpace Working Group, Retrieved January, 2019, from <https://wiki.duraspace.org/display/DSPACE/DSpace+7+Working+Group>

- [13] Stoyanov, S., A. Stoyanova-Doycheva, T. Glushkova, E. Doychev, Virtual Physical Space – An Architecture Supporting Internet of Things Applications, *20th International Symposium on Electrical Apparatus and Technologies (SIELA)*, 3-6 June 2018, DOI: 10.1109/SIELA.2018.8447156
- [14] Glushkova, T., M. Miteva, A. Stoyanova-Doycheva, V. Ivanova, S. Stoyanov, Implementation of a Personal Internet of Thing Tourist Guide, *American Journal of Computation, Communication and Control*, Publication Date: May 10, 2018, Pages: 39-51, ISSN 2375-3943
- [15] Jadex Active Components, Retrieved January, 2019, from <https://download.actoron.com/docs/releases/latest/jadex-mkdocs/getting-started/getting-started>
- [16] Metadata Encoding and Transmission Standard, Retrieved January, 2019, from <https://www.loc.gov/standards/mets/>
- [17] Machine-Readable Cataloging, Retrieved January, 2019, from <https://www.loc.gov/marc/umb/>
- [18] Drupal, Retrieved January, 2019, from <https://www.loc.gov/marc/umb/>

Факултет по математика и информатика

Пловдивски университет „Паисий Хилендарски“

бул. „България“ № 236, 4003 Пловдив, България

e-mails: gig87@abv.bg, dimitarhristov@yahoo.com,
astoyanova@uni-plovdiv.net

ACADEMIC PUBLICATION SYSTEM

Georgi Georgiev, Dimitar Hristov, Asya Stoyanova-Doycheva

Abstract: This article describes and analyzes problems and tasks related to the storage and processing of data in ViPS (Virtual Physical Space) [1]. Existing storage platforms, in particular DSpace [2], and the possibilities of their application, are being explored not only as a publication tool giving greater visibility and accessibility of scientific materials but also in a way that allows the integration of a digital library in ViPS and its interaction with the other components of the space. A functional Web-based academic publishing system created by implementing DSpace REST API [3], which administers faculties and departments, allows creating user collections as well as creating, editing, and deleting publications in a collection. The application is the basis for upgrading with future intelligent functionalities by using the BDI (belief-desire-intention) [4] model and the step towards creating an intelligent heterogeneous data repository.