

**Библиографические форматы и метаданные
как инструмент поиска и представления в электронной
библиотеке «Научное наследие России»**

**Bibliographic Formats and Metadata
as a Search and Representation Instrument
in «Scientific Heritage of Russia» E-library**

**Бібліографічні формати та метадані,
як засіб пошуку та подання в електронній
бібліотеці «Наукова спадщина Росії»**

В. В. Рубцов, Н. Н. Елкина

Библиотека РАН, Санкт-Петербург, Россия

Vladimir Rubtsov and Nadezhda Yelkina

Library of the Russian Academy of Sciences,

St. Petersburg, Russia

В. В. Рубцов, Н. М. Йолкіна

Бібліотека РАН, Санкт-Петербург, Росія

Проводится анализ проблемы адаптирования и использования библиографических форматов и метаданных средствами библиотечного программирования, аргументируется их тесная взаимосвязь с инновационными технологиями при создании единого техпроцесса на конкретном примере реализации проекта электронной библиотеки «Научное наследие России».

Как показывает наш опыт, при создании электронных библиотечных ресурсов наиболее сложным является грамотная координация усилий специалистов различного профиля, что подчеркивает значимость совместного творческого труда высококвалифицированных специалистов – библиотекарей, технологов, программистов, работающих на стыке различных специальностей.

The paper analyses the problem of adapting and using bibliographic formats and metadata by means of library programming, sets forth their close interconnection with innovative technologies in forming a single technical process, and illustrates it by realization of the project of «Scientific Heritage of Russia» e-library. According to our experience, the most difficult thing in creating electronic library resources is good coordination of various specialists' efforts, which stresses the importance of joint creative activity of highly-qualified professionals: librarians, technologists, programmers, who do cross-disciplinary work.

Проводиться аналіз проблеми адаптації та використання бібліографічних форматів і метаданих засобами бібліотечного програмування, аргументується їх тісний взаємозв'язок із інноваційними технологіями при створенні єдиного техпроцесу на конкретному прикладі реалізації проекту електронної бібліотеки «Наукова спадщина Росії».

Як показує наш досвід, при створенні електронних бібліотечних ресурсів найскладнішим є грамотна координація зусиль спеціалістів різного профілю, що підкреслює важливість спільної креативної праці висококваліфікованих спеціалістів – бібліотекарів, технологів, програмістів, які працюють на перетині різних спеціальностей.

Жанр библиографической литературы обогащается сегодня благодаря широкому внедрению в практику вычислительной техники, позволяющей создавать средствами информационных технологий библиографию, в которой форматы и метаданные углубляют и делают доступной для поиска огромное количество ранее недоступных элементов библиографического описания. Это ведет к расширению доступа к огромному количеству документов, следовательно, и скрытой в них информации.

После второй мировой войны человечество почувствовало «информационный голод», увеличилась потребность в выходящей во всем мире информации о печатной продукции. Возрастающее число изданий потребовало улучшения учета информации о мировой печати. В 1950 г. в Париже под эгидой ЮНЕСКО была проведена Международная конференция по улучшению национальных библиографических служб. 70-е годы XX века вошли в историю мировой библиографии как период интенсивной разработки международных проектов, направленных на создание глобальных информационных систем. Каталогизация, или библиографическое описание, стало инструментом, позволяющим отразить отличительные признаки документа в стандартной форме. Активизация деятельности международных организаций, интенсивное внедрение ЭВМ в библиотеках подготовили почву для создания проекта Универсального библиографического учета и ускорили принятие международных стандартов на библиографическое описание. В 1970-х гг. появились более 20 его различных версий, ориентированных на национальные правила каталогизации. Тогда же появилась и машиночитаемая каталогизация.

Задача машиночитаемой каталогизации – объединение всех библиотечных каталогов в единый информационный ресурс. Трансформация языка информационной среды внесла свои коррективы и привела к созданию компьютерной лексикографии – тезаурусов, дескрипторов, индексных (инвертированных) словарей и проч. Модификация информационных концепций и технологий в мировой библиотечной практике не могла не отразиться на практической деятельности библиотек. В целях оптимизации своей деятельности по удовлетворению информационных потребностях читателей в изменившихся условиях библиотеке пришлось решать задачи по внедрению новых информационных технологий.

Таким образом, технология создания библиографических форматов стала отличительной чертой машиночитаемой каталогизации. Всё это неизбежно потребовало для учета обработки документопотока внедрения оригинальных программных продуктов, ориентированных на библиотечно-библиографическую деятельность. Эта деятельность опирается на знания и опыт библиографов, разработчики инженеров и программистов для создания информационных производственных процессов, что обеспечивает новое качество библиографических форматов для описания любых видов документов. Как показывает опыт, наиболее сложным при создании электронных библиотечных ресурсов является грамотная координация усилий специалистов различного профиля, работающих на стыке различных специальностей – библиотекарей, технологов, программистов.

Используя алгоритмы обработки библиографической информации, программист и библиограф создают интегрированные автоматизированные информационно-поисковые системы, превосходящие по качеству традиционные карточные каталоги (алфавитные, предметные, систематические, служебные и проч.).

В 1999 году появился термин «метаданные» и был предложен стандарт DCMI – Dublin Core Metadata Initiative. Термин «метаданные» является, по сути, дальнейшим развитием представления о библиографических форматах и практически с ним коррелируется.

Таблица 1.
Соответствия библиографического формата (UNIMARC)
и элементов Дублинского Ядра

N п/п	Элемент формата «Дублинское ядро» Dublin core		Метка поля в формате UNIMARC	Комментарий
1.	Название	Title	200^a	Совпадает с названием, под которым формально известен ресурс.
2.	Создатель	Creator	730^a 700^a 710^a	Человек, организация или сервис, ответственные за производство содержимого ресурса.
3.	Тема	Subject	606^a 610^a 675^a 686^a	Ключевые слова, фразы, коды классификации
4.	Описание	Description	300^a	Резюме, оглавление, ссылка на графическое представление контента, произвольное описание и другая информация
5.	Издатель	Publisher	210^c	Человек, организация или сервис, делающие ресурс доступным
6.	Внёсший вклад	Contributor	730^a 701^a 702^a 711^a 712^a	Человек, организация или сервис, участвующие в создании содержимого ресурса
7.	Дата	Date	210^d 100^a 005	Дата события в жизненном цикле ресурса
8.	Тип	Type	200^b	Вид содержимого ресурса (носитель)
9.	Формат документа	Format	215^a^c^e	Измерение или физическая характеристика
10.	Идентификатор	Identifier	856^u	Конкретная ссылка на ресурс в данном контексте
11.	Источник	Source	452^a	Ссылка на ресурс, на основе которого составлен данный ресурс
12.	Язык	Language	101^a	Национальный язык содержимого
13.	Отношения	Relation	488^a 311^a	Ссылка на связанный ресурс
14.	Покрытие	Coverage	607^a^x^y^z 610^a	Рубрика: географическое название (рубрика) и временной период (Subject Term)
15.	Авторские права	Rights	801^a^b	Человек, организация или сервис; обычно совпадает с именем человека, названием организации или сервиса

<http://dublincore.org/documents/dces/>

Примером использования библиографических форматов и метаданных может служить проект электронной библиотеки (ЭБ) «Научное наследие РАН», в котором на практике применены информационные и библиотечные технологии, сопутствующие технические, программные и лингвистические средства обеспечения. В рамках единого технологического процесса создания электронной библиотеки «Научное наследие России» создаются форматы и метаданные, поддерживаемые информационными потоками текстовой, графической, мультимедийной информации как инструмент представления и поиска.

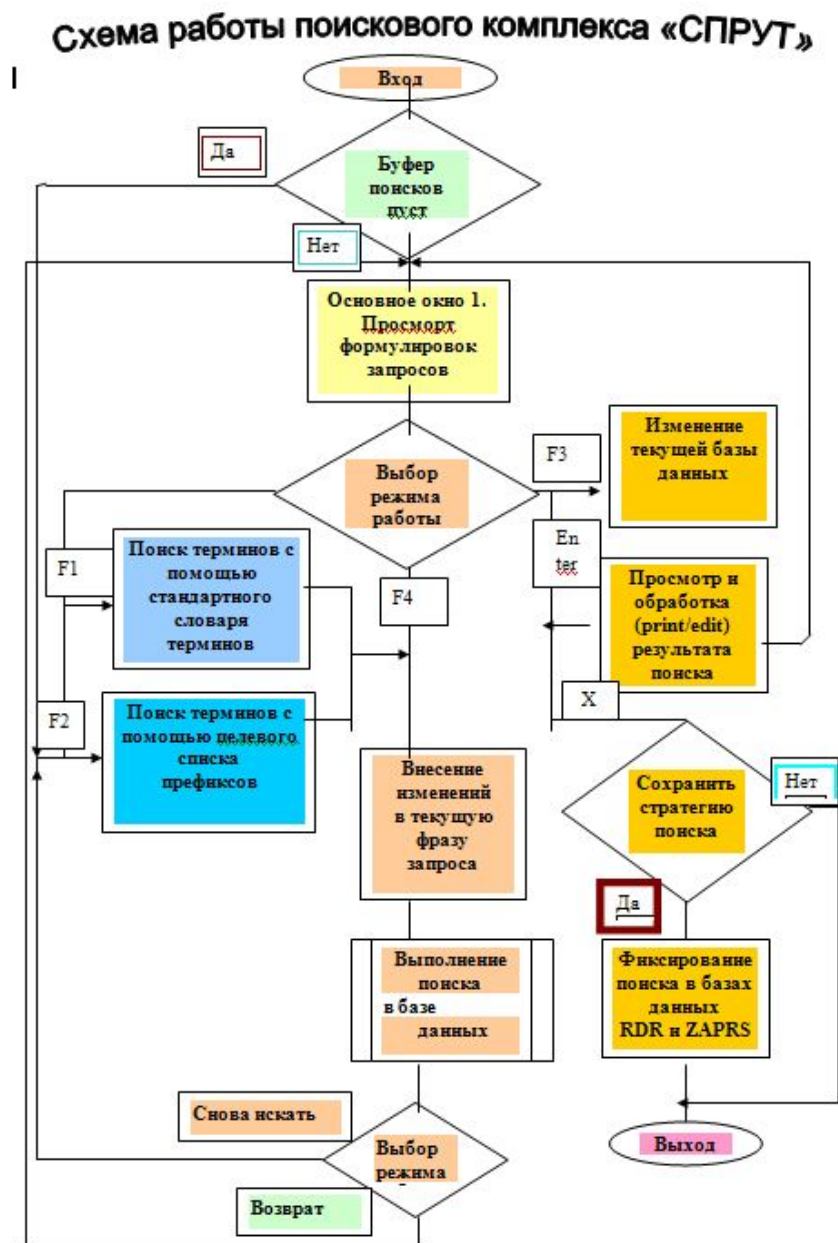


Рис. 1. Схема поискового комплекса «СПРУТ»

Для целей создания электронного каталога ЭБ нами был создан сводный формат хранения информации, который включает набор форматных элементов библиографических реквизитов и/или атрибутов и метаданных. Метаданные, обслуживающие описываемую технологию, в сочетании с международным форматом UNIMARC, согласуются с форматом «Дублинское ядро». Формат детально структурирован, дополнен необходимыми элементами локального использования, и на его основе, как из кирпичиков, создается информация электронных массивов и метаданные. В Библиотеке РАН этим целям служит поисковый автоматизированный комплекс «СПРУТ».

Таким образом, форматно и программно строится новый вид библиографической информации по технологии, которая позволяет сохранить целостность, единство данных в их совокупности. Техническая информация для оцифрованных документов в сочетании с библиографической представляют собой уже новую сущность – полный текст описываемого в электронном массиве документа и его характеристики.

Для разработки этого комплекса оригинальных программ использованы языки: Pascal – внутренний язык программирования в СУБД и Delphi-2007 – для создания самостоятельных сервисных приложений, работающих непосредственно с созданными информационными массивами.

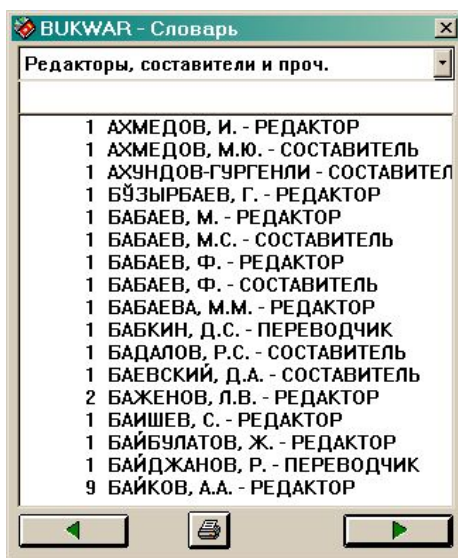


Рис. 2. Фрагмент виртуального списка

Примером использования библиографических форматов и метаданных как инструмента поиска и представления данных может служить оригинальная методика использования гиперссылок в библиографической записи. Они позволяют нам динамически, по запросу генерировать виртуальные сущности, например: поисковые списки по разным критериям.

В общероссийской программе создания электронной библиотеки «Научное наследие России» участвуют также несколько отделов и филиалов Библиотеки РАН. Ими совместно создается рабочий электронный каталог данной коллекции, как самостоятельная часть общего каталога БАН, а также ведется сводный каталог Электронной библиотеки (ЭБ) «Научное наследие» он-лайн через Интернет-систему «Наука в России/SciRus. Информация подготавливается и представляется в двух видах: биобиблиографической и полнотекстовой.

Для автоматизации каталогизирования и процессов диспетчеризации при создании Электронной библиотеки в БАН, создания отчетов выгрузки в электронные таблицы Excel для отправки этих таблиц-отчетов в Москву в качестве документального сопровождения вместе со сканами, создан, отлажен и внедрен комплекс типовых методик и оригинальных сервисных программ:

- Сервис работы в многобазовой среде. Фрагмент Таблицы выбора BUKWAR.FST

```

• 701 0 (v701^a,|v701^b,if v701^4<>" then
• REF->RELAT(L->RELAT('COD:',v701^4),(' - ',v200[1]^a)) fi/)

```

- Сервис работы с включением в записи ссылки на мультимедийные компоненты, в том числе – ссылки на интернет-ресурсы. Фрагмент формата FORM02.PFT

```

• #,link((v999^a), 'OPENFILE ',v999^b)
• v999^a = Руководство по UNIMARC: Библиографический формат
• v999^b = http://www.bnrm.md/sibimol/info/unimarc_rus_IFLA/sec-uni.htm

```

- Сервис контроля полноты и качества ввода информации и генерации списков – диагностических сообщений (методика широко использует также механизм гиперссылок и оригинальные программы). Фрагмент Таблицы выбора SEL01.FST

```

• 9 0 if p(v702) and v702^a=" then 'ошибка в mfn ', mfn fi
• 9 0 if v702^4*0.3='070' then 'ошибка в реляторе mfn ', mfn fi
• 9 0 if p(v700)and p(v710) then 'ошибка в mfn',mfn fi

```

- Программы технологического контроля для представления данных на экране с выдачей диагностических сообщений. Фрагмент формата представления данных FORM03.PFT

```

• if s(v97^x,v97^y)=" then ' . Номер МСЦ отсутствует' fi,
• if v9^a<>v97^x*0.8 then cl13 ' – Ошибка в номере МСЦ !!!' fi,

```

- Программы, автоматизирующие ввод учетных реквизитов, например: присвоение электронных шифров, автоматическое дублирование наиболее ответственных реквизитов, и др. Фрагмент файла настройки BUKWAR.VAL

```

• begin:if mfn=0 then 'a801# ^aRU^brasl_spb^9Библиотека Российской академии
наук. СПб^c',&v801('),#', 'a098#',&dtss('
^a^b(скан)^x'),#', 'a001#',&dtss('EB_г.изд._Otdel_'),#', 'a100#',&v100('),#', fi,
• end:if IsChanged() then 'd005', 'a005#',&v005('),#', fi

```

- Сервисы корректировки разделительных знаков библиографического описания с помощью алгоритмов, которые не могут быть реализованы стандартными средствами. Программа SIZZE7.PAS – корректировка финальных точек, в формате вывода.

```

• (if v300^a<>' then &size7(v300^a) '!' fi)

```

- Программы генерации табличных документов – отчетов, интегрирующих сведения об отосланной информации в виде коротких таблиц (имена и размеры файлов-сканов и проч.).

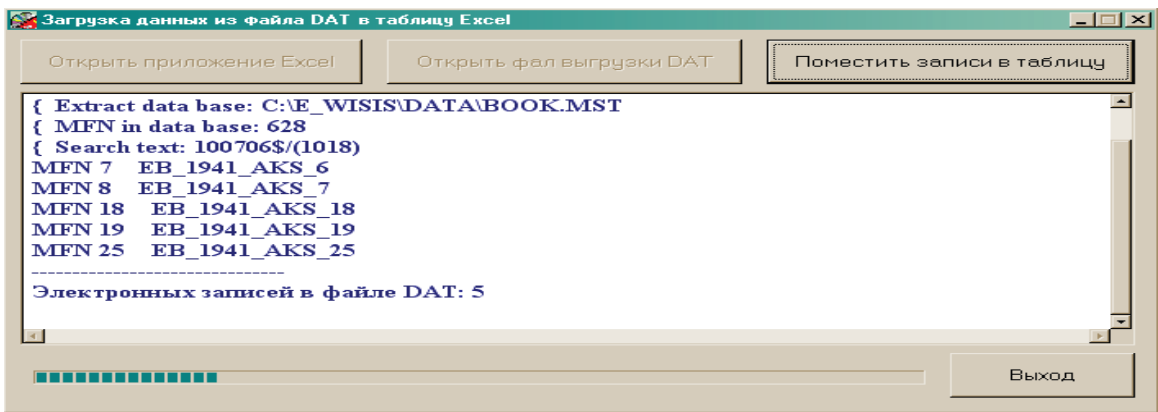


Рис. 3. Протокол программы-обработчика PEXL1.EXE

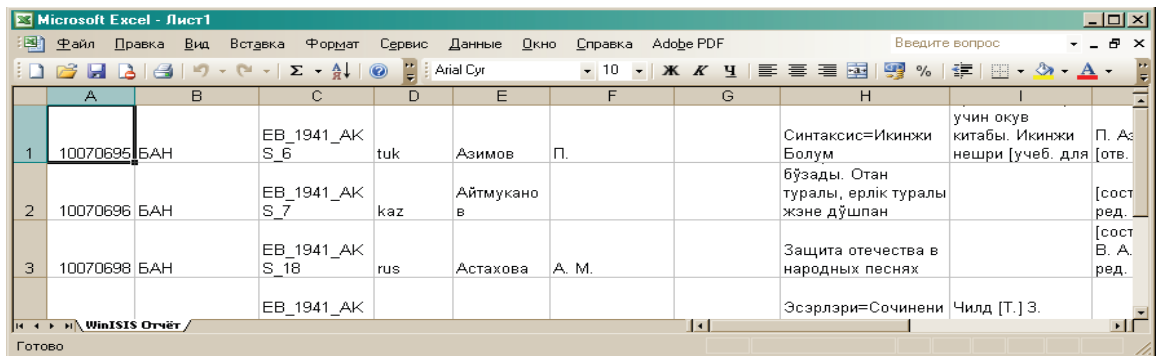


Рис. 4. Окно выгрузки 1 в Excel-таблицу

БАН №	Электронный шифр папок	Количество файлов	Количество Мб
1	EB_1941_AKS_6	167	68.4
2	EB_1941_AKS_7	43	2.33
3	EB_1941_AKS_18	35	5.53
4	EB_1941_AKS_19	331	46
5	EB_1941_AKS_25	23	4.46
Итого:			

Рис. 5. Окно выгрузки 2 в Excel-таблицу

- Методика работы с библиографической записью «вне тела баз данных» для создания коммуникативного обменного формата (селективная выгрузка, редактирование, предзагрузочный контроль, загрузка).

Рис. 6. Фрагмент Рабочего листа ввода структурированных данных в БД

Рис. 7. Фрагмент текстового редактирования структурированного файла «вне тела» БД

Выполнение проекта «Научное наследие России» также подтвердило правомерность использования нами методик библиотечного программирования, что расширило возможности библиографического формата-предшественника, информация полей и подполей которого может рассматриваться как элементы метаданных.

МФН 117 | Формат: ВУКВАЛ | Марк: | Поиск #2: ВУК | -STOP-

Резервное копирование текщей записи перед редактированием
Каталог электронных книг * 1941-1945 гг.

Шифр ЭК: EB_1941_AKS_0000001 = 113 сканов, 18 Мб
 Шифр печатного издания: АкР/6191, ОИАИ
 Сорговод. N.1, 20061220. Номер МСЦ: 10070734. - внесено в св. каталог БЕИ

Абегян, Манук Хачатурович (1865-1944), проф., академик АрмФАН СССР (1943)
«Вишалы» как статуи богини Астхик-Деркето [Мультимедиа] / М. Абегян ; Ин-т истории и матер. культуры Арм. фил. АН СССР. - Ереван : Изд-во АрмФАН СССР, 1941. - 98, [8] с. : ил. - На арм. яз. - Загл. парал. арм., рус.

Коллективные авторы (1)

Язык издания: русский армянский
 Страна издания: СССР

Номер МСЦ: 10070734
Аннотация
 Вишалы - каменные изваяния рыб до 5 м высотой, опит. плодородия. Известны на Кавказе и в Монголии

Персоналия
 * Астхик-Деркето, богиня воды.

Ключевые слова
 * Древнеармянская литература
 * Древнеармянский фольклор
 * Археология
 * Каменные статуи

Географические рубрики
 * РСФСР
 * Кавказ
 * Армения

Иллюстрации:
[Обложка](#)
[Титульный лист](#)
[Фрагменты рисунков 1](#)
[Фрагменты рисунков 2](#)
[Резюме](#)
[Иллюстрация 1](#)
[Иллюстрация 2](#)

Предметизация

АКАДЕМИЯ НАУК СССР - АРМЯНСКИЙ ФИЛИАЛ
 ИНСТИТУТ ИСТОРИИ И МАТЕРИАЛЬНОЙ КУЛЬТУРЫ
 ПРОФ. Д-Р Я. АБЕГЯН
«ВИШАЛЫ»
 КАК СТАТУИ БОГИНИ
 АСТХИК - ДЕРКЕТО
 (На армянском языке)

ՀԻՇԱՊՆԵՐ
 ՎԻՇԱԼՆԵ ԿՈՅՈՒՆԵՐՆԻ ԻՐԵՐԻԼ
 ՍՈՏՆԻԿ-ԳԵՐԿԵՏՈՒ ԳԻՏՈՒԿԻՆ ԻՐԱՎԱՆՆԵՐ

Рис. 8. Страница библиографического формата и метаданных

На 2010 год в БАН библиографировано, внесено в электронный каталог и отсканировано более 1,5 тыс. изданий академической книжной продукции военного времени (1941-1945). Данный этап работы ведется совместно со специалистами Физико-Технического института РАН, а также Института цитологии РАН в Санкт-Петербурге и приурочен к 65-летию Победы в Великой Отечественной войне.

Литература

1. Англо-русский словарь по библиотечной и информационной деятельности / Сост. Дж. Ричардсон; под ред. Э.Р. Сукиасяна, В.В. Зверевича. СПб.: Профессия, 2005. – 267 с.
2. Воройский Ф.С. Информатика. Новый систематизированный словарь-справочник (Вводный курс по информатике и вычислительной технике в терминах). – 2-е изд., перераб. и доп. – М.: Издательство Либерия, 2001. – 536 с.
3. Гордукалова Г.Ф. Анализ информации. Методы, технологии, организация. – СПб.: Профессия, 2009. – 512 с.
4. Елкина Н.Н. К Истории справочной и информационной работы Библиотеки Российской академии наук. – СПб.: БАН, 2007. – 216 с.
5. Пилко И.С. Основы библиотечной технологии: учебно-методическое пособие. – М.: Профиздат, 2003. – 175 с.
6. Рубцов, В. В. Что может ISIS? Взгляд программиста на библиотечные информационные технологии // Петербург. библи. шк. – 2006. – № 1/2. – С. 74; Рубцов В.В. Технологические аспекты работы с постоянно действующими запросами в автоматизированной библиотечной системе (на примере электронного каталога библиотеки ботанического института им. В.Л. Комарова РАН / В.В., Рубцов, Г.Н. Панкратова, М.Г. Батурина, Т.Ф. Рубцова // 2-я Международная научная конференция «Современные проблемы информатизации в системах моделирования, программирования и телекоммуникациях», 2009 / РАЕ. – URL: <http://www.econf.rae.ru/article/4977>

7. Шрайберг Я.Л. Автоматизированные библиотечно-информационные системы России: состояние, выбор, внедрение, развитие / Я.Л. Шрайберг, Ф.С. Воройский . – М.: Либерея, ГПНТБ России, 1996. – 273 с.
8. ЮНИСИСТ: Доклад об исследовании возможностей создания Всемирной системы научной информации, проведенном Организацией Объединенных Наций по вопросам образования, науки и культуры и Международным советом научных союзов = UNISIST: Study Report in the feasibility of a World Science Information System by the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. – Париж: Юнеско, 1971. – 192 с.
9. Mann Thomas. Doing research at the Library of Congress: A guide to subject searching in a closed State library / Thomas Mann. – Washington: Library of Congress. 1994. – 37 p.
10. Tedd, L.A. Digital libraries. Principles and practice in a global environment / L.A. Tedd, A. Large. – Munchen: K.G. Saur, 2005. – 280 p.