

Методология построения электронных библиотек

Проскудина Г.Ю.

I. Состояние проблемы

1. История ЭБ и библиотечной автоматизации
2. Объект исследования – ЭБ
3. Суть проблематики методологического обеспечения построения ЭБ
4. Методологические аспекты, составляющие технологический процесс

История развития ЭБ и библиотечной автоматизации



Области исследований, повлиявшие на появление и развитие ЭБ:

- Информационный поиск
- Высокопродуктивные вычисления
- Электронные каталоги и автоматизированные библиотечные ИС
- Электронные научные публикации
- Гипертекст
- Базы данных: текст и мультимедиа
- Информатизация гуманитарной сферы

История развития ЭБ в мире

В.Буш концепция информационных систем, «MEMEX»

Дж. Ликлайдер «Библиотеки в будущем», идея - связь по сети с ЭБ

«American Memory», 1989... , «National Digital Library», 1990..., программа «Digital Libraries Initiative», 1998

"Global-Info", Германия

eLib, Великобритания

Инициативы DELOS, ЕС

Библиотека Конгресса, формат MARC

«Проект Гутенберг», 1990, М. Харт, ArXiv, 1991, США

Всемирная электронная библиотека

Первые ЭБ научных журналов проекты «Mercury», CORE, «Tulip», 1987-1993; JSTORE, 1995; «High Wire Press», 1995

NSDL, NDLTD, ACM, США

1940 1950 1960 1970 1980 1990 2000 2010

Программы развития ЭБ

США – «American Memory» (с 1989 по наст. время)

США – Библиотека Конгресса, создание национальной электронной библиотеки, NDL, с 1990

США – Инициатива ЭБ (DLI), 1994 и 1998

Великобритания – eLib, 1995

Юнеско – Memory of the World Program, 1992

ЕС – инициативы DELOS, конец 90-х

Исследовательская программа DLI

Основатели: NSF, DARPA и NASA, 1994

6 проектов: 5 CS (computer science) программ (например, в Стенфорде, рождается прототип системы Google), 1 программа LIS (library and information science)

Главная цель: разработать инфраструктуру ЭБ

Результаты: Были построены первые коллекции

Информационные объекты: тексты периодических печатных изданий, изображения, карты, аудио- и видеозаписи, большие наборы данных.

Это происходило одновременно с появлением сети WWW

ЭБ в мире

США – NSDL, NDLTD, ACM...

Япония – Электронная "библиотека XXI столетия"

Германия – "Global-Info" (<http://www.global-info.org>), 1995, 16 университетов и 4 научных сообществ (физики, математики, химии и информатики)

Франция – "Gallica" (<http://gallica.bnf.fr>), 1997, под эгидой Национальной библиотеки Франции

Россия – **Elibrary.ru**: 15 млн научных статей

более 2500 тыс. журналов, имеет РИНЦ. **Mathnet...**

National Science Digital Library (NSDL)

Создана в 2000 NSF, как точка доступа к STEM-контенту (наука, технология, инженерия, математика), собранного из разных ЭБ.

Содержит **метаданные** о ресурсах с других сайтов, их провайдеров, т.е. через NSDL и его сервисы осуществляется поиск и открытый доступ к образовательным ресурсам.

Топ 25 сайтов для преподавания и обучения 2010

The screenshot shows the NSDL website interface. At the top, there is a navigation bar with the NSDL logo, the text "National Science Digital Library", and the tagline "Serving science, technology, engineering, and mathematics education". There are links for "Register Now", "Log In", and "Help", along with a "LEARNER OPTIONS" button. A search bar is also present. Below the navigation bar, there is a horizontal menu with options: "Explore", "About", "Education Level", "Resource Type", "Subject", "Standards", and "Advanced Search". The main content area features a large image of Earth from space. Below the image, there are two main sections: "Search Collections" and "Search By Provider". The "Search Collections" section includes the text "Learn more about the organizations that share their resources with NSDL." The "Search By Provider" section includes the text "Browse an alphabetical list of all participating providers." Below these sections, there is a "Find a Resource" section with a search input field labeled "Enter a keyword or search term...". Underneath, there is a table for "Education Level" with the following categories: Preschool, Middle School, College / Upper Division, Adult Education, Lower Primary, High School, Graduate / Professional, Upper Primary, Community College / Lower Division, and Career / Technical. At the bottom, there is a section for "Resource Type".

ЭБ АСМ

2,3 млн. библ. записей

1,5 млн. резюме

231+ тыс. названий

6+ тыс. издателей

1+ млн. сборников докладов
конференций

880+ тыс. журналов

200 тыс. книг

75 тыс. диссертаций

25 тыс. отчетов

4 тыс. документов RFC

Скачка файлов каждые 3 сек

Contents of the DL

The ACM Guide to
Computing Literature

ACM Computing
Reviews

Platform and Features

ACM Digital Library
Policies



ACM's Flagship
Magazine



The Digital Library

The *ACM Digital Library* (DL) is the most comprehensive collection of full-text articles and bibliographic records in existence today covering the fields of computing and information technology. The full-text database includes the complete collection of ACM's publications, including journals, conference proceedings, magazines, newsletters, and multimedia titles and currently consists of:

- 407,367 Full-text articles
- 2.0+ Million Pages of full-text articles
- 18,000+ New full-text articles added each year
- 44+ High Impact Journals with 2-3 new journals being launched each year
- 275+ Conference Proceedings Titles added each year
- 2,000+ Proceedings Volumes
- 8 Magazines (including the flagship *Communications of the ACM*, the most heavily cited publication in the field of computing according to Thomson-Reuters)
- 37 Technical Newsletters from ACM's Special Interest Groups (SIGs)
- 6,500+ Video files
- 594 Audio files

In addition to the full-text database, the *ACM Digital Library* is heavily integrated with and includes unrestricted access to the *Guide to Computing Literature* bibliography.

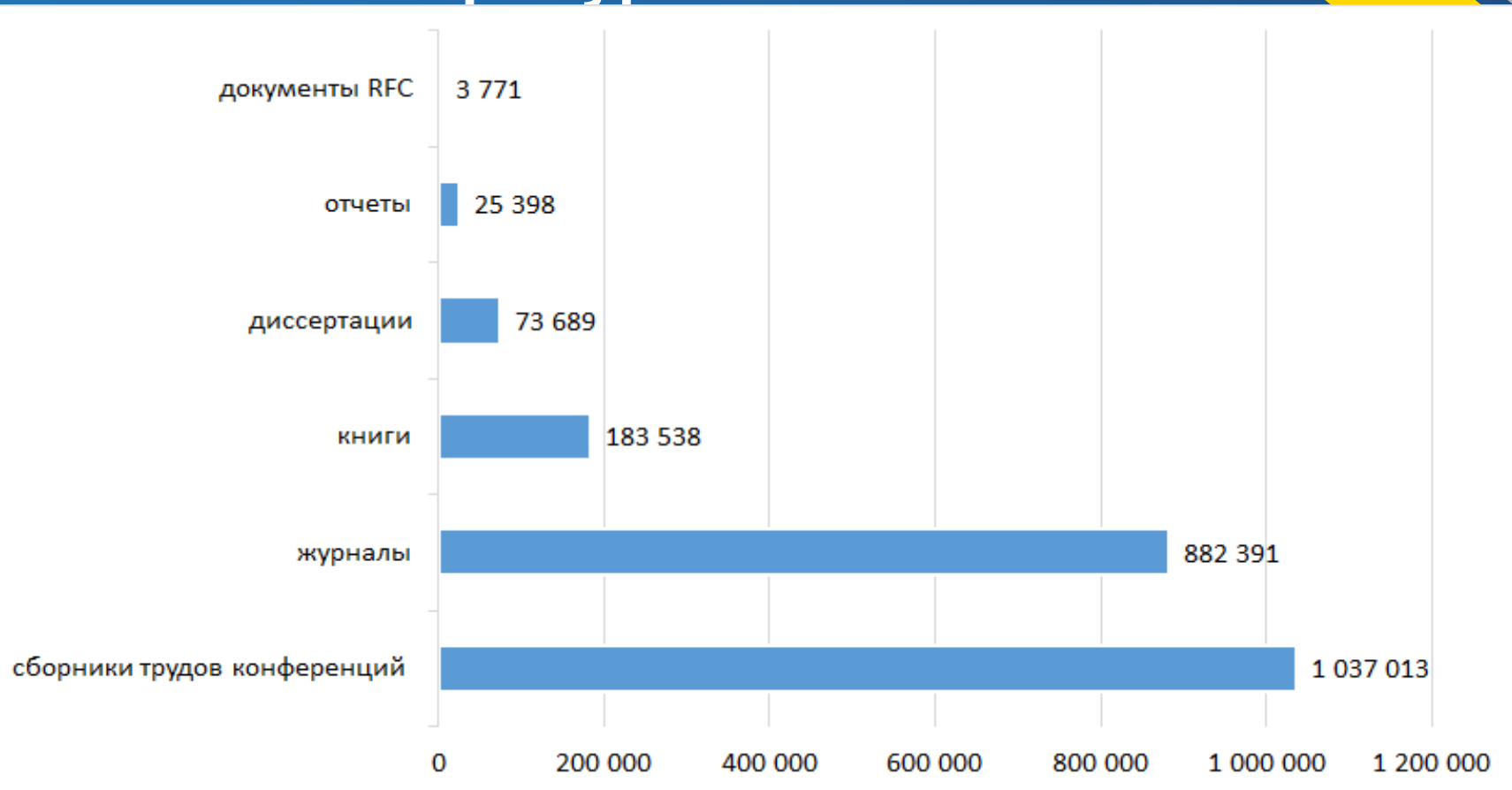
The *ACM Digital Library* includes reference linking through CrossRef, integration with the ACM *Computing Reviews* database, index terms using ACM's 2012 *Computing Classification Scheme (CCS)*, alerting and TOC services, and all export formats including BibTeX, Endnote, and ACM Ref, as well as OpenURL compliance, and COUNTER III and SUSHI Compliant usage statistics.

Free trial



30 day trials are
available for institutions
interested in evaluating
the ACM DL.
[Apply Here...](#)

Распределение по видам хранимых ресурсов в ЭБ АСМ



Объект исследования - ЭБ

- Концептуальные структуры, модели, теории, определения
- Информационная модель ЭБ
- Сервисы
- Архитектура
- Взаимодействие с пользователем, поведение пользователя
- ...

Концептуальные структуры,
модели, теории, определения



Электронные библиотеки – это управляемые коллекции информации, связанные с ними сообщества и сервисы, где информация сохраняется в цифровой форме и доступна по сети.

Концептуальные модели, описывающие сферу ЭБ:

DLRM - Digital Libraries Reference Model, Эталонная

DELOS схемы классификации и оценки

5S

...

Что такое эталонная модель?

Абстрактная структура для понимания значимых **связей** между объектами некоторой среды.

Основана на **небольшом числе объединяющих концепций** и предназначена для разработки более узких моделей конкретной архитектуры для дальнейшей реализации программных систем.

Непосредственно не связана со стандартом, технологией, другими конкретными деталями, ее цель - предоставить **общую семантику**, которую можно использовать в разных реализациях.

Эталонная модель различает

ЭБ – конкретная ЭБ со своими пользователями, правилами, контентом, интернет-сайтом и ведущей организацией. Например: библиотека института программных систем <http://eprints.isofts.kiev.ua>

система ЭБ – ПО, на основе которого создается ЭБ. Например: EPrints

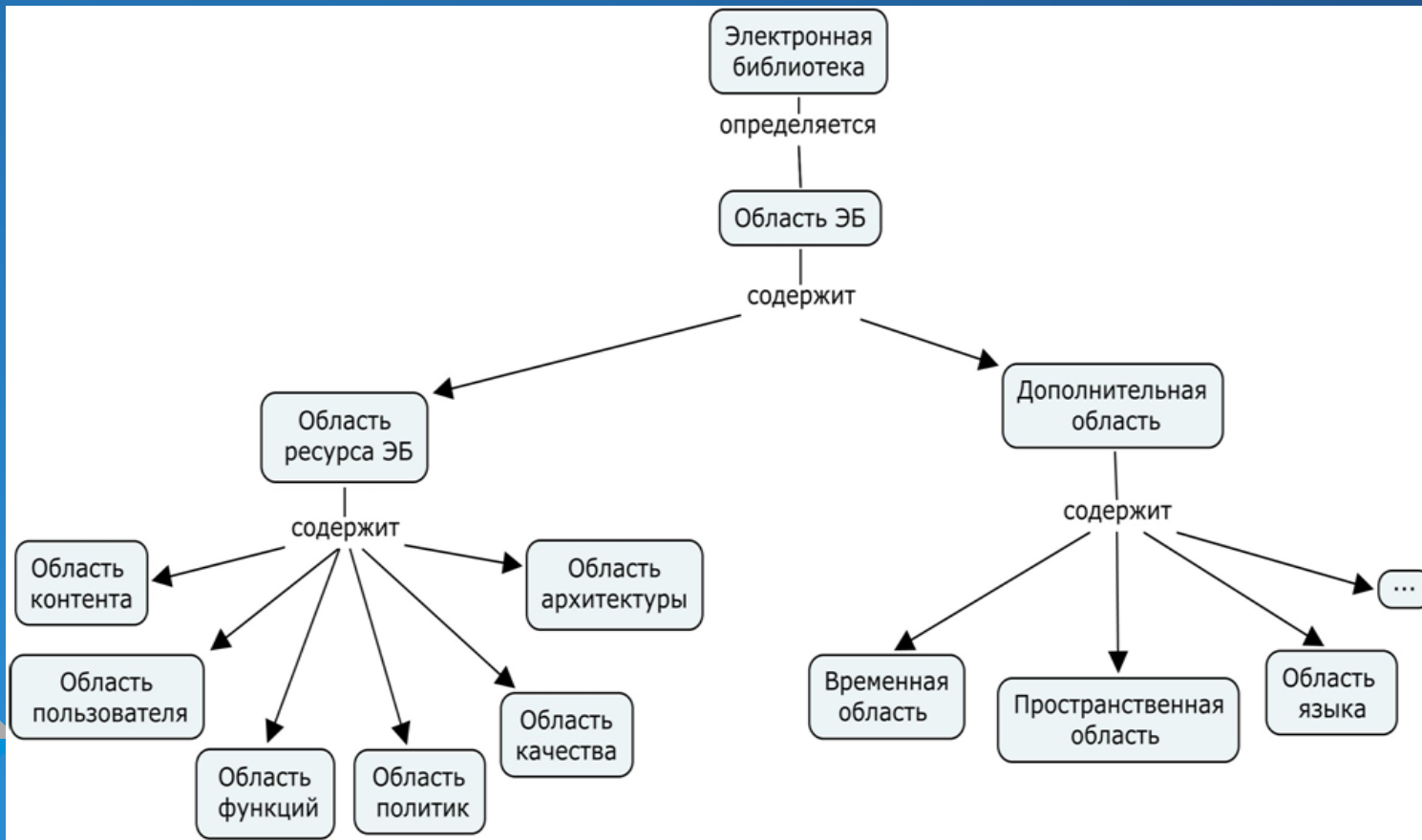
система управления ЭБ – ПО для создания и управления системами ЭБ. Например, OpenDLib

Цель и структура эталонной модели:

разобраться с фундаментальными понятиями, существенными объектами и их связями, стандартными функциональными и структурными блоками и процессами, из которых состоит ЭБ:

- контент
- пользователь
- функциональность
- качество
- политика
- архитектура

Иерархия областей в DLRM



Модель 5S

Потоки (Streams), структуры (Structures), пространства (Spaces), сценарии (Scenarios) и сообщества (Societies)

Минимальная ЭБ: (**R**, **Cat**, **Serv**, **Soc**), где:

R – репозиторий, сервис, содержащий семейство коллекций и специфичные сервисы (получить, сохранить и удалить), чтобы манипулировать коллекциями;

Cat – множество метаданных для всех коллекций репозитория;

Serv – множество сервисов, содержат как минимум сервисы для индексации, поиска и просмотра;

Soc – сообщество

Информационная модель ЭБ

- Цифровые объекты ЭБ;
- Организация информации и знаний;
- Построение коллекций

Цифровые объекты ЭБ

- Текстовые ресурсы
- Мультимедиа
- Изображения
- Форматы файлов, преобразование, миграция

Документ может включать текстовую, графическую, аудио- или видеоинформацию, иметь линейную или нелинейную структуру

Задача: Типы и описание документов, составляющих продукцию научной организации.

Организация информации и знаний

Метаданные

Задача: Описание схем метаданных, формулировка требований к этим описаниям.

Онтологии, классификации, категоризации,
Тематическое описание, контролируемые словари,
тезаурусы, терминология

Задачи: Описание предметной области. Работа с предметными классификациями. Обзор инструментов инженерии онтологий.

Описание и организация объектов конкретного домена

Построение коллекций

Выбор политики

Оцифровка

Сбор информации (харвестинг)

Задача: Методологические аспекты обеспечения харвестинга

Веб-публикации

Краулинг

Сервисы



Задача: Описание функциональной составляющей ЭБ.

Индексация и поиск

Задача: Особенности осуществления поиска и просмотра в ЭБ. Инструкции по просмотру и поиску.

Поиск изображений

Справочные сервисы

Рекомендательные системы

Архитектура



Обзор архитектуры

Прикладное программное обеспечение

Задача: Сравнение и выбор платформы построения ЭБ.

Идентификаторы, handles, DOI, PURL

Протоколы

Интероперабельность

Задача: Методологические аспекты интеграции ЭБ.

Безопасность

Взаимодействие с
пользователем,
поведение пользователя



Информационные потребности, релевантность

Поведение при поиске информации и стратегия поиска

Обмен, сети, совместное использование

Проектирование взаимодействия, оценка юзабилити

Обобщение и визуализация информации

Другие задачи в ЭБ



Долгосрочное хранение

Оценка и изучение пользователей


Оценка информационного поиска

Библиометрия и вебометрика

Задача: Статистика использования ЭБ

Вопросы интеллектуальной собственности

Экономические и социальные вопросы

The background is a solid blue color. In the top-right and bottom-left corners, there are decorative geometric shapes composed of overlapping triangles and quadrilaterals in shades of yellow, magenta, cyan, and grey. The text is centered in the upper half of the slide.

Суть проблематики методологического обеспечения создания ЭБ

Методологическое обеспечение создания ЭБ

Методология создания ЭБ - совокупность принципов, методов, инструментов и средств, применяемых для последовательной разработки ЭБ.

Метод - упорядоченная логическая процедура для решения определенной задачи [Резниченко В.А. “Организация БД и знаний”]

Методологические аспекты процесса создания ЭБ

- оценка источников контента
- состав информационных объектов
- способы их описания (метаданные)
- выбор предметных и других классификаций
- принципы формирования коллекций
- оценка пользователей
- ...

II. Методологическое обеспечение создания ЭБ



Этапы создания ЭБ

1. Цель, планирование и предварительные оценки
2. Определение информационных потребностей организации
3. Определение пользователей и групп пользователей
4. Определение и описание источников информации
5. Сбор, обработка, классификация и описание электронного контента
6. Выбор архитектуры
7. Документирование
8. Разворачивание опытного образца
9. Тестирование и эксплуатация опытного образца
10. Другие задачи, возникающие в процессе эксплуатации

2. Определение информационных потребностей организации (1)

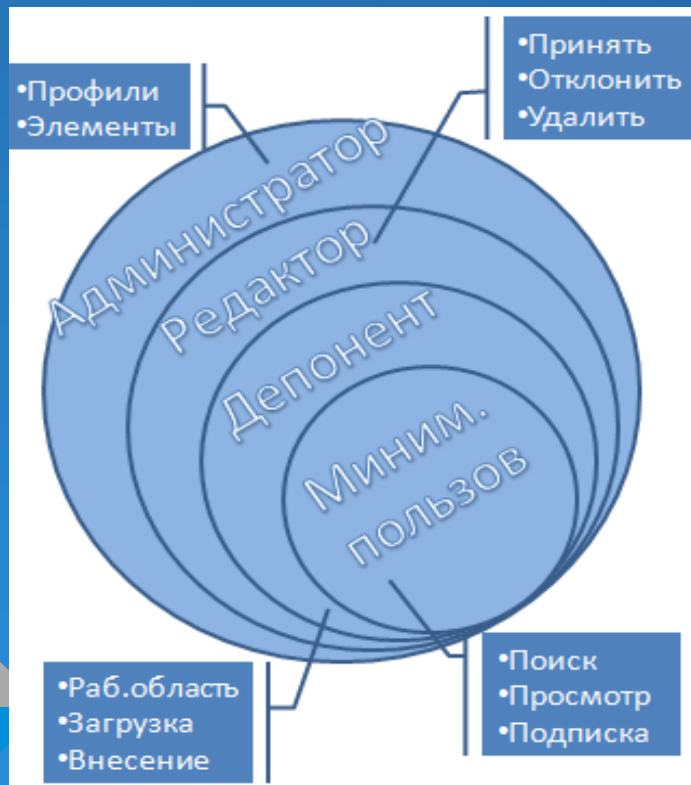
Что предполагается хранить в ЭБ?

- типы документов (наборы данных, бланки, статьи, отчеты, стандарты, патенты,) + их форматы хранения
- описательная информация, стандартные наборы метаданных, расширенные наборы
- каждому типу документов поставить в соответствие - свой набор метаданных
- предметные классификации... и другие стандарты некоторых полей метаданных

2. Определение информационных потребностей организации (2)

- выработать регламент видимости ресурсов и доступа к ним
 - открытый, (будет ли эмбарго ?)
 - закрытый, только для сотрудников организации
 - по подписке
- вопросы языка (языки интерфейса, метаданных, классификаций)
- вопросы авторского права
- принцип организации и описания коллекций, стандарт EAD

3. Определение пользователей и групп пользователей ЭБ



Конечные пользователи:

- создатель контента или депонент
- потребитель контента или читатель
- библиотекарь

Служба поддержки функционирования ЭБ:

- редактор
- администратор
- оператор
- проектировщик
- администратор системы
- разработчик приложений

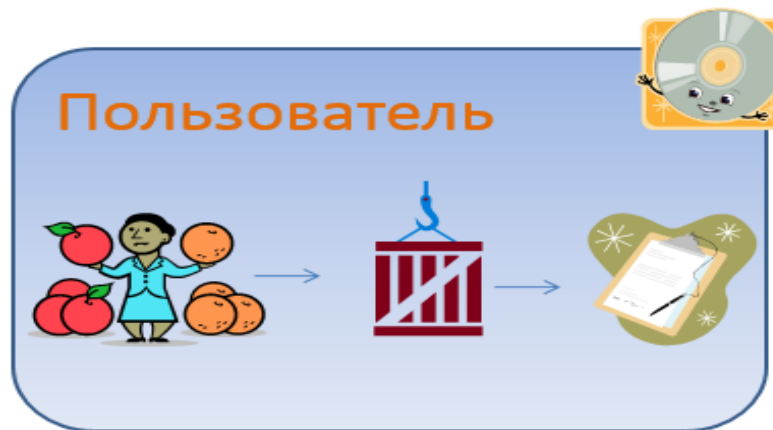
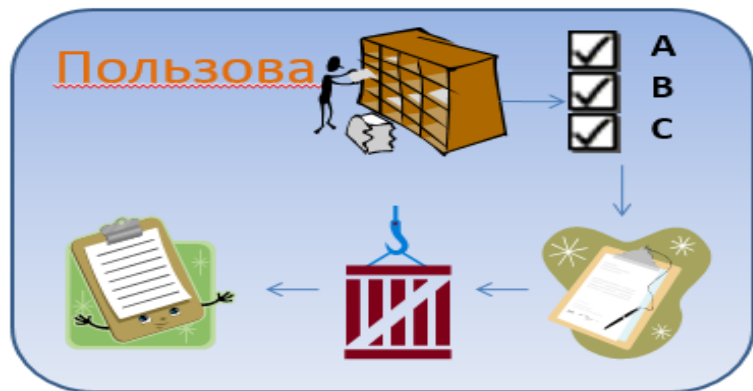
4. Определение и описание источников информации

- оцифровка бумажных источников, если да, нужна инструкция по оцифровке
- авторы вводят свои документы
- ...
- перечень основных сайтов-источников
- вопросы экспорта-импорта, форматы импорта-экспорта

5. Сбор, обработка, классификация и описание электронного контента

- a. описание рабочего процесса внесения
оцифровка
последовательность
отклонение документов
обнаружение плагиата
привлечение авторов и рецензентов, техническое управление
рецензентами и рецензиями
обработка документов, которые присылают авторы
регистрация и защита авторских прав
- b. построение коллекции
- c. сбор электронного контента
- d. обработка и организация цифрового контента
- e. пакетное внесение
- f. управление правами пользователей

Внесение



Редактор



Библиотека

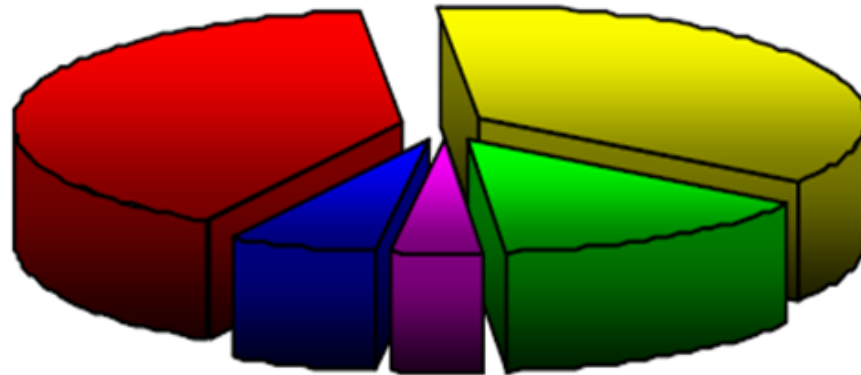


6. Выбор функциональности ЭБ

- поддержка контента
- управление контентом
- поддержка поиска
- контроль доступа и конфиденциальность
- возможности по отчетам и запросам
- долговременное хранение
- взаимодействие (или интероперабельность)
- интерфейс пользователя
- соответствие стандартам
- автоматические инструменты
- поддержка и обслуживание

7. Выбор программно-технических средств

Распределение в мире



- Dspace
- Eprints
- Greenstone
- Fedora
- Others

7. Документирование

Список документов

...



9. Разворачивание, тестирование и эксплуатация ЭБ

- инструкции по установке, настройке, эксплуатации
- резервное копирование
- интеграция
- статистика
- включение в международные реестры, наукометрические БД
- обновление версий ПО
- ---

10. Другие задачи, возникающие в процессе эксплуатации ЭБ (1)

- планирование и разработка стратегии развития
- регистрация (например, журнала, в ISSN и PubMed)
- создание инфраструктуры, наблюдение и корректировки (масса разных действий и вариантов, часто повторяющихся во времени)
- регистрация DOI
- поисковая оптимизация, SEO. Авторы желают, чтобы их статьи были найденными
- локальное хранение и архивация
- обновление и перенос площадок
- отслеживание изменений в области авторского права
- редактирование содержания
- работа с другими типами информационных ресурсов иллюстрации, мультимедиа...

10. Другие задачи, возникающие в процессе эксплуатации ЭБ (2)

- верстка и дизайн
- миграция форматов
- модерация комментариев
- внедрение и поддержка связанных служб
- предложение дополнений
- анализ и отслеживание злоупотреблений. Защита ЭБ и ее данных
- расширенные управленческие функции
- ориентиры и обмен мнениями