

РАЗРАБОТКА ИНТЕРФЕЙСА ПОИСКА ИНФОРМАЦИИ В ЕДИНОМ ХРАНИЛИЩЕ МУЛЬТИМЕДИЙНЫХ ДАННЫХ ИЗ ПОЛЕВЫХ ЭТНОГРАФИЧЕСКИХ ЭКСПЕДИЦИЙ

Левшун Дмитрий Сергеевич

*младший научный сотрудник лаборатории проблем компьютерной безопасности
Санкт-Петербургского института информатики и автоматизации Российской академии наук
199178, РФ, Санкт-Петербург, 14-я линия В.О., д. 39
E-mail: levshun@comsec.spb.ru*

Чечулин Андрей Алексеевич

*старший научный сотрудник лаборатории проблем компьютерной безопасности
Санкт-Петербургского института информатики и автоматизации Российской академии наук
199178, РФ, Санкт-Петербург, 14-я линия В.О., д. 39
E-mail: chechulin@comsec.spb.ru*

DEVELOPMENT OF THE INFORMATION SEARCH INTERFACE IN UNIFIED REPOSITORY OF MULTIMEDIA DATA FROM FIELD ETHNOGRAPHIC EXPEDITIONS

Dmitry Levshun

*junior research fellow at Laboratory of Computer Security Problems
of the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science
199178, St. Petersburg, 14th Liniya, 39*

Andrey Chechulin

*senior research fellow, Ph.D. at Laboratory of Computer Security Problems
of the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science
199178, St. Petersburg, 14th Liniya, 39*

АННОТАЦИЯ

Данная работа посвящена разработке интерфейса поиска информации в едином хранилище мультимедийных данных из полевых этнографических экспедиций. Объектом исследования являются существующие подходы к организации процесса поиска, их основные достоинства и недостатки. В статье предлагается возможная архитектура системы поиска информации в едином хранилище мультимедийных данных, построенная на основе выбранных решений.

ABSTRACT

This paper is devoted to the development of the information search interface in unified repository of multimedia data from field ethnographic expeditions. The object of investigation is existing approaches to organize the search process, their advantages and disadvantages. The possible architecture of the information search system in the unified repository of multimedia data, built on the basis of selected solutions, is suggested.

Ключевые слова: единое хранилище мультимедийных данных; поисковый интерфейс; система поиска информации; этнографические экспедиции, полевые экспедиции.

Keywords: unified repository of multimedia data; search interface; information search interface; ethnographic expeditions; field expeditions.

Введение

В современном мире, цифровые средства записи являются неотъемлемой частью полевых этнографических экспедиций, позволяя антропологам, этнографам и фольклористам привозить разнообразные цифровые данные: аудио- и видеозаписи, фотографии, – объемы которых настолько значительны, что единственным способом их обработки и архивации за приемлемое время также становятся цифровые технологии.

При этом одной из ключевых проблем все также является организация хранения и поиска данных из полевых этнографических экспедиций. Доклад, посвященный проблеме организации хранения данных из этнографических экспедиций, был представлен авторами на Юбилейной XV Санкт-Петербургской Международной Конференции «Региональная Информатика-2016» [7]. Для решения проблемы поиска данных необходимо разработать подходы и инструменты для отображения и поиска мультимедийных

материалов, представляющих собой результаты прошедших этнографических экспедиций. Как правило, при работе с поисковыми системами такого типа, ключевым фактором является удобство поиска, т.к. если формирование запроса будет представлять собой достаточно сложную задачу, то исследователи, вероятно, откажутся работать с подобной системой [3].

Существующие подходы к поиску информации

В основе любого подхода к поиску информации в мультимедийных хранилищах данных лежит схема поисковой деятельности, направленная на решение основных этапов задачи поиска и ответов на следующие вопросы: *кто ищет, что ищет, где ищет, каким средствами ищет и каковы требования к результатам поиска* [1]. Рассмотрим основные этапы решения задачи поиска более подробно:

- **Формирование запроса:** важно отметить, что любой процесс поиска начинается не с наличия запроса, а с потребности в получении информации; этап формирования запроса представляет собой выражение данной потребности в форме некоторого запроса.

- **Определение зоны поиска:** прямое или косвенное указание расположения информационных ресурсов, интересных пользователю; позволяет ограничить пользователя от получения заведомо бесполезной информации.

- **Анализ запроса пользователя:** после того как потребность в получении информации была выражена в виде некоторого запроса, начинает работу система поиска, основная задача которой – создать на основе поискового запроса "понятные" системе поисковые образы (что искать) и сценарии (как искать), а также определить зону поиска (где искать).

- **Выполнение поиска:** на основе предыдущего этапа, поисковая машина уже "знает" что, где и как искать, т.е. может реализовать необходимые процедуры поиска; при этом важно отметить, что полученный на данном этапе результат "понятен" именно поисковой машине, а потому нуждается в дополнительной обработке для представления пользователю.

- **Анализ результатов поиска:** представляет собой оценку соответствия пользовательскому запросу каждого найденного на предыдущем этапе информационного объекта; при этом важно отметить, что от качества данной оценки зависит эффективность работы поисковой системы с точки зрения конечного пользователя [6].

- **Формирование и выдача результатов поиска:** данный этап основан на анализе результатов поиска, а также учитывает особенности восприятия информации человеком; при этом результаты поиска представляются в удобной для навигации форме, сопровождаются информацией о степени соответствия найденного ресурса пользовательскому запросу, а также содержат краткую характеристику их содержания [2].

- **Переформирование запроса:** данный этап необходим в ситуации, когда потребность в нахождении некоторой информации по итогам решения задачи поиска не была удовлетворена и необходимо выполнить вышерассмотренные этапы ещё раз.

Отметим, что возможность получения адекватного потребностям пользователя результата поиска определяется не только поисковыми технологиями, типами, объёмом и качеством доступных мультимедийных данных, качеством телекоммуникаций или вычислительными ресурсами системы. Не менее важным является отражение в структуре и постановке поисковых задач информации о том, *кто, что и где* ищет.

Так, например, особенности пользователя системы во многом определяют характер требований к результатам поиска, а также возможность со стороны поисковой системы удовлетворить его информационные потребности. При этом для описания пользователя поисковой системы выделяют: *априорный портрет пользователя, уточненный портрет пользователя и права доступа пользователя*.

Априорный портрет пользователя в контексте единого хранилища мультимедийных данных из полевых этнографических экспедиций определяется, прежде всего, функциональной ролью пользователя, которой соответствуют определенные цели поиска (сторонний исследователь и участник экспедиции будут нуждаться в различных информационных объектах [5, 9]). Таким образом, сформировав априорный портрет пользователя, можно заранее определить допустимые зоны поиска, а также, при возможности, упростить работу по формированию запроса и сценария поиска.

Уточненный портрет пользователя может быть сформирован только на основе обратной связи с пользователем (этап анализа результатов поиска), которая заключается в оценке пользователем результатов поиска как точных или не точных, полезных или бесполезных, избыточных, достаточных или недостаточных, и т.д.

Права доступа пользователя являются составной частью априорного портрета пользователя и учитываются при формировании зоны поиска. Например, некоторые информационные объекты могут быть доступны только ограниченному кругу лиц из соображений защиты авторского права.

В рамках работы с поисковой системой именно пользователь определяет, что именно он хочет найти. При этом поисковая машина соответствующим образом строит сценарий поиска, анализирует результаты и формирует поисковую выдачу. Существует несколько основных аспектов, отражающих информационные объекты, которые ищет пользователь: тип результатов поиска, тип содержательных требований к результатам поиска, требования к характеристикам информационных объектов.

По типу результатов поиска различают коллекции информационных объектов, сформированные вне поисковой машины (базы данных, базы знаний, сайты и т.д.); подборки информационных объектов, сформированные поисковой машиной; вторичные

информационные объекты (метаданные первичных информационных объектов).

Важно отметить, что типы содержательных требований к результатам поиска определяются не только желаниями конечного пользователя, но и возможностями поисковой системы. Наиболее распространены следующие варианты формирования содержательных требований: на основании свободного текста, на основании ключевых слов, на основании логических выражений, на основании определенных полей описания информационных объектов, на основании уникальных идентификаторов, на основании принадлежности к определенной коллекции информационных объектов, на основании типа информационного объекта, на основании связи информационного объекта с конкретным пользователем, на основании тегов, на основании временных или географических меток.

При этом существует два принципиально различных подхода (каждый из которых обладает определенными недостатками) к формированию пользователем содержательных требований: на основании предложенных поисковой системой вариантов и самостоятельно. Действия пользователя в рамках определенной системы предполагают небольшое число разных и понятных пользователю вариантов на каждом этапе выбора (при этом количество этапов выбора также должно быть минимальным). В противном случае, формирование запроса будет представлять собой достаточно сложную задачу, что вероятно приведет к отказу пользователя работать с данной системой. Самостоятельное формирование запроса, а именно небольшого по объему списка слов, отражающего интересы пользователя, – также непростая задача, требующая определенных навыков и опыта. Для организации поиска материалов в едином хранилище мультимедийных данных из полевых этнографических экспедиций было принято решение использовать комбинацию данных подходов, предоставляя пользователю возможность самостоятельно формировать запрос, а также ограничивать поисковую выдачу на основании предложенных поисковой системой вариантов.

Рассмотрим возможную архитектуру системы поиска информации в едином хранилище мультимедийных данных из полевых этнографических экспедиций, построенную на основе выбранных решений.

Архитектура интерфейса поиска информации

Для удовлетворения информационной потребности пользователя, интерфейс поиска информации в мультимедийных хранилищах данных должен быть

эффективным – точным, избирательным и выразительным [4]. Точность запроса определяет возможность найти по запросу требуемую информацию. Избирательность запроса определяет возможность получить только требуемую информацию и ничего кроме нее. Выразительность запроса определяет возможность детального описания информационной потребности.

В качестве результата поиска, как правило, предоставляются ссылки на мультимедийные данные или их коллекции, в которых имеется то, что искомое в рамках поискового запроса. На основе полученных отдельных ссылок и связанных с ними мультимедийных материалов система формирует поисковую выдачу для предоставления информации пользователю.

Разработка архитектуры интерфейса поиска осуществляется на основе системы управления контентом (CMS) Omeka [10]. Данная CMS представляет собой компьютерную программу, которая обеспечивает совместный процесс создания, редактирования и управления хранилищем данных. При этом CMS обеспечивает хранение данных, контроль целостности данных, соблюдение режима доступа к данным, а также предоставляет данные в виде, удобном для навигации и поиска. Таким образом, использование готового CMS-решения для управления хранением и поиском данных, позволяет при помощи средств CMS регулировать процесс построения архитектуры хранилища мультимедийных данных и интерфейса поиска, а также последующую их поддержку.

Рассмотрим разработанную архитектуру интерфейса поиска информации в едином хранилище мультимедийных данных из полевых этнографических экспедиций более подробно.

Просмотр коллекций мультимедийных данных (рисунок 1): предоставляет возможность просмотра всех коллекций материалов из полевых этнографических экспедиций (связанных между собой мультимедийных данных), отражающих определенное событие в рамках этнографической экспедиции. Интерфейс предполагает возможность сортировки коллекций по временным меткам (дата добавления), а также названию.

Кроме того, в рамках поисковой выдачи предоставляется только краткое описание коллекции, поэтому интерфейс предоставляет возможность перехода к более подробному описанию посредством ссылки.

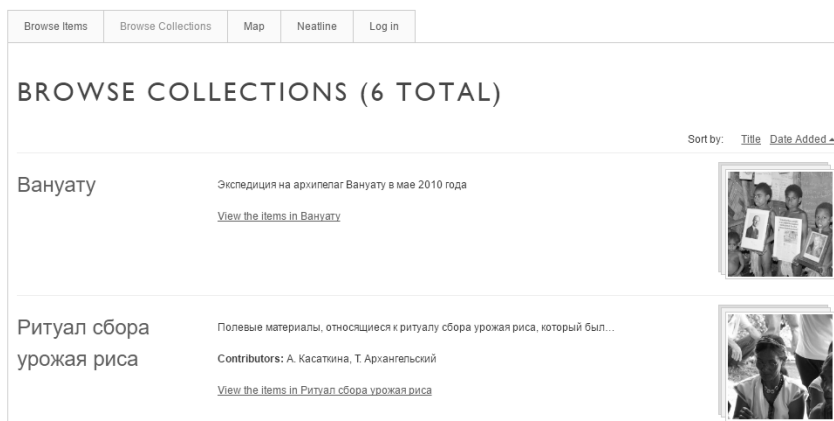


Рисунок 1. Просмотр коллекций мультимедийных данных

Просмотр мультимедийных данных (рисунок 2): предоставляет возможность просмотра всех мультимедийных данных из полевых этнографических экспедиций. Интерфейс предполагает возможность сортировки мультимедийных данных по временным меткам (дате добавления), участнику экспедиции и названию. Кроме того, в рамках поисковой выдачи предоставляется только краткое описание мультимедийного объекта, поэтому интерфейс

предоставляет возможность перехода к более подробному описанию посредством ссылки. Важно отметить, что количество мультимедийных данных, представленных на одной странице, ограничено. Поэтому для удобства навигации, поисковый интерфейс поддерживает возможность перехода между страницами поисковой выдачи (на следующую страницу, на предыдущую страницу, к определенной странице).

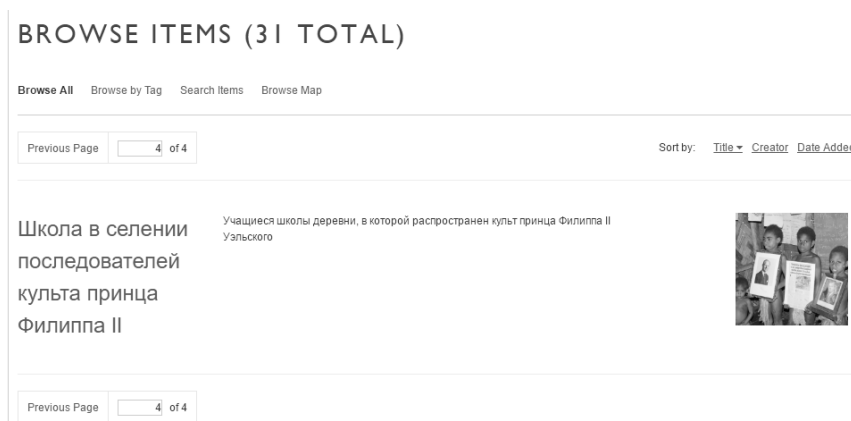


Рисунок 2. Поиск по всем мультимедийным данным

Поиск мультимедийных данных на основе тегов (рисунок 3): предоставляет возможность поиска мультимедийных материалов из полевых этнографических экспедиций, отмеченных специальным тегом. Для

удобства навигации, интерфейс поиска отображает теги в зависимости от частоты их использования (чем больше мультимедийных данных отмечены соответствующим тегом, тем больше его размер).

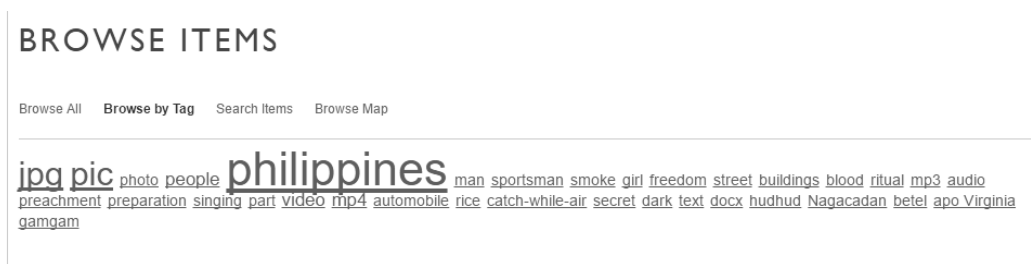


Рисунок 3. Поиск мультимедийных данных на основе тегов

При этом при переходе по конкретному тегу, пользователь перейдет к интерфейсу просмотра мультимедийных данных, поисковая выдача которого будет ограничена только отмеченными соответствующим тегом материалами.

Поиск мультимедийных данных на основе географических меток (рисунок 4): предоставляет

возможность поиска мультимедийных материалов на основе их географических меток на интерактивной карте мира. Отметим, что в рамках поисковой выдачи предоставляется только краткое описание выбранного мультимедийного материала, поэтому интерфейс предоставляет возможность перехода к более подробному описанию посредством ссылки.



Рисунок 4. Поиск мультимедийных данных на основе географических меток

При этом для удобства навигации количество отображаемых на интерактивной карте мультимедийных материалов ограничено, а поисковый интерфейс предоставляет возможность перехода между страницами поисковой выдачи (на следующую страницу, на предыдущую страницу, к определенной странице).

Поиск мультимедийных данных на основе требований (рисунок 5): предоставляет возможность поиска по ключевым словам, вводимым пользователями вручную, на основе логического выражения для ограничения поисковой выдачи на основании содержимого конкретных полей мультимедийных материалов (отношение принадлежности, наличия, соответ-

ствия и т.д.), по диапазону уникальных идентификаторов мультимедийных материалов, по принадлежности мультимедийных материалов к определенному событию, произошедшему в рамках этнографической экспедиции, по типу мультимедийного материала (текст, видео, аудио, звук), по участнику этнографической экспедиции, по тегам, вводимым пользователем вручную, и географической принадлежности. Кроме того, можно ограничить поисковую выдачу путем отображения только публичных (или наоборот) мультимедийных данных, а также путем отображения только рекомендованных (или наоборот) мультимедийных данных.

SEARCH ITEMS

Browse All Browse by Tag **Search Items** Browse Map

Search for Keywords

Narrow by Specific Fields

Language does not contain

Select Below Select Below

Add a Field

Search by a range of ID#s (example: 1-4, 156, 79)

Search By Collection

Ритуал сбора урожая риса

Search By Type

Text

Search By User

Andrey Chechulin

Search By Tags

Public/Non-Public

Only Public Items

Featured/Non-Featured

Only Featured Items

Geographic Address

Geographic Radius

10

(miles)

Search for items

Рисунок 5. Поиск мультимедийных данных на основе требований

При этом при запуске процесса поиска, пользователь перейдет к интерфейсу просмотра мультимедийных данных, поисковая выдача которого будет ограничена в соответствии с поисковым запросом.

Заключение

Для организации поиска материалов в едином хранилище мультимедийных данных из полевых этнографических экспедиций было принято решение использовать комбинацию подходов к формированию пользователем содержательных требований, предоставляя пользователю возможность самостоятельно формировать запрос, а также ограничивать поисковую выдачу на основании предложенных поисковой системой вариантов.

Разработанная архитектура интерфейса поиска информации в едином хранилище мультимедийных

данных из полевых этнографических экспедиций позволяет удовлетворить информационную потребность пользователя, т.к. полученный интерфейс поиска информации является точным, избирательным и выразительным.

В рамках дальнейших исследований, планируется заполнение единого хранилища мультимедийных данных материалами из полевого архива экспедиций, доступного руководителю и исполнителю проекта [8]. Кроме того, планируется оптимизация структуры единого хранилища и поисковой выдачи, а также расширение списка тегов на основе взаимодействия с участниками экспедиций, которые являются непосредственными авторами используемых мультимедийных материалов.

Список литературы:

1. Афонин А.А., Крейнс М.Г. Поиск образовательных информационных ресурсов: принципы, архитектура, реализация // Сборник научных статей "Интернет-порталы: содержание и технологии". Вып. 1. ГНИИ ИГТ "Информика". - М.: Просвещение, 2003. - С.584-634.
2. Донцов Д.О. Алгоритм генерации тезаурусных расширений для корпоративного информационного поиска // Труды СПИИРАН. 2013. Вып. 30. С. 189-203.
3. Касаткина А.К., Чечулин А.А. Мультимедийная база данных по фольклору Ифугао (Филиппины): постановка задач и выбор решений // Всероссийская научно-практическая конференция "Малочисленные этносы в пространстве доминирующего общества: практика прикладных исследований и эффективные инструменты этнической политики". г. Кемерово, 17–18 октября 2014 г. Сборник научных статей, г. Кемерово - ООО «Практика», 2014. С.312-317.
4. Курчинский Д.Н., Палей Д.Э., Смирнов В.Н. Электронная библиотека ВУЗа - как инструмент автоматического формирования учебных мультимедийных коллекций. // Ярославский государственный университет им. П. Г. Демидова. / [Электронный ресурс]. - Режим доступа: URL: <http://www.nsu.ru/xmlui/bitstream/handle/nsu/8905/12.pdf> (дата обращения: 29.03.2017).
5. Левшун Д.С., Чечулин А.А. Математическая модель описания информационных объектов разных типов для организации поиска материалов в едином хранилище мультимедийных данных из полевых этнографических экспедиций // Инновации в науке: сб. ст. по матер. LXIII междунар. науч.-практ. конф. – Новосибирск: СибАК, № 60, 2016. С.21-29.
6. Левшун Д.С., Чечулин А.А. Постановка задачи построения единого хранилища мультимедийных данных из полевых этнографических экспедиций // Журнал «Технические науки — от теории к практике». Изд. НП "СибАК", №46, 2015, с. 25-30.
7. Левшун Д.С., Чечулин А.А. Сравнение подходов к построению баз данных для организации поиска материалов в едином хранилище мультимедийных данных из полевых этнографических экспедиций // Юбилейная XV Санкт-Петербургская Международная Конференция «Региональная информатика-2016» («РИ-2016»). 26-28 октября 2016 г. Материалы конференции. СПб., 2016. С. 330-331.
8. Станюкович М.В. PĪLIPINAS MUNA! ФИЛИППИНЫ ПРЕЖДЕ ВСЕГО, к 80-летию Геннадия Евгеньевича Рачкова. Сер. "Маклаевский сборник" ответственный редактор и составитель М.В. Станюкович. // Издательство: Музей антропологии и этнографии им. Петра Великого (Кунсткамера) РАН, Санкт-Петербург, 2011, - С.500-510.
9. Gary Marchionini. Exploratory search: from finding to understanding. // Communications of the ACM, New York, USA. Volume 49 Issue 4. April 2006. P.41-46.
10. International site of open source Content Management System Omeka. / [Электронный ресурс]. - Режим доступа: URL: <https://omeka.org> (дата обращения: 29.03.2017).