

Роль коллекционных предметных метаданных в предметном доступе к электронным коллекциям в порталах

РОЛЬ КОЛЕКЦІЙНИХ ПРЕДМЕТНИХ МЕТАДАНИХ У
ПРЕДМЕТНОМУ ДОСТУПІ ДО ЕЛЕКТРОННИХ КОЛЕКЦІЙ У
ПОРТАЛАХ

ROLE OF COLLECTION-LEVEL SUBJECT METADATA IN
SUBJECT ACCESS TO DIGITAL COLLECTIONS IN
AGGREGATIONS

О. Л. Завалина

Университет Северного Техаса, США

Предметный доступ

- Одна из наиболее важных сфер исследования в областях теории информационного поиска и информационного поведения пользователей (Hjørland, 1997)
- Подходы к предметному доступу (Cochrane, 1985) :
 - систематический (классификационные номера)
 - тематический (предметные рубрики)
 - естественный (заголовок документа, аннотация)
- Осуществляется с помощью метаданных
 - «структурированные данные об информационном объекте, которые поддерживают выполнение этим объектом своих функций» (Greenberg, 2005)

Предметные метаданные

Метаданные, описывающие содержание информационного объекта

1. свободнотекстовые метаданные
 - аннотации к журнальным статьям
2. метаданные нормализованной лексики
 - предметные рубрики:
 - тематические («Архитектура»)
 - географические («Крым»)
 - хронологические («19-й век»)
 - жанровые («энциклопедия»)

Составляющие предметного доступа



Коллекционные метаданные

Метаданные, описывающие целые коллекции информационных объектов как единое целое

- Долгое время используются в архивах
 - «опись фонда»
- В последнее десятилетие широко применяются в электронных порталах
 - портал «Память Америки» <http://memory.loc.gov>
 - сотни электронных коллекций (например, «Золотой век джаза в фотографиях»
<http://memory.loc.gov/ammem/wghtml/wghome.html>

Коллекционные предметные метадаанные (КПМ)

**КПМ = коллекционные метадаанные (КМ)
предметные метадаанные (ПМ)**

Описывают содержание целой коллекции
информационных объектов как единого
целого

Пример КПМ (портал «Открытие истории»)

Collection Information

Title: 1893 Chicago World Columbia Exposition: A Collection of Digitized Books

Collection Home: <http://illinois.harvest.granger.uiuc.edu/results.asp?searchtype=collectioncontent&collID=70533&collname=1893%20Chicago%20World%20Columbia%20Exposition.%20A%20Collection%20of%20Digitized%20Books> #

Items: [Browse 6 Items in Collection](#)

Description: Souvenir guides, photo albums, directories, and reports, pertaining to the 1893 World Columbian Exposition held in Jackson Park on Chicago's south side. The collection is a subset of the University of Illinois Digitized Books Collection.

Library of Congress Subject Headings: World's Columbian Exposition (1893 : Chicago, Ill.).

Subjects: States and Regions
American Culture
American Culture

GEM Subjects: Arts
Architecture
Social Studies
State history
United States history

Geographic Coverage: North and Central America (continent)
United States (nation)
Midwest U.S. (general region)
Chicago, Illinois (city)
Illinois (state)

Time Period: 1850-1899

Objects Represented: Books and pamphlets

Format: text/html
application/pdf
image/vnd.djvu
image/jpeg, text/plain

Language: eng

Audience: General public
Genealogists/History Enthusiasts
K-12 students
Undergraduate Students
Staff at peer/partner organizations
K-12 teachers and administrators
School Researchers/Professors/Students

Постановка проблемы исследования

- Растет число крупномасштабных порталов, которые
 - объединяют в себе сотни (иногда тысячи) электронных коллекций
 - используют КПМ для обеспечения предметного доступа
- Ранее не проводилось исследований, которые бы
 - оценивали применение КПМ
 - анализировали взаимодействие пользователей порталов с КПМ в процессе предметного доступа
- Назрела необходимость тестирования модели «Функциональных требований к библиографическим записям» (ФРБР) в различных контекстах (Zhang & Salaba, 2007)
 - контекст предметного поиска (Riesthuis & Žumer, 2004)

Задачи исследования

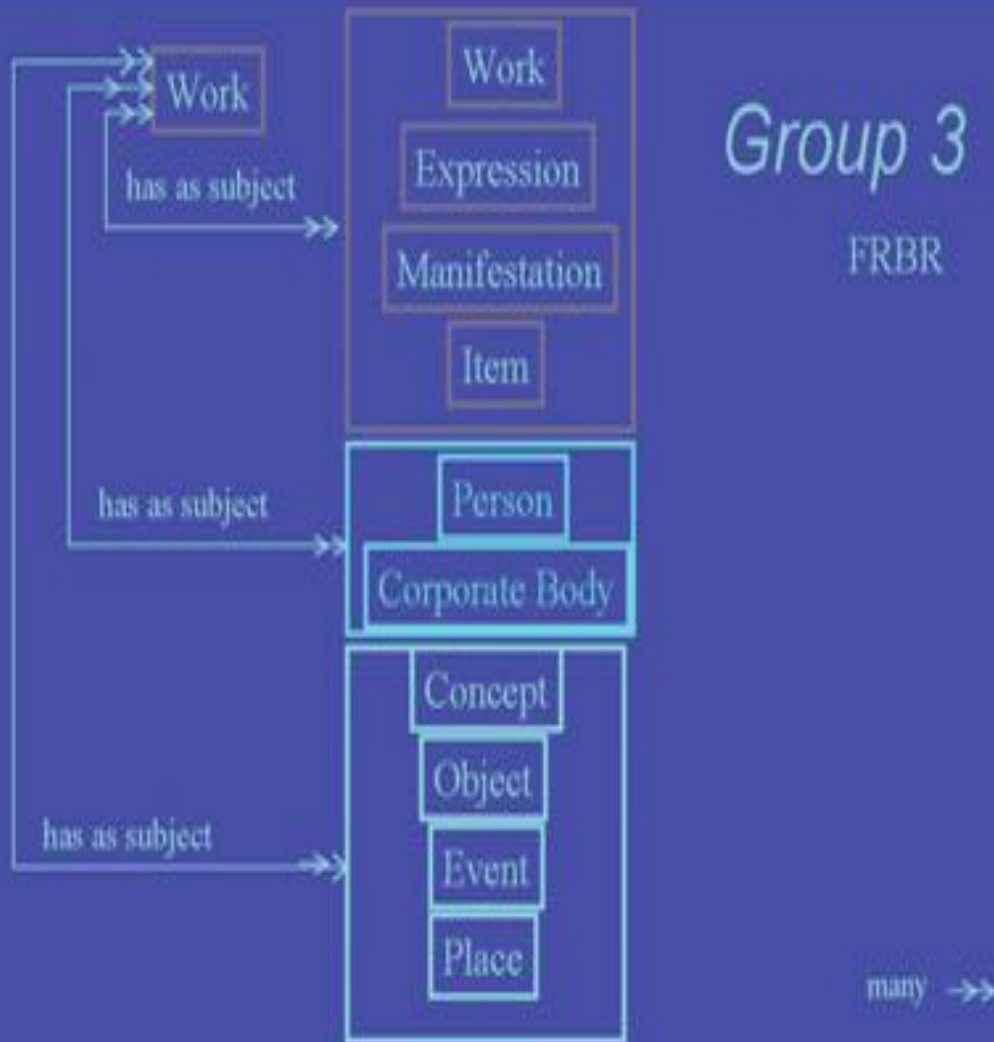
Какова роль коллекционных предметных метаданных в предметном доступе к электронным коллекциям, объединенным в порталы?

- Как КПМ применяются в порталах, объединяющих электронные коллекции?
- Как пользователи взаимодействуют с порталами?
- Каким образом КПМ помогают пользователям находить электронные коллекции?
- Как информационное (в частности, поисковое) поведение пользователей в порталах соотносится с моделью «Функциональных требований к библиографическим записям» (ФРБР)?

Методы исследования

1. Сравнительный контент-анализ представительной выборки коллекционных предметных метаданных в трех крупномасштабных порталах, объединяющих электронные коллекции США и Европы
 - «Память Америки» (<http://memory.loc.gov>)
 - «Открытие истории» (<http://imlsdcc.grainger.uiuc.edu/history>)
 - «Европейская библиотека» (www.theeuropeanlibrary.org)
2. Анализ представительной выборки транзакций пользователей в портале «Открытие истории» в течение одного года
3. Интервью и наблюдения за пользователями-историками в процессе взаимодействия с двумя порталами, объединяющими исторические электронные коллекции США:
 - «Память Америки»
 - «Открытие истории»

Концептуальная основа исследования: предметные категории и связи модели ФРБР (Tillett, 2004)



КАТЕГОРИИ

- произведение (work)
- выражение (expression)
- воплощение (manifestation)
- физическая единица (item)
- лицо (person)
- организация (corporate body)
- концепция (concept)
- физический предмет (object)
- событие (event)
- место (place)

СВЯЗЬ

- имеет своим предметом («has as subject»)

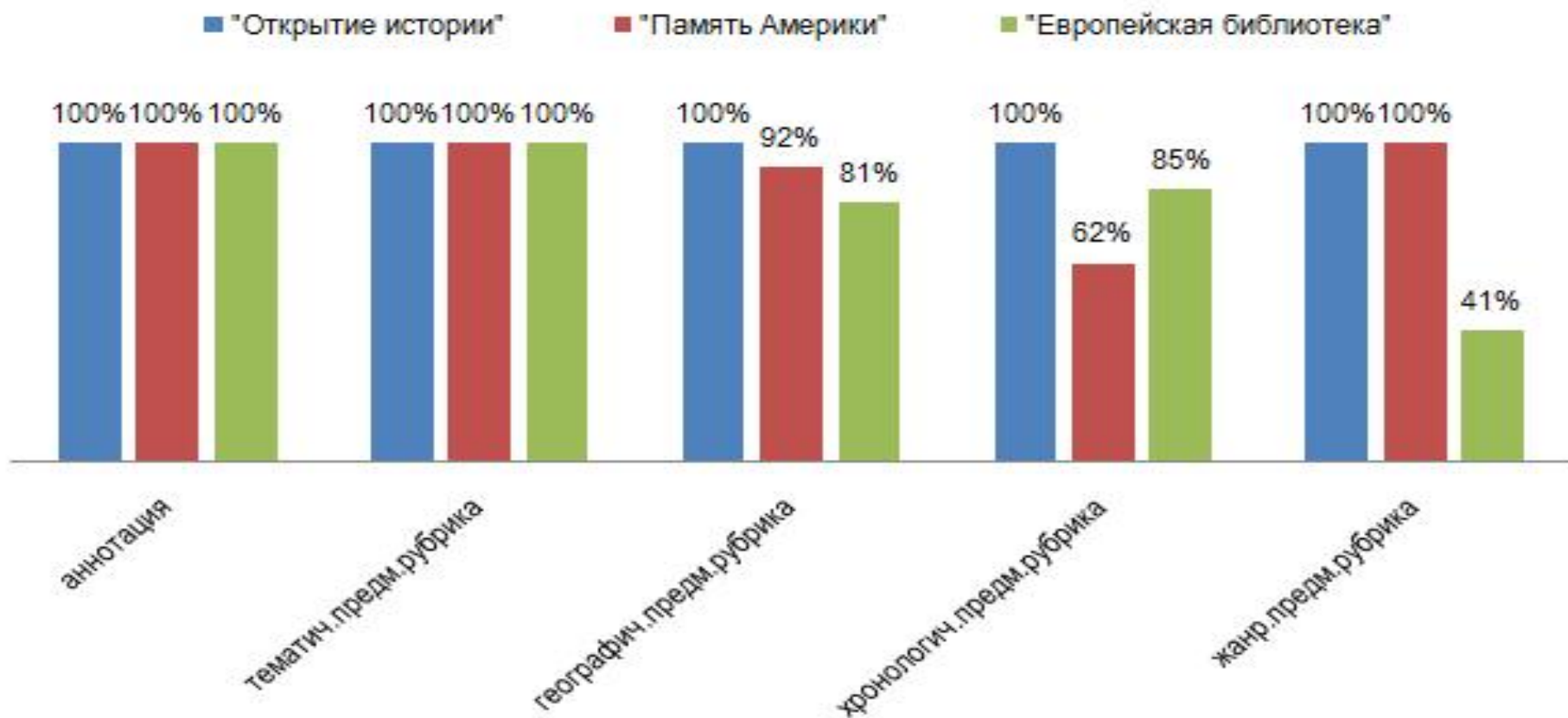
Результаты исследования

Качество КПМ

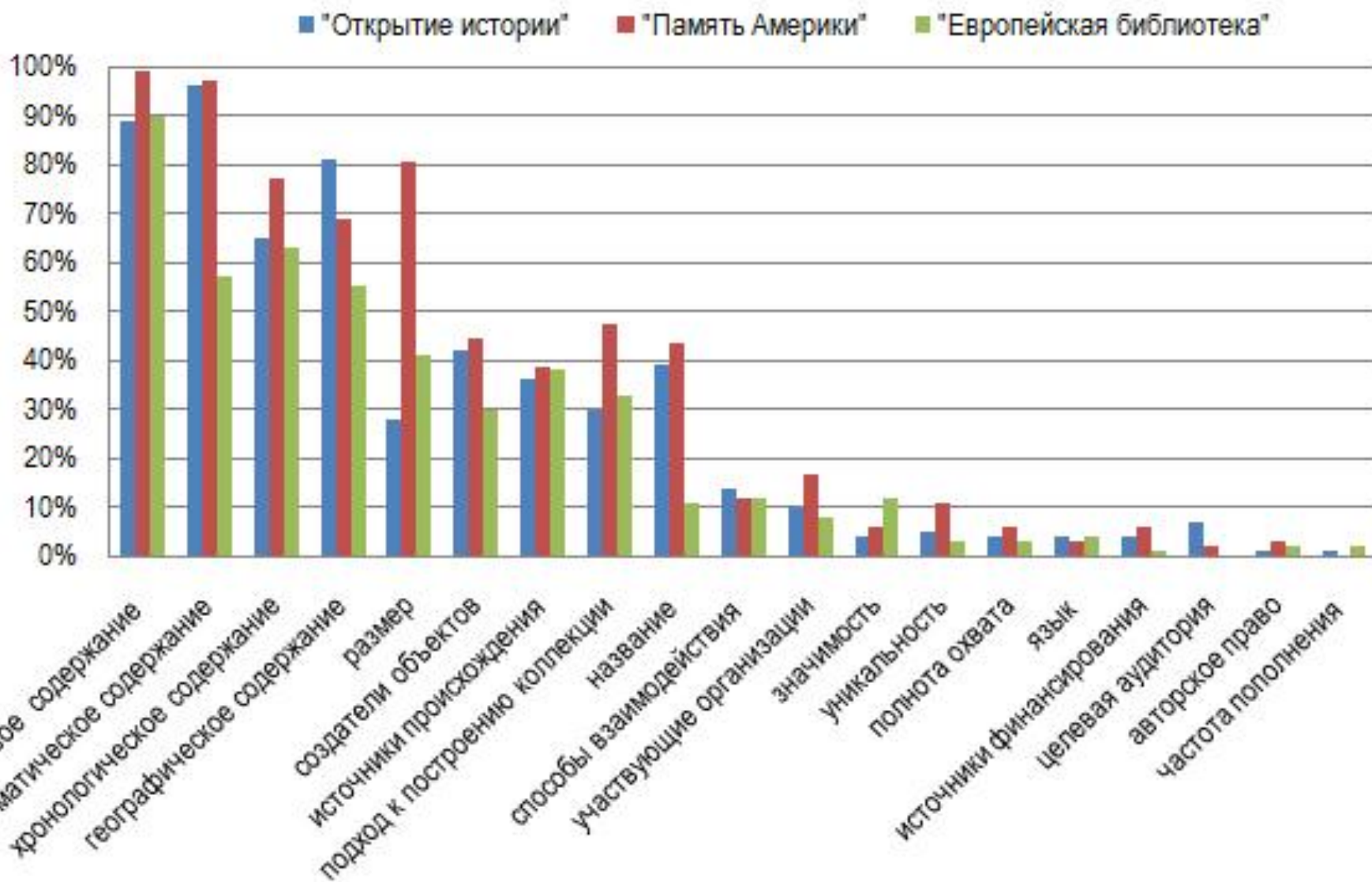
- оказывает существенное влияние на предметный доступ в порталах, объединяющих электронные коллекции
- **характеристики качества КПМ:**
 - постоянство применения основных элементов КПМ в библиографических записях
 - разнообразие характеристик электронной коллекции, представленных в свободнотекстовых метаданных
 - взаимодополняемость информации, представленной в двух видах КПМ:
 - свободнотекстовых КПМ
 - КПМ нормализованной лексики.

Постоянство применения элементов КПМ

- Стабильно высокое (100%) для аннотаций и тематических предметных рубрик
- Колеблется (41%-100%) для хронологических, географических и жанровых предметных рубрик



19 характеристик электронных коллекций, представленных в свободнотекстовых КПМ



Взаимодополняемость свободнотекстовых КПМ и КПМ нормализованной лексики

- **Односторонняя дополняемость** (свыше 70% библиографических записей в выборке)
 - информация, содержащаяся в элементе метаданных, дополняет информацию, содержащуюся в одном или нескольких других элементах
 - например, аннотация дополняет географическую и хронологическую предметные рубрики
- **Двусторонняя дополняемость** (около 40%)
 - наиболее часто (29% библиографических записей) дополняет друг друга информация в свободнотекстовой аннотации и тематической предметной рубрике
- **Дублирование** одной и той же информации разными элементами коллекционных предметных метаданных (6%)

Взаимодействие пользователей с порталами и метаданными: анализ транзакций

Вид взаимодействия пользователей с порталом	Количество случаев в выборке
Просмотр коллекционных метаданных	1760
Просмотр метаданных, описывающих отдельные информационные объекты	368
Просмотр коллекций по тематическому признаку	953
Просмотр коллекций по географическому признаку	533
Просмотр коллекций по жанровому признаку	487
Просмотр коллекций по создавшим коллекции организациям	311
Просмотр коллекций по названию	153
Просмотр отдельных информационных объектов	4388
Поиск на уровне коллекций	880
Поиск на уровне отдельных информационных объектов	1860

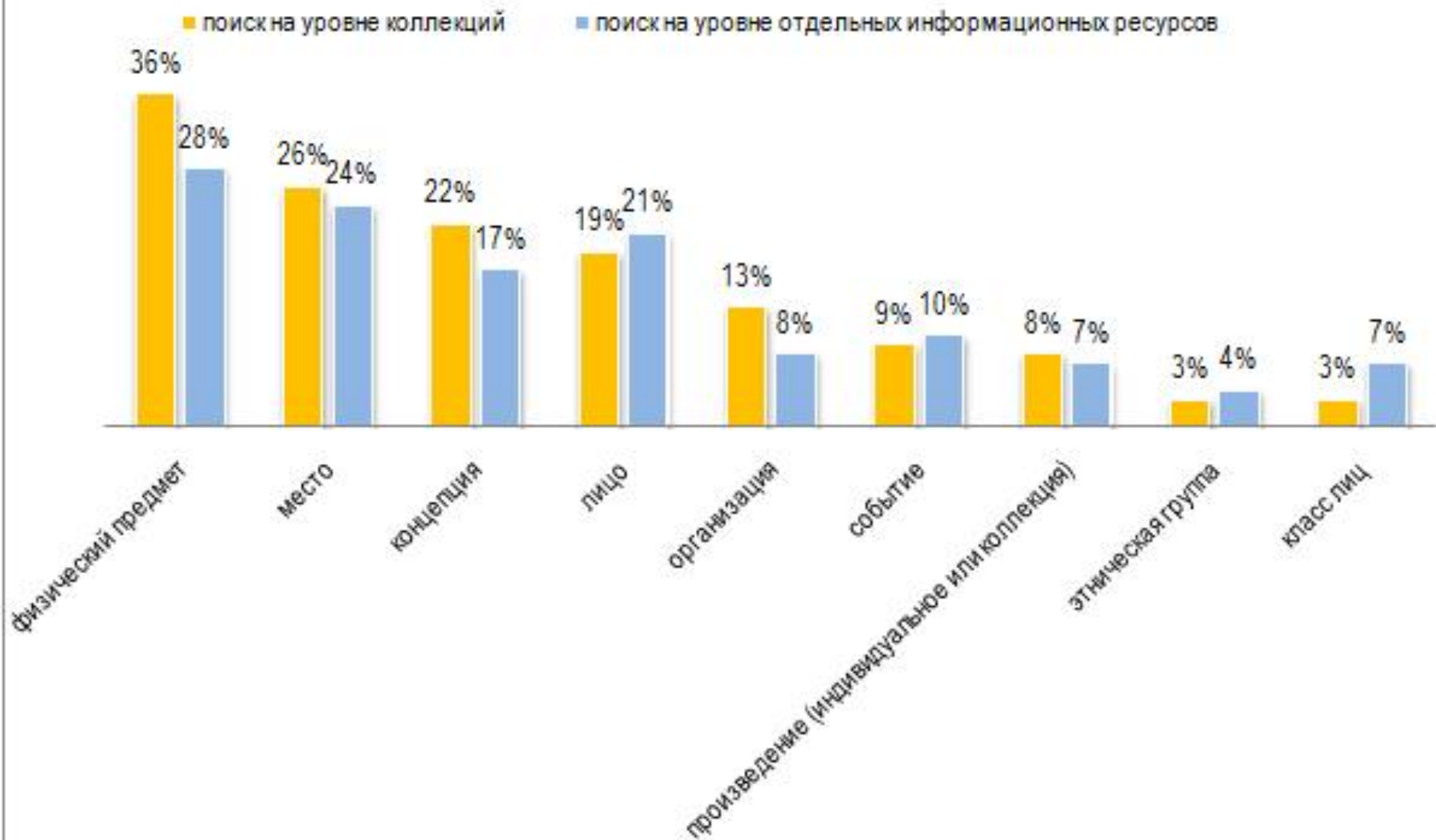
КПМ и результативность поиска в порталах

- Поиск на уровне коллекций наиболее часто (в 74% поисковых запросов) приносит результаты (т.е., находит одну или несколько электронных коллекций, соответствующих поисковому запросу) благодаря информации, представленной в **свободнотекстовых аннотациях**.
- Пользователи находят значительное число электронных коллекций (41%) исключительно благодаря **предметным рубрикам нормализованной лексики**:
 - хронологическим (13%)
 - географическим (12%)
 - тематическим (11%)
 - жанровым (5%).

Мнение пользователей

- В коллекционных метаданных особенно важна информация об:
 - источниках происхождения коллекции,
 - ее размере,
 - **жанровом, тематическом, географическом и хронологическом содержании коллекции.**
- Структурированные детальные коллекционные метаданные, содержащие множество элементов метаданных нормализованной лексики, предпочтительнее укороченных

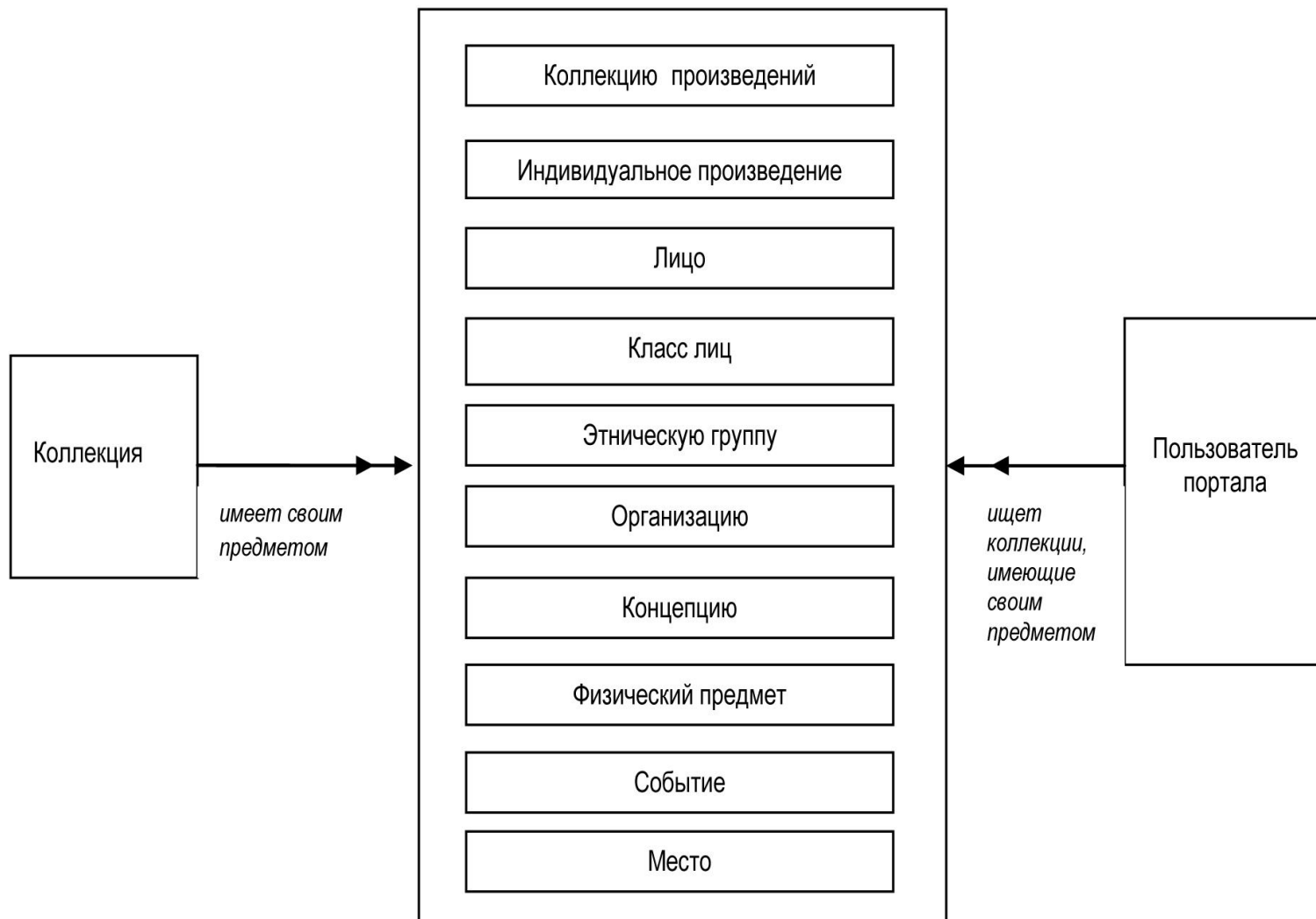
Модель ФРБР и поисковые запросы пользователей в порталах



Результаты исследования

- Предоставляют эмпирическое подтверждение ценности КПМ для пользователей порталов, объединяющих в себе электронные коллекции
- Раскрывают передовой опыт применения КПМ в порталах-первопроходцах, который создатели новых порталов могут принять в качестве ориентира
- Позволяют создать на основе модели ФРБР модель «объект-связь» предметного доступа в порталах на уровне коллекций

Модель «объект-связь» предметного доступа в порталах на уровне коллекций



Нерешенные вопросы на будущее

- Существует ли разница в применении КПМ между порталами
 - разного масштаба
 - международные
 - национальные
 - региональные
 - областные
 - разного тематического фокуса
 - гуманитарные науки
 - общественные науки
 - точные науки, технология и медицина
- Существует ли разница в информационном поведении пользователей порталов
 - разного масштаба
 - разного тематического фокуса?

Спасибо за внимание!
Дякую за увагу!
Thank you for attention!

Вопросы?
Питання?
Questions?

Список цитируемой литературы

- Cochrane, P. (1985). *Redesign of Catalogs and Indexes for Improved Online Subject Access: selected papers of Pauline A. Cochrane*. Phoenix, AZ: Oryx.
- Greenberg, J. (2005). Metadata and the World Wide Web. In *Encyclopedia of Library and Information Science*, pp. 1876-1888. New York: Marcel Dekker.
- Hjørland, B. (1997). The concept of subject or subject matter and basic epistemological positions. In *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science*. (pp. 55-103). Westport CT: Greenwood Press.
- Riesthuis, G., & Žumer, M. (2004). FRBR and FRANAR: subject access. *8th International ISKO Conference*.
<http://www.ucl.ac.uk/isko2004/sysweb/4bRiesthuisZumer.ppt>.
- Tillett, B. (2004). *What is FRBR?*
<http://www.loc.gov/cds/downloads/FRBR.PDF>.
- Zhang, Y., & Salaba, A. (2007, August/September). Critical issues and challenges facing FRBR research and practice. *Bulletin of the American Society for Information Science and Technology*.