

ДОСТУПНОСТЬ ТЕХНОЛОГИЙ В НАУКОВЕДЧЕСКИХ ИССЛЕДОВАНИЯХ

Д.Е. Прокудин

Санкт-Петербургский государственный университет

Университет ИТМО

Санкт-Петербург

Развитие информационного общества приводит к постепенной информатизации различных научных дисциплин и направлений. Применение подходов и методов научных исследований, основанных на использовании информационно-коммуникационных технологий (ИКТ), привело к формированию междисциплинарного направления «Цифровая гуманитаристика» (Digital Humanities), которое активно развивается и в России. От этого направления берёт своё начало историческая информатика, которая ориентирована на использование компьютерных методов и информационных технологий в исторических исследованиях [1][2][3]. Эти же подходы и методы всё чаще используются и в науковедческих исследованиях, исследованиях по истории науки и научного наследия. К основным устоявшимся направлениям цифровой гуманитаристики [6], используемым в этих исследованиях, можно отнести:

- отбор цифровых информационных ресурсов;
- отбор и извлечение информации (данных) из цифровых информационных ресурсов;
- количественная и качественная обработка информации (данных), полученной из отобранных источников в соответствии с целями и задачами исследования;
- визуализация полученных результатов обработки данных, в том числе связанная с пространственным представлением (например, представление в геоинформационных системах (ГИС)).

В свою очередь, реализация этих направлений базируется на:

- доступе к цифровым информационным ресурсам;
- использовании специализированного программного обеспечения или других программных средств (например, сетевых и облачных сервисов) для извлечения, обработки и визуализации данных.

Ценностью любого исследования является, в том числе, возможность его воспроизводимости и, соответственно, тиражируемости, т.е. возможность использования подходов, методов и инструментов другими учёными и исследователями. Иначе говоря, одним из основных принципов, которые должны быть положены в основу выбора подходов, методов и инструментов проведения исследования, является принцип доступности. Такой подход к организации и проведения научных исследований полностью вписывается в парадигму «Открытой науки» [4, 7]. В соответствии с этим можно выделить несколько основных уровней доступности:

- доступность, основывающаяся на свободно распространяемой информации. В этом плане прежде всего выделяются цифровые информационные ресурсы открытого доступа [5], которые предоставляют открытый доступ к научным публикациям;
- доступность программного обеспечения, которое является свободно распространяемым (Open Source), или различных web-ориентированных информационных проектов, доступных по публичным лицензиям;
- доступность, основанная на достаточной простоте использования, т.е. для применения соответствующих технологий и инструментов нет необходимости привлекать специалистов или обладать специальными знаниями в области информационно-коммуникационных технологий, что позволяет использовать их любым исследователем-гуманитарием без специальной подготовки.

В рамках проводимого исследования по применимости методов и подходов цифровой гуманитаристики в исследованиях научного наследия Г.Ф. Гаузе как основной учитывался именно принцип доступности, что позволило отобрать методы, технологии и инструменты с учётом доступности их применения другими исследователями в своих научных целях. Кроме этого, такой подход позволил систематизировать и сгруппировать цифровые информационные ресурсы, системы и программные приложения по признаку доступности.

Цифровые информационные ресурсы можно разделить на:

1. Проприетарные (коммерческие). Доступ к ним и, соответственно, к данным (например, полным текстам) осуществляется по коммерческой подписке для организаций, сотрудники которых пользуются ими в своей работе бесплатно. К таким ресурсам относятся: полнотекстовые базы научной информации (JSTOR, Academic Search Ultimate и пр.), цифровые платформы ведущих мировых научных издательств (Springer, Elsevier, SAGE, Taylor & Francis и пр.), реферативные базы данных (Scopus, WoS). Необходимо также отметить, что большинство этих ресурсов позволяют свободно (без подписки) осуществлять поиск научных публикаций, получать доступ к их метаданным, а также обладают инструментами автоматического экспорта метаданных в стандартных форматах. Также отличительной особенностью коммерческих ресурсов является

наличие (хоть и в небольшом количестве) публикаций открытого доступа – так называемый «золотой» открытый доступ (Gold Open Access).

2. Открытого доступа (представляют категорию Green Open Access). К ним в основном относятся институциональные и тематические репозитории (архивы открытого доступа). Среди российских ресурсов можно выделить проект открытой научной библиотеки «Киберленинка» (<https://cyberleninka.ru>), однако у этого ресурса достаточно ограниченный поисковый механизм и отсутствует возможность выгрузки метаданных публикаций.

3. Смешанные, которые размещают публикации в открытом доступе, на коммерческой основе, только метаданные публикаций и ссылки на публикации, размещённые на сторонних ресурсах. Как яркий пример можно привести Научную электронную библиотеку (<https://elibrary.ru/>), являющуюся самым крупным научным цифровым информационным ресурсом в России. Она обладает мощными поисковыми и аналитическими инструментами, но у неё отсутствуют механизмы автоматической выгрузки метаданных и пакетной выгрузки публикаций, размещённых в открытом доступе.

Такой же подход применим и классификации программного обеспечения, которое можно разделить на следующие категории:

1. Коммерческое, использование которого доступно далеко не каждому учёному или исследователю. Например, к основному инструментарию, используемому в том числе и при проведении науковедческих исследований, можно отнести популярный офисный пакет MS Office, библиографический менеджер EndNote, систему статистической обработки данных SPSS.

2. Условно бесплатное, производители которого предоставляют базовый функционал бесплатно, а расширенный – на коммерческой основе. Для индивидуальной работы исследователя вполне достаточным бывает базовый бесплатный функционал. Как примеры можно перечислить: библиографический менеджер Mendeley, система подготовки и демонстрации мультимедийных презентаций Prezi, веб-ориентированная система подготовки диаграмм Draw.io, пакет разработчика с локальным веб-сервером MAMP.

3. Бесплатное, распространяемой по различным публичным лицензиям (например, GPL, Apache, Creative Commons и пр.) – офисный пакет Libre Office, веб-ориентированный набор инструментов анализа текстов Voyant-tools, приложение для анализа и визуализации библиографической информации VOSviewer, система обработки и визуализации геоданных Mapbox и т.п.

Также при использовании различных информационных систем и платформ необходимо учитывать наличие у них интерфейса прикладного программирования (application programming interface, API), который позволяет использовать их данные или функционал в других прикладных программных системах. Например, через API сетевой базы CrossRef возможно либо выгрузить метаданные публикаций (у которых есть DOI, Digital Object Identifier) в один из стандартных форматов представления, либо автоматически загрузить для обработки в приложение анализа и визуализации библиографической информации VOSviewer. А через API сервиса Mapbox можно создать карту средствами базовой версии пакета MAMP с динамической кластеризацией геоданных.

Учёт принципа доступности при проведении исследования позволил отобрать и использовать только те цифровые информационные ресурсы, системы, платформы и программные приложения, которые не требуют при их освоении и использовании специализированных знаний в области информационных технологий или привлечения соответствующих специалистов. Такие инструменты обладают простым в использовании и дружелюбным по отношению к пользователю интерфейсом, а также подробной документацией и примерами использования.

Однако, для того, чтобы принцип доступности в полной мере позволял использовать цифровые информационные ресурсы, системы, платформы и программные приложения в науковедческих исследованиях широким кругом учёных и исследователей, необходимо, чтобы каждый учёный при представлении результатов своих исследований в том числе описывал используемые им инструменты цифровой гуманитаристики и методики их применения.

ЛИТЕРАТУРА

1. Бородкин Л. И., Гарскова И. М. Историческая информатика: перезагрузка? // Вестн. Перм. ун-та. Сер. История. 2011. №2 (16). С. 5-11.
2. Владимиров В. Н. Историческая информатика: пути развития // Вестник ТГПУ. 2006. №1.
3. Гарскова И. М. Основные направления развития исторической информатики в конце XX-начале XXI в // Вестник Московского университета. Серия 8: История. 2010. №. 6. С. 75-103.
4. Паринов С.И. Развитие электронных библиотек – путь к Открытой Науке // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всероссийской научной конференции RCDL'2009. Петрозаводск: КарНЦ РАН, 2009. С. 225-234. URL: http://rcdl.ru/doc/2009/225_234_Invited-2.pdf

5. Цветкова В.А., Павловска Е. И снова об открытом доступе к информационным ресурсам как перспективной модели распространения научного знания // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. – М.: ГПНТБ России, 2012. URL: <http://www.gpntb.ru/win/inter-events/crimea2012/disk/027.pdf>.
6. Arthur P. Exhibiting history: the digital future // *Journal of the National Museum of Australia*. 2008. Vol. 3, № 1.
7. David P. A. Understanding the emergence of 'open science' institutions: functionalist economics in historical context // *Industrial and Corporate Change*. 2004. Vol. 13, Iss. 4. P. 571—589. DOI:10.1093/icc/dth023.