

Міжнар. наук. конф., м. Київ, 4–6 жовт. 2016 р. – Режим доступу : <http://conference.nbuv.gov.ua/report/view/id/837>

13. Тарасенко Н. Розкриття бібліотечних фондів засобами соціальних мереж як фактор наповнення інформаційного простору [Електронний ресурс] / Н. Тарасенко // Наукові праці Національної бібліотеки України ім. В. І. Вернадського. – 2014. – Вип. 39. – С. 205–218. – Режим доступу : http://nbuv.gov.ua/UJRN/nbnbuimviv_2014_39_19

УДК 028:005.591.6

ВОЗМОЖНОСТИ АНАЛИЗА ЧИТАТЕЛЬСКИХ ИНТЕРЕСОВ В СОВРЕМЕННОЙ СИСТЕМЕ АВТОМАТИЗАЦИИ БИБЛИОТЕК

Панарин Геннадий Сергеевич,
ведущий программист ООО «ЭйВиДи-системы»,
главный библиотекарь Научной библиотеки Харьковского
национального технического университета сельского хозяйства
им. П. Василенко

Рассмотрены примеры анализа читательских интересов, возможных областей применения полученных данных. Названы основные проблемы и перспективы развития инструментов анализа спроса на документы из фонда, с которыми столкнулся автор.

Ключевые слова: САБ ИРБИС64, автоматизация библиотек, анализ данных, читатели, интересы читателей, кластер, вектор интересов, рекомендательные сервисы.

МОЖЛИВОСТІ АНАЛІЗУ ЧИТАЦЬКИХ ІНТЕРЕСІВ У СУЧАСНІЙ СИСТЕМІ АВТОМАТИЗАЦІЇ БІБЛІОТЕК

Панарин Геннадій Сергійович,
провідний програміст ТОВ «ЕйВіДі-системи»,
головний бібліотекар Наукової бібліотеки Харківського національного
технічного університету сільського господарства ім. П. Василенка

Розглянуто приклади аналізу читацьких інтересів, можливих сфер застосування отриманих даних. Указано основні проблеми та перспективи розвитку інструментів аналізу запитів на документи з фонду, із якими зіткнувся автор.

Ключові слова: САБ ІРБІС64, автоматизація бібліотек, аналіз даних, читачі, інтереси читачів, кластер, вектор інтересів, рекомендаційні сервіси.

POSSIBILITIES OF ANALYSIS OF READING INTERESTS IN THE MODERN AUTOMATIC SYSTEM OF THE LIBRARY

Panarin Gennady,

leading programmer of OOO Avid-systems,

Chief Librarian of Scientific Library

Kharkiv P. Vasylenko National Technical University of Agriculture

The article deals with examples of analysis of readers' interests, possible areas of application of the received data. The main problems and perspectives of developing instruments for analyzing the demand for documents from the fund, which the author encountered, is called.

Keywords: SAB IRBIS64, automation of libraries, data analysis, readers, readers interests, cluster, interest vector, advisory services.

Современная система автоматизации библиотечных (САБ) процессов – это мощный и многогранный инструмент, который создан для того, чтобы освободить персонал библиотеки от монотонной и сложной статистической, аналитической и технологической работы. Но далеко не всегда этот инструмент используется не то что бы полностью, но и вообще стоит вопрос о нерациональном его использовании. А ведь и сами САБ на данный момент чаще всего реализуют почти исключительно классические библиотечные технологии, почти не выходя за рамки общепринятых задач и технологий. Другими словами, в большинстве случаев САБ – это довольно продвинутый калькулятор на столе отдела комплектования и кафедры выдачи. Впору вспомнить о том, что забивать гвозди микроскопом можно, но вот КПД и экономический эффект такого мероприятия оставляют желать лучшего. Именно по этому, так интересно решать нетривиальные задачи, возникающие на стыке технологий. Двум примерам таких новых идей, появившихся у библиотекарей, я и хочу посвятить данную статью.

В процессе выполнения автоматизированного обслуживания читателей в САБ ІРБІС64 формируются электронные формуляры читателей, в которых хранится список всех выданных читателю книг. В предыдущих версиях САБ ІРБІС64 технология обслуживания имела существенное ограничение по количеству хранимых в

формуляре данных. В связи с этим ограничением многие библиотеки удаляли информацию о тех книгах, которые читатели уже вернули. В последних версиях появилась возможность архивировать эти данные, тем самым облегчая работу САБ с сохранением данных о многолетних интересах читателей. И этот массив интересов оказался настоящим кладом данных для анализа.

Выданная в САБ книга – это сформированная связь между записью с метаданными о выдаваемом документе и записью с метаданными о читателе. Это прямая связь между книгой и читателем. Она позволяет предположить, что данному читателю интересны документы из вашего фонда, написанные тем же автором, той же тематики, с теми же ключевыми словами. Уже этого достаточно, что бы в первом приближении начать рекомендовать читателю документы, которые входят в область его интересов. Проблема остается в том, что в большинстве случаев в САБ заносится весьма ограниченное количество данных, редко выходящих за пределы формирования «печатной каталожной карточки», и построенная модель интересов читателя по одной книге является не жизнеспособной.

Для решения этой проблемы достаточно обратиться к архиву выдаваемых книг и связать выданные книги через читателя. Это опосредованная связь документов уже может дать гораздо больше информации о том, что именно необходимо читателю. Для решения такой задачи я сформировал служебную базу данных, в которой для каждой пары выданных читателю книг сформировал связывающую запись. Эта запись в самом примитивном виде связывает два библиографических описания. Это первый шаг к рекомендации читателям литературы по интересующей их тематике. Уже эта сформированная связь может в будущем помочь другому читателю перейти от одной книги к другой. Для повышения объективности рекомендаций и отсеиванию случайных связей была выполнена обработка всех данных о выданных книгах. И это дало неожиданно точный результат. Обработка данных велась по следующему интуитивному алгоритму:

- 1) из записи читателя получался список всех прочитанных книг;
- 2) каждая из книг списка сопоставлялась с другой книгой из этого же списка;
- 3) при составлении пары книг производилась проверка на существование такой пары в базе анализа, в случае существования

пары ее коэффициент повышался на единицу, а при отсутствии создавалась новая запись с коэффициентом «1».

Такой примитивный алгоритм требует достаточно высоких затрат вычислительного времени, но он дает неплохой результат: в базе анализа появились миллионы пар книг, которые брались разными читателями и имели коэффициенты от единиц (такая пара книг выдавалась всего несколько раз, может считаться не существенной) до многих десятков и сотен (устойчивая связь между книгами, пара книг очень популярна).

Полученная база со связями между книг может стать замечательным подспорьем для огромного количества разнообразных анализов. Самым первым и естественным порывом становится использование этого массива для формирования рекомендательного сервиса для читателей. Ведь для этого достаточно организовать показ читателю книг-пар из приведенной базы с самыми высокими коэффициентами. Дальнейшая кластеризация документов может объединить книги, по группам даже в том случае, когда они сами не формировали между собой прямой пары связей. Для этого достаточно сформировать группы, переходя от одной книги в паре к другой и выполняя поиски других пар с этими же книгами и коэффициентами связи не ниже, чем у обрабатываемой пары. Говорить о порогах чувствительности здесь сложно. Самым простым вариантом будет сформировать кластеры либо от самых низких коэффициентов к высоким, либо от высоких к низким, а потом визуально просмотреть перечень полученных кластеров и принять решение о том уровне, на котором стоит остановиться.

Нормальной ситуацией является факт того, что читатели могут интересоваться книгами из разных кластеров – у каждого из нас много разных интересов. Но есть и еще один интересный аспект. База читателей это не только «что» читает читатель, но и «когда». Кроме самого факта его интересов можно анализировать перемещение его интересов из одной области в другую – вектор интересов. В том случае, если переход большого количества читателей из одной возрастной группы в другую сопровождается переходом его из одной группы интересов в другую, можно говорить об устойчивом векторе интересов. Пример естественного такого вектора: читатели студенты четко работают с группой учебных материалов в процессе обучения, переходя от одной группы тем и учебников к другим на протяжении многих лет обучения. Но ведь такие же векторы могут просматриваться и вне образовательных библиотек.

Другим интересным результатом формирования базы анализа оказался тот факт, что кластеризация документов по интересам читателей выявила сильное расхождение с действительностью качества, глубины систематизации и предметизации фонда: в каждой достаточно крупной группе с высокой степенью связей между книгами разделы знаний и предметные рубрики почти всегда совпадали или незначительно различались, и документы, которые оказались в этих группах, но сильно отличались по этим параметрам, чаще всего были ошибочно закодированы. Таким образом, сами читатели опосредованно выполняют роль экспертов в области содержания документов, помогая выявлять ошибки в предметизации. И тут нет ничего противоземного: практика – критерий истины.

А ведь тут же есть место и для работы отдела комплектования. Ведь каждый кластер, полученный при анализе, характеризует количество выданных книг, частотой их выдачи и группой читателей, которые этим кластером пользуются. Благодаря этому можно получить общую картину интересов читателей библиотеки, куда эти интересы двигаются.

Что же может быть дальше? Как минимум – создание полноценного рекомендательного сервиса для читателей и встраивание его в САБ. Это должно быть обязательно, если мы не хотим потерять читателей, которым нравится подход Google Amazon «возможно вы имели в виду» и «с этими документами часто просматривают». Для того, что бы предлагать массовому читателю именно то, что ему хочется, у нас есть все данные. Не хватает только алгоритмов и технологий. Но над этим работа ведется.

Другой важной проблемой является то, что в фондах библиотек есть большое количество книг-аналогов, либо книг, в значительной части совпадающих по содержанию. Каждая из этих книг в отдельности может привлекать незначительное количество интересов читателя. Но ведь читателя в первую очередь интересует не обложка и год издания, а те материалы и информация, которые под ними находятся. Таким образом, важным шагом к повышению релевантности кластеров интересов является формирование связей между книгами-аналогами – в дальнейшем связи должны строиться не по документам, а по их содержанию.

В связи с тем, что сам процесс анализа данных достаточно трудоемкий, а работает качественно только при большом количестве входных данных, идеальным направлением развития этой технологии была бы кооперация ряда разнопрофильных библиотек с

обслуживанием их выходных данных в рамках единой системы. Мы много уже осилили в области кооперации: сводные каталоги, корпоративную роспись, ЭДД и прочее. Пора подходить к следующему этапу коллективной обработки данных и почаще фантазировать о том, что бы еще сделать для читателей.

УДК 027.7:004.771

ІНФОРМАЦІЙНО-БІБЛІОТЕЧНИЙ СЕРВІС: ПЕРСПЕКТИВИ РОЗВИТКУ

Петленко Ксенія Юрївна,
студентка

Харківського державного університету харчування та торгівлі

Досліджено інноваційний розвиток бібліотеки вищого навчального закладу на основі комунікації студентів та науковців. Звернуто увагу на розвиток інформаційно-бібліотечного сервісу наукової бібліотеки ХДУХТ та розглянуто перспективи сучасного функціонування бібліотеки на підґрунті інституціонального репозитарію.

Ключові слова: наукова бібліотека вищого навчального закладу, інституціональний репозитарій, наукова комунікація.

ИНФОРМАЦИОННО-БИБЛИОТЕЧНЫЙ СЕРВИС: ПЕРСПЕКТИВЫ РАЗВИТИЯ

Петленко Ксения Юрьевна,
студентка

Харьковского государственного университета питания и торговли

Исследовано инновационное развитие библиотеки высшего учебного заведения на основе коммуникации студентов и ученых. Уделено внимание развитию информационно-библиотечного сервиса научной библиотеки ХГУПТ и рассмотрены перспективы современного функционирования библиотеки на основе институционального репозитария.

Ключевые слова: научная библиотека высшего учебного заведения, институциональный репозитарий, научная коммуникация.