

ДИГИТАЛНЕ БИБЛИОТЕКЕ – РАЗГРАНИЧЕЊЕ ПОЈМОВА*

Цветана Крстев

Катедра за библиотекарство и информатику
Филолошки факултет, Београд

Сажетак

Рад покушава да смести истраживања у области дигиталних библиотека првенствено у информатичка истраживања, и то, пре свега, истраживања у области формата за представљање информација, савремених система за управљање базама података и проналажења информација. Истакнут је такође и значај стандардизације у овој области, почев од стандардизације формата за представљање информација и метаподатака па до стандардизације протокола. У раду је посебан акценат стављен на значај истраживања везана за анализу садржаја информација која обухватају области обраде природних језике, анализе слика и препознавања говора. Посебно је указана потреба за заштитом интелектуалне својине кроз развој хардверских и софтверских система заштите. У раду су представљени и пионирски пројекти у овој области који су отворили пут многим другим пројектима.

Кључне речи: дигиталне библиотеке, проналажење информација, анализа садржаја, дистрибуирани системи, здружено претраживање

DIGITAL LIBRARIES – DEFINITION OF CONCEPTS

Cvetana Krstev

Library and Information Science Department
Faculty of Philology, Belgrade

Abstract

In this paper we try to locate the research related to digital libraries in the computer science fields, such as definition of formats for information representation, development of modern database management systems and information retrieval. The importance of standardization in this field is stressed, from standardization of formats for representation of information and metadata to protocol standardization. The importance of research related to content analysis, including natural language processing, image analysis and speech recognition is particularly emphasized. This paper indicates the necessity of intellectual right protection through the development of software and hardware protection systems. The pioneer projects in this field are as well presented those that have traced the path for numerous other projects.

Key words: digital libraries, information retrieval, content analysis, distributed systems, federated search

1. ДЕФИНИЦИЈЕ

Премда је појам дигиталних библиотека ушао у употребу у деведесетим годинама он не губи на актуелности. Дигиталне библиотеке су и даље у жижи информатичких интересовања, о чему говоре многе конференције, часописи и бројни значајни национални и међународни пројекти који се посвећују проблемима њихове израде, коришћења и улоге. И поред свог несумњивог присуства у жижи интересовања истраживача и корисника током дугог времена, сам појам дигиталне библиотеке још увек није прецизно омеђен: дефинисању овог појма и даље су посвећене дискусије и округли столови на конференцијама о дигиталним библиотекама.

Може се поставити питање одакле потиче ова немогућност да се појам дигиталне библиотеке, на први поглед прихватљив и лаицима, прецизније одреди. Један могући одговор на то питање лежи можда у чињеници да дигиталне библиотеке представљају, у неку руку, остварење различитих визија, како из дела научне-фантастике, тако и из озбиљне научне литературе о свеобухватном бележењу људског знања и његовој широкој доступности. Овај амбициозни циљ се нужно мора посматрати из различитих перспектива, а његово

* Рад је изложен на Стручном скупу *Информациона технологија и библиотекарство*, одржаном у оквиру Седме скупштине Заједнице библиотека универзитета у Србији, 1. јуна 2001. године на Технолошко-металуршком факултету у Београду.

остварење мора се поверити истраживачима из различитих области. Поменимо само неке од релевантних области: управљање базама података, мултимедијални информациони системи, проналажење информација, обрада природног језика, интеракција између човека и рачунара, подручја мреже и комуникације и библиотекарство.

Можда се појам дигиталних библиотека може лакше описати, моделирати, дизајнирати, имплементирати, користити и евалуирати ако се сагледа у светлу следећих пет апстрактних концепата које су предложили Фокс и Сорнила са *Virginia Polytechnic Institute and State University* [5]:

токови, појам који се користи да опише текст, мултимедијални садржај, и друге секвенције апстрактних појмова: на пример, описе протокола, интерактивне дијалоге, дневнике сервера (*server logs*) и дискусионе листе.

структуре, које се односе на структуре података у информатичком смислу, базе података, хипертекстуалне мреже и друге концепте из области претраживања информација, као што су инвертоване датотеке, датотеке сигнатура, MARC слогови и тезауруси.

простори покривају не само дводимензионалне и тродимензионалне просторе, виртуелну реалност већ и друге вишедимензионалне форме које као једну компоненту укључују и време;

сценарија која обухватају дизајнирање и спецификацију интеракције између човека и рачунара, али такође описују процесе, процедуре, функције, сервисе и трансформације укључене у дигиталне библиотеке – једном речју аспекте њиховог активног живота и временског трајања.

друштва која покривају оне аспекте дигиталних библиотека који су од значаја за ауторе, библиотекаре, каталогизаторе и друге чуваре информација. Ови аспекти су посебно значајни у последње време од када се и антрополози, комуниколози, психолози и социолози укључују у истраживања везана за дигиталне библиотеке. Понуђене дефиниције дигиталних библиотека треба процењивати према њиховом сагледавању ових пет аспеката дигиталних библиотека. Међу дефиницијама има врло кратких, какав је рецимо поднаслов специјалног броја престижног рачунарског часописа *ACM Communications* из 1998 који је посвећен дигиталним библиотекама "Дигиталне библиотеке – глобални опсег, неограничени приступ", па до оних знатно развијених, какав је дефиниција Бејкера из 1996 [2]:

Дигиталне библиотеке конструишу, односно прикупљају и организују, заједнице корисника. Функционалне могућности дигиталних библиотека подржавају информационе потребе заједнице. Дигиталне библиотеке су проширење, надградња и интеграција разноврсних информатичких институција, као и физичких простора, у којима се ресурси селекују, прикупљају, организују, чувају или им се обезбеђује приступ као подршка заједници корисника.

За све оне кориснике који су укључени у процесе проналажење информација, дигиталне библиотеке представљају проширење таквих система које је обогато мултимедијалном разноврсношћу и удруженим системима (енгл. *federated systems*). У сваком случају, дигиталне библиотеке морају да подржавају велике колекције докумената, као и могућности њиховог претраживања, каталогизирања и индексирања. Истраживања у области дигиталних библиотека се, стога, у значајној мери преклапају са текућим истраживањима у области проналажења информација, и она обухватају вишејезичку обраду, претраживање мултимедијалног садржаја, визуелизацију информација, руковање великим дистрибуираним колекцијама докумената, стандарде и архитектуре, те ће о свим овим аспектима бити укратко речи.

2. АРХИТЕКТУРЕ

Дигиталне библиотеке могу бити део инфраструктуре Интернета обезбеђујући различите сервисе у склопу система који су намењени другим задацима. Али, с друге стране, оне могу бити и независни системи и у том случају, да би уопште биле изграђене, морају имати независну архитектуру. Заправо, многе данашње дигиталне библиотеке се склапају од постојећих делова: машина за претраживање, прелистача мреже, система за управљање базама података и алатки за руковање мултимедијалним документима.

Једна предложена архитектура дигиталне библиотеке је колекција дигиталних објеката, од којих сваки има садржај, полугу за повлачење (енгл. *handle*), која је неко име или идентификатор, а може такође имати својства, сигнатуру и дневник обављених трансакција. Дигиталним објектима се придружују и метаподаци. Спремишта дигиталних објеката обезбеђују њихову сигурност и протоколе за приступ. Што се тиче

остварење мора се поверити истраживачима из различитих области. Поменимо само неке од релевантних области: управљање базама података, мултимедијални информациони системи, проналажење информација, обрада природног језика, интеракција између човека и рачунара, подручја мреже и комуникације и библиотекарство.

Можда се појам дигиталних библиотека може лакше описати, моделирати, дизајнирати, имплементирати, користити и евалуирати ако се сагледа у светлу следећих пет апстрактних концепата које су предложили Фокс и Сорнила са *Virginia Polytechnic Institute and State University* [5]:

токови, појам који се користи да опише текст, мултимедијални садржај, и друге секвенције апстрактних појмова: на пример, описе протокола, интерактивне дијалоге, дневнике сервера (*server logs*) и дискусионе листе.

структуре, које се односе на структуре података у информатичком смислу, базе података, хипертекстуалне мреже и друге концепте из области претраживања информација, као што су инвертоване датотеке, датотеке сигнатура, MARC слогови и тезауруси.

простори покривају не само дводимензионалне и тродимензионалне просторе, виртуелну реалност већ и друге вишедимензионалне форме које као једну компоненту укључују и време;

сценарија која обухватају дизајнирање и спецификацију интеракције између човека и рачунара, али такође описују процесе, процедуре, функције, сервисе и трансформације укључене у дигиталне библиотеке – једном речју аспекте њиховог активног живота и временског трајања.

друштва која покривају оне аспекте дигиталних библиотека који су од значаја за ауторе, библиотекаре, каталогизаторе и друге чуваре информација. Ови аспекти су посебно значајни у последње време од када се и антрополози, комуниколози, психолози и социолози укључују у истраживања везана за дигиталне библиотеке. Понуђене дефиниције дигиталних библиотека треба процењивати према њиховом сагледавању ових пет аспеката дигиталних библиотека. Међу дефиницијама има врло кратких, какав је рецимо поднаслов специјалног броја престижног рачунарског часописа *ACM Communications* из 1998 који је посвећен дигиталним библиотекама "Дигиталне библиотеке – глобални опсег, неограничени приступ", па до оних знатно развијених, какава је дефиниција Бејкера из 1996 [2]:

Дигиталне библиотеке конструишу, односно прикупљају и организују, заједнице корисника. Функционалне могућности дигиталних библиотека подржавају информационе потребе заједнице. Дигиталне библиотеке су проширење, надградња и интеграција разноврсних информатичких институција, као и физичких простора, у којима се ресурси селекују, прикупљају, организују, чувају или им се обезбеђује приступ као подршка заједници корисника.

За све оне кориснике који су укључени у процесе проналажење информација, дигиталне библиотеке представљају проширење таквих система које је обогато мултимедијалном разноврсношћу и удруженим системима (енгл. *federated systems*). У сваком случају, дигиталне библиотеке морају да подржавају велике колекције докумената, као и могућности њиховог претраживања, каталогизирања и индексирања. Истраживања у области дигиталних библиотека се, стога, у значајној мери преклапају са текућим истраживањима у области проналажења информација, и она обухватају вишејезичку обраду, претраживање мултимедијалног садржаја, визуелизацију информација, руковање великим дистрибуираним колекцијама докумената, стандарде и архитектуре, те ће о свим овим аспектима бити укратко речи.

2. АРХИТЕКТУРЕ

Дигиталне библиотеке могу бити део инфраструктуре Интернета обезбеђујући различите сервисе у склопу система који су намењени другим задацима. Али, с друге стране, оне могу бити и независни системи и у том случају, да би уопште биле изграђене, морају имати независну архитектуру. Заправо, многе данашње дигиталне библиотеке се склапају од постојећих делова: машина за претраживање, прелистача мреже, система за управљање базама података и алатки за руковање мултимедијалним документима.

Једна предложена архитектура дигиталне библиотеке је колекција дигиталних објеката, од којих сваки има садржај, полугу за повлачење (енгл. *handle*), која је неко име или идентификатор, а може такође имати својства, сигнатуру и дневник обављених трансакција. Дигиталним објектима се придружују и метаподаци. Спремишта дигиталних објеката обезбеђују њихову сигурност и протоколе за приступ. Што се тиче

идентификатора дигиталних објеката, једну шему је предложио OCLC (Online Computer Library Center, Inc.) у облику постојаних URL, скраћено PURL (*Persistent URL – PURL*) (www.purl.org). Ова врста локатора представља врсту надградње над постојећим URL којом се жели избећи претерана везаност URL-а за хардвер. Наиме, PURL је URL али уместо да показује директно локацију ресурса на Интернету, он показује адресу једног уметнутог сервиса – purl.oclc.org – који пресликава PURL у физичку адресу URL коју враћа клијенту који је испоставио захтев [27].

Велики пројекти дигиталних библиотека посвећују пажњу управо питањима њихове архитектуре. Главна брига Стенфордског пројекта предузетог у оквиру Подстицаја дигиталним библиотекама (*Digital Library Initiative*) у периоду од 1994-1998 је интероперабилност [13]. Овим пројектом је предвиђено да се различити информатички ресурси повезују посредством одговарајућих посредника, а затим да се користе преко дељеног буса, кроз разноврсне интерфејсе. Ти информатички ресурси, дакле, не би морали да буду у истом формату као што је случај са вебom где сва документа користе исти HTML формат.

3. МОДЕЛИ ДОКУМЕНАТА, РЕПРЕЗЕНТАЦИЈЕ И ПРИСТУП

Документа представљају саму срж како проналажења информација тако и дигиталних библиотека. С обзиром на ефикасност, посебно када се ради о руковању милијардама докумената у простору који се мери гигабајтима, терабајтима или петабајтима, компресија докумената је од суштинског значаја. Премда су на овом подручју остварени значајни резултати, конвертовање великог броја докумената коришћењем висококвалитетних репрезентација, на пример, PDF, DVI, PC, може да постане недозвољиво скупо, посебно имајући у виду цену претраживања, осим ако пронађена јединица није изузетно популарна. Ово говори да се стварном коришћењу докумената и њиховом животном веку мора посветити изузетна пажња.

Следећа карактеристика докумената која се мора узети у обзир је њихов, и даље најчешће, природнојезички садржај. Премда постоји доста разлога да се енглески језик прихвати као језик глобалне комуникације, посебно у области науке и технике, предвиђа се да ће управо развој дигиталних библиотека увећати присуство и других језика на Интернету. Многе државе или институције ће пожелети да преко дигиталних библиотека промовишу своје културно наслеђе и богатство. Европска Унија, на пример, подстиче равноправно коришћење језика свих земаља чланица. Коришћењем могућности дигиталних библиотека, а пре свега паралелног, или здруженог, претраживања, требало би да се повећа и приступ таквим документима, јер корисници желе да добију информације о одређеној теми из свих расположивих извора, независно од језика.

У основи проблема вишејезичности лежи проблем кодирања на нивоу карактера, који се релативно успешно решава 16-битним Unicode-ом. Његова примена у пракси је још увек далеко од задовољавајуће, па се често као алтернативно решење нуди могућност превлачења, по потреби, фонтова са сервера или вратница. Ова могућност задовољава, наравно, само потребе визуелизације јединица дигиталне библиотеке. Решавање проблема претраживања вишејезичких колекција је много захтевније. Превођење упита је у пракси практичније од превођења докумената, посебно ако дигитална колекција покрива више језика [17]. За превођење упита могу да се користе посебни речници, тезауруси или семантичке мреже [15] за лоцирање речи или фраза у документима на другом језику. О значају овог питања говори и чињеница да су Европска Унија и NSF (*National Science Foundation*) САД-а 1997. године оформили радну групу за вишејезични приступ информацијама у дигиталним библиотекама с циљем да се реше питања складиштења, приступа и приказа информација на било ком светском језику [11].

Следећи аспект садржаја докумената је његов мултимедијални карактер. Вишеструки токови мултимедијалних докумената морају се на неки начин синхронизовати. Ово је предмет новог мрежног стандарда SMIL (*Standard Multimedia Integration Language*) (www.w3.org/AudioVideo/). SMIL је једна реализација XML која дозвољава интеграцију скупа независних мултимедијалних објеката у синхронизовану мултимедијалну презентацију. С друге стране, и на мултимедијални садржај се могу применити многе технике из области проналажења информација. Ту, пре свега, мислимо на коришћење визуелних упита, на анализу визуелног садржаја и на приказ добијеног резултата у контексту [3]. Пример садржаја су боја, текстура, облик, позиција, доминантне ивице у слици, региони, а пример визуелног упита у области медицине био би: “Пронађи друге слике које имају тумор са текстуром сличном овој” [4]. Овакви упити наликују на упите у базама података пуног текста, то јест, изводе се без корисника.

Истакнути пример је пројекат QVIC (*Query By Image Content*) развијан у фирми IBM у оквиру кога се истражују методе упита великих база података слика у чијој је основи садржај слика (www.qbic.almaden.ibm.com). На QVIC систему се заснива веб-страница музеја Ермитаж која се сматра једном од најбољих у Русији. Корисник може, рецимо, да пронађе слике из ове чувене колекције у којима преовлађују боје по његовом избору.

Оваква врста истраживања је посебно занимљива за оне колекције које су настале дигитализацијом старих књига, рукописа или слика. За њих је од првенственог значаја да се оствари такав квалитет репродукције који би био довољно добар за истраживачке потребе али не и за комерцијалну употребу или пиратерију. Као вид заштите дигитализованог материјала данас се користе и врсте водених жигова који ефикасно одвраћају од недозвољене употребе. Пример овакве дигиталне библиотеке су колекције уметничког и историјског материјала које је развио IBM [8]. Међу њима су колекција *Bibliotheca Vaticana Apostolica* из Ватикана и *Archivo General de Indias* из Севиле.

Када се говори о документима питање њихове структуре је незаобилазно. У ствари, структурирана документа представљају токове над које је надграђена једна или више структура, обично коришћењем ознака уметнутих у ток, а понекад и коришћењем одвојене спољашње структуре налик хипертекстуалним показивачима. Структура се обично надграђује над документа када је потребно побољшати њихову вредност уграђивањем одређеног знања. За описивање ове структуре најчешће се користи SGML, или у новије време XML, јер већина докумената спада у неку препознату логичку структуру за коју је већ развијена дефиниција типа документа DTD (*Document Type Definition*) [12]. Колекција DTD-а за књижевне и друге текстове из области друштвених наука развијена је у оквиру пројекта Подстицаја кодирању текста TEI (*Text Encoding Initiative*) (www.tei-c.org) [22]. Важно је такође истаћи да се метаподаци који по правилу прате дигиталне објекте у колекцији и који се обично бележе у UNIMARC формату могу такође представити и у облику SGML/XML документа [7], и то било независно било у саставу оригиналног документа (xml.coverpages.org/marc.html).

Структура је веома значајна и са становишта претраживања. Претраживање и по структури и по садржају доноси изузетну предност системима за проналажење информација какав је OpenText, или PAT (etext.lib.virginia.edu/pat.html) [9]. Пример оваквог претраживања у речничкој бази података је упит: “Пронађи све застареле термине из физике који су у употребу ушли између 1860. и 1930. године, а потичу из грчког или се завршавају на *-isation*”. У овом упиту наведене године, затим “физика”, “грчки” и суфикс су кључеви за претраживање садржаја, док се етимологија и датум најранијег појављивања реферишу на структуру речничке одреднице која омеђује упит по садржају.

За дигиталне библиотеке су важне и друге врсте структуре, као анотације које укључују аспект сценарија јер су оне резултат интеракције корисника са дигиталном колекцијом.

Премда термин дигиталне библиотеке обухвата и сасвим мале или ограничене приватне колекције које подржава одређени систем и које нуде прописане сервисе, већина дигиталних библиотека се најчешће простире преко више рачунара који се шире у физичком и логичком простору. На тај начин, руковање колекцијама које су по својој природи дистрибуиране је предуслов развоја технологије дигиталних библиотека.

Од посебног је значаја рад са великим бројем дигиталних библиотека које су независно конструисане тако да су њихови информатички системи ефективно хетерогени. Једна могућност експлоатације је да се смањи функционалност за крајњег корисника чиме се ауторима дигиталних библиотека даје већа слобода у дизајнирању. Интеграција се овде своди на проналажење заједничког садржаоца, што је најчешће једноставна претрага, система који подржавају различите комуникационе протоколе, упитне језике, подржавају разноврсне токове и структуре, или их комбинују на различите начине с аспекта репрезентације и метаподатака. На другом крају је могућност повећавања функционалности здружених система интелигентнијим коришћењем рачунарских ресурса сервера и клијента. Примена ове могућности претпоставља развој описног језика за сваку дигиталну библиотеку и развој удруженог система који би могао да га разуме.

Као трећа могућност појављује се коришћење протокола Z39.50. Ово је ознака америчког националног стандардног протокола за проналажење информација који је 1997. усвојен и као међународни стандард под ознаком ISO 23950 (www.ukoln.ac.uk/dlils/z3950). Овај стандард који је потекао из библиотечког окружења из потребе да се повежу разноврсни библиотечки каталози, дозвољава једном клијенту да комуницира за једним или више система за проналажење, или једном систему за проналажење да комуницира са другим сличним

системима. Да би ова могућност могла да се користити у контексту дигиталних библиотека потребно је да се обезбеди подршка Z39.50 протоколу и на страни сервера и на страни клијента.

Истакнути пример овакве примене пружа пројекат CIMI (*Computer Interchange of Museum Information*) [16]. У оквиру овог пројекта користи се Z39.50 протокол да би се омогућио једнообразан приступ постојећим и настајућим дигиталним колекцијама као и великим спремиштима информатичких ресурса културног наслеђа. Ови ресурси обухватају физичке предмете и из њих изведене дигиталне објекте, описне записе начињене ради управљања колекцијом, библиографске записе, документе пуног текста, *online* алате какви су тезауруси или нормативне листе имена уметника и слично. У оквиру овог пројекта развијена је снажна Z39.50 клијент/сервер апликација надграђена над постојећим музејским системима и базама података. То омогућава кориснику да један упит, који се заснива на приступним тачкама везаним за музејске информације какве су назив објекта, име уметника, материјал, порекло, и слично, проследи ка више сервера, а резултати претраге на њима враћају се кориснику.

Многе апликације захтевају здружено претраживање (енгл. *federated search*) које подразумева прикупљање тражених информација обиласком расположивих ресурса, затим интелигентан избор најбољих ресурса за тражење и на крају спајање и приказивање резултата добијених из различитих извора. Један од првих пројеката у овом смеру је иницијатива *Wide Area TEchnical Reprot Service (WATERS)* коју је у САД подржала NSF чији је циљ омогућавање приступа техничким извештајима из области рачунарства из целог света које би се остварило с минимумом улагања и с тежиштем на локалној контроли [6].

Ако дигитална библиотека садржи колекције врло вредних слика, или ако су њене колекције велике и добро организоване те ако нуде корисне сервисе за манипулацију информацијама, какви су претрага, уређивање, извештавање, апстраховање, и слично, онда се обично подразумева неки начин плаћања за приступ њеним фондовима. Иако се раније приступ научним информацијама није сматрао робом као што је данас случај, дигиталне библиотеке морају на неки начин да управљају интелектуалном својином. Оне морају имати сервисе које подржавају ауторска права и поштују договоре, уговоре и законе. Премда техничка страна чини само један аспект овог проблема, кључно је обезбедити поузданост система. Проналажење информација се није никада претерано бавило питањем сигурности, али она је од фундаменталног значаја за дигиталне библиотеке. Као што је криптовање суштинско за електронску трговину, тако је важна примена дигиталних “водених жигова”, хардверских брава, специфичног интерфејса [13] и дугих, јачих, механизма заштите да би дигиталне библиотеке заштитиле интелектуалну својину и контролисале начин приступа за разне врсте корисника.

4. ПРОТОТИПОВИ И ПРОЈЕКТИ

Већ годинама су истраживања на подручјима проналажење информација, хипертекста, мултимедијалних документа и библиотечке аутоматизације врло жива. Изградња првих аутентичних прототипова дигиталних библиотека везује се за пројекте ENVISION и CORE, започетих 1991. године. Дигитална библиотека настала из пројекта ENVISION заснива се на колекцији докумената из области рачунарства и у њој су нове технике визуелизације укључене у постојећи систем за проналажење. У оквиру пројекта CORE изграђен је прототип дигиталне библиотеке заснован на колекцији часописа из области хемије. Колекција садржи и дигитализовану слику и SGML означену верзију сваког чланка. У оквиру пројекта је осим саме колекције израђен и тестиран разноврстан кориснички интерфејс.

Најзначајнији пројекат посвећен дигиталним библиотекама у САД је Подстицај дигиталним библиотекама DLI (*Digital Library Initiative*), чија се прва фаза одвијала у периоду од 1994-1998. године [19, 13, 5]. Овај пројекат су финансирале NSF (*National Science Foundation*), DARPA (*Defense Advanced Research Projects Agency*) и NASA (*National Aeronautics and Space Administration*). Прва фаза покрива врло широк опсег тема, а превасходно оне везане за индексирање, здруживање и решавање семантичких проблема. Пројекат је био подељен у шест потпројеката, а за сваки од њих био је задужен по један универзитетски центар. То су:

- Потпројекат *University of Illinois at Urbana-Champaign* који је био оријентисан ка изградњи дигиталне библиотеке са много корисника (10.000) која се заснива на здруженим разноврсним колекцијама научних докумената (100.000). Сваки документ у колекцији је представљен пуним текстом који је ручно SGML означен. Највише напора је уложено у дефинисање семантичког простора [20].

- Пројекат *University of California at Berkley* је био оријентисан ка развоју нове технологије за дигиталне библиотеке које се заснивају на великим колекцијама (300.000 страница, 560.000 фотографија из области животне средине, као и 1Гб података из исте области). Нове технологије су обухватале упите на основу садржаја слика, екстракцију база података из докумената, мултивалентна документа и природно-језичку обраду (лексичку дезамбигуацију) [28].
- Пројекат *Carnegie Mellon University* је истраживао могућности да се манипулише новим медијима, прецизније говорећи, изградњу система са интелигентним приступом аудио и видео материјалу [25]. Пројекат се заснива на дигиталном видео запису који се састоји од 1000 сати видео траке једне јавне телевизијске станице. У оквиру пројекта су развијане различите технологије које треба да обезбеде приступ оваквом материјалу, као што су анализа слика, препознавање говора, препознавање лица и природно језичка обрада (препознавање имена људи и њихово повезивање са сликом).
- Главни садржај дигиталне библиотеке која је развијена у оквиру пројекта *University of California at Santa Barbara* није текст. Овај пројекат се концентрисао на индексирање мапа коришћењем техника обраде слика. Технологије које су развијане за ове потребе су просторно индексирање и проналажење и обрада слике. Као прототип је изграђена дистрибуирана дигитална библиотека *Alexandria* (www.alexandria.ucsb.edu) за материјале који се реферишу преко географских појмова, какви су мапе, сателитски снимци, дигитализоване фотографије из ваздуха и слично.
- Инфраструктурни пројекат *Stanford University* се бавио проблемима интероперабилности [18]. Иако је развијена и дигитална библиотека која се састоји од колекције докумената из области рачунарства главни напор је уложен у развој оних аспеката мрежне технологије и система за управљање базама података који су од значаја за развој дигиталних библиотека. За имплементирање приступа информацијама и протокола плаћања, користили су протокол CORBA који дозвољава да компоненте система буду написане на различитим језицима и за различите рачунарске платформе.
- Пројекат *University of Michigan* се концентрисао на изградњу инфраструктуре за пружање библиотечких сервиса преко мреже [1]. Циљ пројекта није била само изградња ових сервиса већ, пре свега, обезбеђивање услова за проширивање дигиталних библиотека новим колекцијама и сервисима према потреби. Други циљ пројекта је образовање кроз развој “Јавне Интернет библиотеке” и система који је намењен школама. Овај је кооперирао и са другим сличним пројектима, какав је JSTOR чији је циљ стварање дигиталне библиотеке пре свега фреквентно коришћених научних часописа и у дигиталном и у текстуалном облику која би се могла претраживати и по формалним атрибутима и по садржају [10].

За време трајања прве фазе пројекта DLI активности и интересовања везана за дигиталне библиотеке су значајно нарасле, без сумње и под његовим утицајем јер су резултати рада на овом пројекту били видљиви кроз презентације на вебу, чланке у часописима и радове на конференцијама. Кроз многе дискусије о резултатима овог и других пројеката је установљено да је за даљи напредак у развоју дигиталних библиотека важно укључивање стручњака и из других области, а не само из рачунарства и информатике.

Године 1998. је отпочео рад на другој фази пројекта Подстицај дигиталним библиотекама, познат као DLI-2, кога подржавају осим NSF, DARPA и NASA, и *National Library of Medicine, Library of Congress, National Endowment for the Humanities* и многе друге институције (www.dli2.nsf.gov). У овој фази пројекат је мање технички оријентисан него раније и подржава разна истраживања везана за животни циклус информација почев од креирања садржаја, приступа и коришћења па до конзервирања и архивирања чиме се људски фактор укључује као битан за развој дигиталних библиотека.

Осим ових пројеката, у САД-у је рађено, а и даље се ради на великом броју пројеката из ове области. Поменимо само неке. Пројекат Конгресне библиотеке из Вашингтона који предвиђа дигитализацију око 5 милиона јединица из њихове колекције. Веома је значајан и пројекат “Умрежене дигиталне библиотеке теза и дисертација” (*Networked Digital Library of Theses and Dissertations – NDLTD*) чији је циљ побољшање квалитета високог образовања кроз ширење могућности за дељење знања [23, 24]. За информације о осталим пројектима треба консултовати неки од пописа на Вебу, какав је каталог електронских текстова у области хуманистичких наука - CETH (www.ceth.rutgers.edu). Многе информације нуди и матична страница познатог стручњака у овој области Едварда Фокса (ei.cs.vt.edu/~h3004fox/Overview.html).

Поред ових активности које су се одвијале и данас се одвијају у САД-у, и у многим другим државама су програми изградње дигиталних библиотека започети или се припремају. Врло су живе активности на Новом

Зеланду (Пројекат јавних библиотека заснован на проналажењу пуног текста), Немачкој (Пројекат *MeDoc* чији је циљ стимулисање коришћења електронских медија у академском образовању и у истраживању), Мађарској (Пројекат формирања и управљања глобалном дистрибуираном дигиталном библиотеком у Централној Европи која покрива област рачунарства), Бразилу (Пројекат организовања колекције из области заштите животне средине), Кореји (Пројекат изградње дигиталне библиотеке опште намене).

У Европи, најживље активности су у Великој Британији која је покренула пројекат *eLib - Electronic Libraries Programme*. За разлику од пројекта DLI, средства из овог пројекта се усмеравају на велики број релативно малих пројеката чији је циљ да се за потребе високог образовања у Великој Британији оформи тело видљивих електронских ресурса и сервиса. Међу њима су пројекат дигитализације колекције новина штампаних пре 1800. године и пројекат “Кантерберијске приче” (*Canterbury Tales*) који обухвата сканирање свих рукописа и штампаних издања овог познатог Чосеровог дела која су настала пре 1500. године и која укључује и транскрипције и коментаре (www.cta.dmu.ac.uk/projects/ctp/). У Француској су активности на овом пољу такође веома живе. Може се рећи да је једна од првих дигиталних библиотека настала управо у Француској укуцавањем око 2000 дела француске класичне литературе са циљем стварања корпуса који прати историјски развој француског језика. Тако је настао Информатички трезор француског језика - *Le Trésor de la Langue Française Informatisé* (atilf.inalf.fr/). Данас Француска национална библиотека предузима многе пројекте дигитализације и формирања дигиталних библиотека.

5. СТАНДАРДИ - ПРОТОКОЛИ И МЕТАПОДАЦИ

Пошто је у току изградња многих дигиталних библиотека широм света који се заснивају на различитим истраживачким, развојним и комерцијалним приступима, коришћење стандарда је неопходно да би се остварила интероперабилност и размена података. Срж подршке здруженим дигиталним библиотекама је договор о протоколима за комуникацију између рачунара.

У овом домену најзначајнији је већ поменути протокол Z39.50 који је настао 80-тих година као одговор на раст броја библиотечких каталога у машински читљивом облику и потребе за успостављањем међубиблиотечке позајмице кроз удаљено претраживање каталога. Овај протокол је брзо ушао у интензивну употребу: поменимо на њему заснован систем WAIS који је био врло популаран пре настанка веба. У употреби су и неки мање захтевни протоколи који су специјално израђени да подрже здружено претраживање дистрибуираних дигиталних библиотека, на пример *Dienst* (www.cs.cornell.edu/cdlrg/dienst/DienstOverview.htm). Недостатак примена свих ових протокола је што се и сервер и клијент морају прилагодити њиховој примени.

Премда се метаподаци могу користити не само за опис докумената већ и за опис колекција, целих дигиталних библиотека и њихових сервиса, у пракси се они најчешће користе за опис дигиталних објеката па имају сличну улогу као и записи библиотечких каталога. Отуда је кодирање у MARC формату често полазна тачка за опис метаподатака. Проблем с MARC-ом је што, с једне стране, постоје многе његове верзије, а с друге стране, он подразумева рад с бинарним слоговима који нису, као такви, читљиви. Други начини кодирања, какав нуди SGML/XML су у том погледу погоднији. Тако је у 2000. године у Француској настао BiblioML као XML апликација намењена репрезентовању библиографских података ([BiblioML, xml.coverpages.org/biblioML.html](http://BiblioML.xml.coverpages.org/biblioML.html)). Коришћењем XML-а могу се имплементирати и друге шеме за опис метаподатака, каква је Даблинско језгро (*Dublin Core*, dublincore.org), шема кодирања настала у оквиру OCLC [26]. Даблинско језгро је једноставна широкоприхваћена шема која има само 15 језгровних елемената који се могу користити за опис дигиталних објеката. Међу језгровним елементима 7 служи за опис садржаја (*Title, Subject, Description, Source, Language, Relation, Coverage*), 4 за опис интелектуалне својине (*Creator, Publisher, Contributor, Rights*) и 4 за рад са апстрактним дигиталним објектима (*Data, Type, Format, Identifier*). Треба истаћи и да постоји могућност превођења из MARC-а у Даблинско језгро и обрнуто, као и пресликавање из Даблинског језгра у Z39.50 протокол.

С једне стране су традиционалне библиотеке са крутим системом каталогизирања, а с друге стране је WWW који у највећој мери игнорише проблем каталогизирања и метаподатака. WWW конзорцијум је увидео недостатке овакве политике 1996. године је отпочео рад на оквиру за опис ресурса *Resource Description Framework – RDF* (www.w3.org/RDF/). RDF представља скуп конвенција за опис метаподатака који користи XML и који обезбеђује оквир за размену разноврсних метаподатака [14]. RDF се може користити као заглавље

унутар дигиталних објеката SGML/HTML/XML кодираних, или, алтернативно, може представљати опис ускладиштен независно од њих.

6. ЗАКЉУЧАК

С великим поуздањем се може рећи да истраживања и пројектовања у области дигиталних библиотека још увек нису достигла зенит. Многа питања, посебно она везана за стандардизацију, потпуније сагледавање семантике објеката и квалитетније услуживање корисника још увек остају отворена. Врло жива истраживања у области анализе садржаја (природни језик, слика, звук) остају веома значајна за даљи развој дигиталних библиотека.

На основу досадашњих искустава у свету с правом се може рећи да су истраживања у области дигиталних библиотека, пре свега, истраживања у данас најактуелнијим информатичким областима. Искуства ипак показују да је за њихов пуни успех и укључивање у разне форме људске делатности, пре свега образовање и истраживање, у њихову изградњу потребно укључити и стручњаке разних друштвених профила, укључујући и библиотекарe.

Досадашња искуства, такође, показују да највећи број дигиталних библиотека настаје интеграцијом и надградњом над постојећим ресурсима, то јест над ресурсима који су већ у дигиталном запису или су на одговарајући начин описани записима у машински читљивом облику. Чињеница да се у том случају дигиталне библиотеке не морају градити из основа свакако олакшава и убрзава њихову изградњу, а коначан резултат чини квалитетнијим. У противном, први корак у изградњи дигиталне библиотеке је прикупљање дигиталних објеката за шта је потребно, пре свега знање, а затим време и знатна финансијска средства. То је сасвим разумљиво ако се има у виду да је у питању дигитализација објеката, ручно SGML/XML означавање, или припрема метаподатака, или све то заједно.

На крају се може парафразирати мишљење Клифа Мекнајта са Одсека за информатику и библиотекарство *University of Loughborough* да је за изградњу дигиталне библиотеке сарадња од виталног значаја, да је успешна дигитална библиотека, баш као и традиционална библиотека, крајње колаборативни подухват.

ЛИТЕРАТУРА

- [1] Atkins, D. E., Birmingham, W. P., Durfee, E. H., Glover, E. J., Mullen, T., Rundensteiner, E. A., Soloway, E., Vidal, J. M., Wallace, R., Wellman, M. P.: *Toward Inquiry-Based Education Through Interacting Software Agents*, *IEEE Computer*, 29(5), 1996
- [2] Baker, J.: *UCLA-NSF Social Aspects of Digital Libraries Workshop*, January 1996, (<http://www.gslis.ucla.edu/DL/>)
- [3] Chang, S-F, Smith, J. R., Meng, H. J., Wang, H., Zhong, D.: *Finding Images/Video in Large Archives*, *D-Lib Magazine*, February 1997
- [4] Faloutsos, Ch.: *Multimedia IR: Indexing and Serching*, *Modern Information Retrieval*, Baeza-Yates, R. and Ribeiro-Nero, B. (eds.), Addison-Wesley, 1999, pp. 345-365
- [5] Fox, E. A., Sornil, O.: *Digital Libraries*, *Modern Information Retrieval*, Baeza-Yates, R. and Ribeiro-Nero, B. (eds.), Addison-Wesley, 1999, pp. 415-432
- [6] French, J., Fox, E., Maly, K., Selman, A.: *Wide Area Technical Reprot Service: Technical Reports Online*, *Communications of the ACM*, 38(1), 1995, pp.
- [7] Gaynor, E.: *From MARC to Markup: SGML and Online library Systems*, *ALCTS Newsletter*, 7(2), 1996.
- [8] Gladney, H. M., Mintzer, F., Schiattarella, F., Bescós, Treu, M.: *Digital Access to Antiquities*, *Communications of the ACM*, 41(4), 1998, pp. 49-57
- [9] Gonnet, G.H, Tompa, F.W: *Mind your Grammar: a New Approach to Modelling Text*, *Proceedings of the 13th International Conference on Very Large Data Bases*, 1987, 339–346
- [10] Guthrie, K.: *JSTOR: Providing new access to old research*, *Proceedings of the ICC/IFIP Conference on Electronic Publishing '98: Towards the Information-Rich Society*, Budapest, Hungary, 1998

- [11] Klavans, J.L.; Schauble: NSF-EU Multilingual Information Access, *Communications of the ACM*, 41(4), 1998, p. 69
- [12] Крстев, Ц: Формати за представљање информација: отворена питања, *ИНФОтека*, 2000, год. 1, бр. 2, 51–57
- [13] Lesk, M.: *Practical Digital Libraries — Books, Bytes & Bucks*, Morgan Kaufmann, San Francisco, 1997
- [14] Miller, E.: An Introduction to the Resource Description Framework, *D-Lib Magazine*, 4(5), 1998
- [15] Miller, G., Beckwith, R., Fellbaum, Ch., Gross, D., Miller, K.: Introduction to WordNet: An on-line lexical database, *International Journal of Lexicography*, 3(4):235-244, 1990
- [16] Moen, W. E.: Accessing Distributed Cultural Heritage Information, *Communications of the ACM*, 41(4), 1998, p. 45-48
- [17] Oard, D. W.: Serving Usres in Many Languages, *D-Lib Magazine*, 3(12), 1997
- [18] Paepcke, A., Cousins, S. B., Garcia-Molina, H., Hassan, S. W., Ketchpel, S. P., Röscheisen, M., Winograd, T.: Using Distributed Objects for Digital Library Interoperability, *IEEE Computer*, 29(5), 1996
- [19] Schatz, B., Chen, H.: Building Large-Scale Digital Libraries, *IEEE Computer*, 29(5):22-26, 1996, (<http://computer.org/computer/dli/>)
- [20] Schatz, B., Mischo, W. H., Cole, T. W., Hardin, J. B., Bishop, A. P., Chen, H.: Federating Diverse Collections of Scientific Literature, *IEEE Computer*, 29(5), 1996
- [21] Smith, T. R.: A Digital Library for Geographically Referenced Materials, *IEEE Computer*, 29(5), 1996
- [22] Sperberg-McQueen, C. M., Burnard, L. (eds.): *Guidelines for Electronic Text Encoding and Interchange of Machine Readable Texts*, TEI P3, ACM-ACL-ALLC, Chicago, Oxford, 1994
- [23] Suleman, H., Atkins, A., Gonçalves, M. A., France, R. K., Fox, E. A., Chachra, V., Crowder, M., Young, J.: Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 1: Mission and Progress, *D-Lib Magazine*, 7(9), 2001.
- [24] Suleman, H., Atkins, A., Gonçalves, M. A., France, R. K., Fox, E. A., Chachra, V., Crowder, M., Young, J.: Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 2: Services and Research, *D-Lib Magazine*, 7(9), 2001.
- [25] Wactlar, H. D., Kanade, T., Smith, M. A., Steven, S. M.: Intelligent Access to Digital Video: Informedia Project, *IEEE Computer*, 29(5), 1996
- [26] Weibel, D.: The State of the Dublin Core Metadata Initiative, *D-Lib Magazine*, 5(4), 1999
- [27] Weibel, S., Jul, E., Shafer, K.: PURLs: Persistent Uniform Resource Locators.
<http://purl.oclc.org/OCLC/PURL/SUMMARY.HTML>
- [28] Wilensky, R.: Toward Work-Centered Digital Information Services, *IEEE Computer*, 29(5), 1996

Association of College and Research Libraries

Slavic and East European Section

NEWSLETTER, No. 18, 2002

или

SEES NEWSLETTER

ГЛАСИЛО

*Секције за славистичке и источноевропске студије
Заједнице високошколских и истраживачких библиотека
која је део Заједнице америчких библиотека (American
Library Association)*

Од 2002. г. главни уредник је Тери Тикл Милер,
Универзитет у Мичигену.

Објављује се једном годишње.

Број 18 за 2002. SEES гласила доноси прилоге у следећим
рубрикама:

1. Конференције, 2. Извештаји, 3. Стипендије, 4. Набавка,
5. Библиотекари на новим радним местима, 6. Библиотеке,
7. Библиографије нових публикација,
8. Истраживања у току

Уредништво часописа ИНФОТЕКА добија SEES гласило из Библиотеке
за славистичке и источноевропске студије на Калифорнијском универзитету
разменом за часопис Заједнице библиотека универзитета у Србији

SEES гласило је на располагању библиотекарима библиотека
чланица Заједнице у Читаоници за периодичку у Универзитетској
библиотеци "Светозар Марковић".