

О создании Межведомственной библиографической лаборатории для исследования дальнего чтения

Леонов В.П.,

доктор педагогических наук, профессор,
Библиотека Российской академии наук,
valleo@mail.ru

Анотация. Принципиальное значение в работах по цифровому анализу текстов (Digital Humanities — DH) приобрела концепция «дальнего чтения» (Distant Reading — DR), предложенная профессором Стэнфордского университета США Ф. Моретти. В докладе рассматривается возможность на основе методологии дальнего чтения изучать большие массивы библиографических данных с использованием формализованного анализа, количественных и библиографических методов. Для координации и планирования работы специалистов гуманитарных и технических областей необходимо создание Межведомственной библиографической лаборатории. Она может быть организована в учреждениях РАН с участием крупнейших библиотек страны, где сосредоточены разнообразные информационные ресурсы.

Ключевые слова: дальнее чтение, Межведомственная библиографическая лаборатория, Франко Моретти, цифровой анализ текстов, библиографическая информация, методология дальнего чтения.

On Establishing the Interagency Bibliographic Lab for Study the Distant Reading

Leonov V.P.,

Dr. Sci. (Pedagogy), professor, Russian
Academy of Sciences Library,
valleo@mail.ru

Abstract. The concept of distant reading developed by Professor Franco Moretti (Stanford University, USA) is vital in publications on digital humanities (DH). The present article considers the possibility, based on distant reading methodology, to study large amount of bibliographical data using formal method, quantitative and bibliographical methods of analysis. To coordinate and plan the work of specialists in humanitarian and technical fields it is necessary to establish the Interagency bibliographical lab. It can be established in the Russian Academy of Sciences institutions with participation of the Russian largest libraries where there is a concentration of various information resources.

Keywords: distant reading, Interagency Bibliographic Lab, Franco Moretti, digital humanities, bibliographic information, distant reading methodology

DOI: 10.31432/1994-2443-2020-15-3-40-44

Цитирование публикации: Леонов Валерий Павлович. О создании Межведомственной библиографической лаборатории для исследования дальнего чтения // Информация и инновации. 2020, Т.15, № 3. с. 40–44. DOI: 10.31432/1994-2443-2020-15-3-40-44

Citation: Leonov Valerii Pavlovich. On establishing the Interagency bibliographic lab for study the distant reading // Information and Innovations 2020, T.15, № 3. p. 40–44. DOI: 10.31432/1994-2443-2020-15-3-40-44

Активное использование в гуманитарных отраслях, в том числе в социологии, литературоведении, книговедении, библиографоведении, библиографии, цифровых и автоматизи-

рованных технологий, а также необходимость решения прикладных задач для изучения обстоятельств и фактов их истории и теории привели к появлению большого числа ресурсов

различного назначения. Деятельность по их формированию, обеспечению к ним доступа, сохранности, возможность многократного использования требует координации и управления [1–3].

Принципиальное значение в работах по цифровому анализу текстов, известное под общим названием Digital Humanities — DH¹, приобрело понятия «дальнее чтение» (Distant Reading). Его ввел в науку о литературе профессор Стэнфордского университета США Франко Моретти [4]. В качестве дополнения к пристальному чтению (close reading), предпочитаемому вдумчивое чтение конкретного текста, Моретти предложил новый подход, ориентированный на исследование доступного корпуса текстов с использованием компьютерных методов. «...Решающее значение, — пишет он, — имело возращение важности количественных методов для моих исследований в Стэнфорде, которое в конце концов привело к созданию Литературной лаборатории в 2010 году... Количественные данные... представили такое огромное количество эмпирического материала, что я оказался к нему совершенно не готов... Сейчас ... результаты этих опытов начинают обретать форму, и период отсутствия теории заканчивается. На самом деле мир Digital Humanities, цифровых гуманитарных наук, понемногу начинает ощущать потребность в обобщающей теории для нового литературного архива» [4, с. 175–176].

Как отмечают переводчики книги «Дальнее чтение» О. Собчук и А. Шеля, «работы Франко Моретти возродили интерес к макроскопическому взгляду на литературу. Это значит: объектом исследования становится не одно или несколько произведений, как раньше, а десятки, сотни и даже тысячи. В этом океане информации (океане не только букв, но и цифр) Моретти ищет закономерности» [5, с. 88].

Междисциплинарность дальнего чтения проявляется в сотрудничестве специалистов разных гуманитарных и технических областей знания. Помимо историков, философов, социологов, филологов и специалистов по информационным технологиям, в таких исследованиях уже успешно принимают участие библиотекари,

библиографы, архивисты, музейные работники. А проблемы, которые поднимает Ф. Моретти, оставляют возможность для выбора. Говоря о междисциплинарных исследованиях, положенных в основу DH, логично полагать, что они станут эффективными, если будут иметь общий объект, на который направлены методы различных наук [6–8]. В нашем случае в качестве такого объекта предлагаю рассматривать *библиографическую информацию*, имеющую сложную видовую структуру. Полагаю, что настало время в дополнение к литературоведческим исследованиям изучать библиографическое дальнее чтение, под которым понимаются способы представления автоматизированной обработки библиографической информации. Этот объект объединит формализованный анализ, количественные и библиографические методы [9].

Следует иметь в виду, что в начале междисциплинарного сотрудничества каждый из представителей наук остается как бы на своей территории и только со временем сам объект изменяет его знание и методологию. Иногда такого времени требуется достаточно много, прежде чем специалист окажется «в проблеме» [10]. «...Мне кажется, — подчеркивает филолог И. Пильщиков, — я после многих лет работы в той сфере, которую сейчас начали называть Digital Humanities, стал своего рода специалистом «по переговорам» между филологами и программистами. Я помогаю коллегам с двух сторон этого дигитального «фронта» найти взаимопонимание [11, с. 31–32].

И далее на вопрос о поддержке государственных и профессиональных исследований в оцифровке И. Пильщиков убедительно пояснил: «... Когда государство начинает что-то поддерживать, оно регламентирует... В России был период, когда ... создание электронных библиотек поддерживали и государство, и научные фонды, государственные и частные... рубеж столетий был кратковременным золотым веком для направления Digital Humanities... Но затем поддержка крайне сузилась... В результате профессиональные электронные библиотеки оказались практически брошены без системной поддержки ... исключения единичны» [11, с. 33–34].

Используя подход и опыт Ф. Моретти по изучению фактов и истории на основе мето-

¹ Цифровые (компьютерные) методы в гуманитарных науках.

логии дальнего чтения, в рамках библиографии целесообразно создание службы управления оцифровкой. В качестве первого шага это может быть Межведомственная библиографическая лаборатория [12]. Предлагается начать работу по ее созданию в учреждениях РАН с участием крупных библиотек страны, где сосредоточены самые разнообразные информационные ресурсы.

Одна из результативных попыток классификации информационных ресурсов гуманитарного профиля, выявленных в процессе мониторинга академических организаций, представлена в статье А.Б. Антопольского [13]. Статистика показала, что электронные версии изданий различаются глубиной ретроспективы и полнотой: полными текстами, аннотациями, содержанием выпусков, сводными указателями. Таким образом, одно и то же издание может быть представлено на сайте учреждения и в Интернете в разных формах и с разной полнотой.

Приведу некоторые примеры. Современный опыт создания цифровых коллекций представлен в проекте Института мировой литературы МГУ им. М.В. Ломоносова «Сравнительная поэтика и сравнительное литературоведение» для изучения взаимосвязи русской и зарубежной поэзии. Уже сейчас пользователям доступно более двух тысяч стихотворений, 60 научных изданий. Общее число авторов, чьи произведения представлены в системе, — 180 исследователей и 480 поэтов и переводчиков. Суть инновации проекта — в способах их представления [14].

Всю информацию представляют четыре взаимосвязанных подсистемы:

- 1) Корпус параллельных текстов;
- 2) Цифровая библиотека комментированных изданий поэтических переводов и их оригиналов, а также книг и статей по сравнительной поэтике;
- 3) Энциклопедия «Систематизированные библиографические сведения о поэтах, переводчиках и исследователях-компаративистах»;
- 4) Тезаурус «Структурированный глоссарий терминов, встречающихся в научной литературе с описанием их значения и примерами их употреблений».

Интерфейс системы реализован на трех языках — русском, английском и испанском.

Цель проекта — объединить возможности цифровых библиотек и параллельных корпусов и превратить полученный инструмент в семантизированный гипертекст, в «семантическую сеть».

Успешную попытку применить статистические методы дальнего чтения (Ф. Моретти) для анализа заглавий книг осуществила аспирант Высшей школы печати и медиаиндустрии Московского политехнического университета Е. А. Тутатина. Суть ее эксперимента состояла в следующем. Из отобранных 5165 заглавий была составлена библиографическая база, включающая название книги, имя автора, год издания и жанр (роман, рассказ, повесть). Книги распределялись по направлениям: детектив, сентиментальная проза, фэнтези. Проверке подвергались издания, которые за последние 20 лет входили в списки бестселлеров газеты «The New York Times» (за исключением произведений А. Кристи и А. Конан Дойла).

Гипотеза исследования заключалась в том, что заглавия бестселлеров формируют «моду» на заглавия в издательском процессе. Как только книга становится бестселлером, ее переводят на множество языков, появляются экранизации, и тогда заглавие начинают копировать. Результаты анализа представлены в виде таблиц и содержат интересную информацию для использования в издательской деятельности, т.е. позволяют понять, как заглавие влияет на популярность книги, какие заглавия являются модными и какие могут возникать при этом ошибки [15].

...Межведомственная библиографическая лаборатория по изучению дальнего чтения может быть построена по плану создания единого цифрового пространства научных знаний (ЕЦПНЗ), предложенному А.Б. Антопольским и коллегами [1–2]. Ее структура состоит из двух частей — постоянная и динамическая (базис и надстройка). Постоянная (базис) формируется из имеющихся в России ресурсов — библиотечных, библиографических, архивных, музейных, энциклопедических, словарных, т.е. базис вырабатывается из существующих систем, прошедших экспертизу профессионалов. Динамическая (надстройка) часть предполагает анализ новой информации, отобранной

экспертами, которую необходимо изучать. Кроме этого, проект формирования Межведомственной библиографической лаборатории должен координироваться специальным советом из ведущих ученых, определяющих этапы ее создания и поддержки. Имеются в виду трудовые и финансовые затраты на поиск, оцифровку, сохранение и использование библиографической информации, полученной методами Digital Humanities.

Если коллеги поддержат идею создания Межведомственной библиографической лаборатории, то результаты ее деятельности могут быть впечатляющими.

ЛИТЕРАТУРА

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестн. РАН. 2019. Т. 89. № 7. С. 728–734.
2. Антопольский А.Б. О создании центра лингвистических ресурсов РАН // Известия Российской академии наук. Сер. литературы и языка. 2019. Т. 78, № 4. С. 5–12.
3. Герасименко А.Ю. Формирование единого информационного пространства научной библиотеки // Библиосфера. 2019. № 4. С. 78–84.
4. Моретти Ф. Дальнее чтение / Ф. Моретти; пер. с англ. А. Вдовина, О. Собчука, А. Шели. Москва: Изд-во Института Гайдара, 2016. 344 с.
5. Собчук О. (Дальнее) чтение и (культурная) эволюция / О. Собчук, А. Шеля // НЛО. 2018. № 2 (150). С. 88–93.
6. Таллер М. Дискуссии вокруг Digital Humanities // Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании. 2012. № 1. С. 5–13.
7. Володин А.Ю. Digital Humanities (цифровые гуманитарные науки): в поисках самоопределения // Вестник Пермского университета. Сер.: История. 2014. № 3 (26). С. 5–12.
8. Можая Г.В. Digital Humanities: цифровой поворот в гуманитарных науках / Г.В. Можая // Гуманитарная информатика: сб. статей / под ред. Г.В. Можайевой. 2015. Вып. 9. С. 8 — 23.
9. Леонов В.П. Вавилонская библиотека в контексте дальнего чтения // Библиография и книговедение. 2020. № 1. С. 77–83.

10. Соколова Е. В. Истолковать и/или исчислить: о возможности комбинированного метода на стыке герменевтики и digital humanities / Е. В. Соколова // Современная наука о литературе: сб. ст. / РАН, ИНИОН. Москва, 2018. С. 142–155.

11. Маяцкий М. Digital Humanities — это что-то новое или мы уже давно этим занимаемся? Интервью с Игорем Пильщиковым // Логос. 2015. Т. 25, № 2 (104). С. 14–36.

12. Леонов В.П. Дальнее чтение как стратегия точного библиографоведения // Научные и технические библиотеки. 2019. № 10. С. 56–67.

13. Антопольский А.Б. О классификации информационных ресурсов академических организаций социогуманитарного профиля // Взаимовлияние информационно-библиотечной среды и общественных наук: сб. ст. / РАН, ИНИОН. М., 2018. С. 48–61.

14. Полилова В.С. Цифровое будущее истории мировой литературы // Троицкий вариант — Наука. 2019. № 25. С. 14.

15. Тутатина Е.А. Заглавия книг: современные тенденции в книгоиздании // Библиография и книговедение. 2020. № 1. С. 111–119.

REFERENCES

1. Antopol'skii A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N. [O edinom cifrovom prostranstve nauchnykh znaniy // Vestn. RAN. 2019. T. 89. № 7. S. 728–734.
2. Antopol'skii A.B. O sozdanii centra lingvisticheskikh resursov RAN // Izvestiya Rossijskoj akademii nauk. Ser. literatury i yazyka. 2019. T. 78, № 4. S. 5–12.
3. Gerasimenko A.YU. Formirovanie edinogo informacionnogo prostranstva nauchnoj biblioteki // Bibliosfera. 2019. № 4. S. 78–84.
4. Moretti F. Dal'nee chtenie / F. Moretti; per. s angl. A. Vdovina, O. Sobchuka, A. Sheli. Moskva: Izd-vo Instituta Gajdara, 2016. 344 s.
5. Sobchuk O. (Dal'nee) chtenie i (kul'turnaya) evolyuciya / O. Sobchuk, A. Shelya // NLO. 2018. № 2 (150). S. 88–93.
6. Taller M. Diskussii vokrug Digital Humanities // Istoricheskaya informatika. Informacionnye tekhnologii i matematicheskie metody v istoricheskikh issledovaniyah i obrazovanii. 2012. № 1. S. 5–13.
7. Volodin A.YU. Digital Humanities (cifrovye gumanitarnye nauki): v poiskah samoopredeleniya

// Vestnik Permskogo universiteta. Ser.: Istoriya. 2014. № 3 (26). S. 5–12.

8. Mozhaeva G.V. Digital Humanities: cifrovoy povorot v gumanitarnykh naukah / G.V. Mozhaeva // Gumanitarnaya informatika: sb. statej / pod red. G.V. Mozhaevoy. 2015. Vyp. 9. S. 8 — 23.

9. Leonov V.P. Vavilonskaya biblioteka v kontekste dal'nego chteniya // Bibliografiya i knigovedenie. 2020. № 1. S. 77–83.

10. Sokolova E. V. Istolkovat' i/ili ischislit': o vozmozhnosti kombinirovannogo metoda na styke germenevtiki i digital humanities / E. V. Sokolova // Sovremennaya nauka o literature: sb. st. / RAN, INION. Moskva, 2018. S. 142–155.

11. Mayackij M. Digital Humanities — eto chto-to novoe ili my uzhe davno etim zanimaemsya? Interv'yu s Igorem Pil'shchikovym // Logos. 2015. T. 25, № 2 (104). S. 14–36.

12. Leonov V.P. Dal'nee chtenie kak strategiya tochnogo bibliografovedeniya // Nauchnye i tekhnicheskie biblioteki. 2019. № 10. S. 56–67.

13. Antopol'skij A.B. O klassifikacii informacionnykh resursov akademicheskikh organizacij sociogumanitarnogo profilya // Vzaimovliyanie informacionno-biblioteknoy sredy i obshchestvennykh nauk: sb. st. / RAN, INION. M., 2018. S. 48–61.

14. Polilova V.S. Cifrovoe budushchee istorii mirovoj literatury // Troickij variant — Nauka. 2019. № 25. S. 14.

15. Tutatina E.A. Zaglaviya knig: sovremennyye tendencii v knigoizdanii // Bibliografiya i knigovedenie. 2020. № 1. S. 111–119.

