

Старопечатные книги Пермской земли и технологии распознавания текстов

Ю. Р. Айдаров, Г. П. Волгирева

Пермский государственный университет, Пермь, Россия

На сегодняшний день Пермская земля счастливо разрешилась блестящим описанием всех старопечатных книг XVI-XVII веков [Кириллические 2003]. Издание выполнено в рамках научной программы «Московский государственный университет им. М. В. Ломоносова – Российской провинции» и региональной программы «Книжные памятники Прикамья». В работе приняли участие как московские специалисты археографы, так и местные. Руководителем проекта является доктор исторических наук И. В. Поздеева. Подобные региональные проекты научных описаний выполнены в Твери и Ярославле.

В Пермском каталоге описан 371 экземпляр 215 кириллических изданий, хранящихся в десяти хранилищах семи городов Пермской области. Большая часть (70 %) всего территориального собрания хранится в музеях: в Пермском – 92 экземпляра, в Березниковском – 72, Чердынском – 60, Соликамском – 11, Кунгурском – 4, Чайковском – 2, Ильинском – 1, в Пермской художественной галерее – 23. В Пермской областной универсальной научной библиотеке и в Пермском государственном педагогическом университете хранятся 99 экземпляров кириллических печатных книг XVI – XVII в. При этом Пермский региональный фонд оказался в два раза больше, чем предполагалось до начала описания. Более того, в пермских книгохранилищах старопечатных книг сохранилось значительно больше, чем в прославленной своими книжными богатствами Тверской земле.

С точки зрения технологии и методов распознавания текстов старопечатных книг изданный научный каталог содержит уникальную информацию о непосредственном процессе печатания книг и об участии тех или иных лиц в процессе их издания. В первую очередь это касается изданий Московского Печатного двора. Так, начиная с издания Триоди Постной 1621 г. в описании читаем подобную информацию: «Печаталась на двух станах «мастера попа Никона и никитинской печати»; наборщики Иван Плотников, Иван Москатыльников, «пособляли допечатывать» наборщики Афанасий

Петров и Иван Данилов. Вышла из дела 05.12.1621 г. (см.: РГАДА. Ф. 1182. Д. 1. Л. 346). Тираж в указе 1000 экз.; «вышли из дела» 1020 экз. Себестоимость 1 руб. 8 алт. 4 ден.; цена – без прибыли; цена простого переплета 10 алт., переплет с золотом на трех книгах стоил 1 руб. 7 алт. В 1632 г. на Печатном дворе одна книга в тетрадах стоила 1 руб. 25 алт., в простом переплете на 12 алт. дороже. (Там же. Д. 1. Л. 346; Д. 14. Л. 446 об.). Имеются сведения о рассылке по челобитным и продаже 973 экз.» [Кириллические 2003]. Нередко также указывается резчик букв, например: «Часовник с кафизмами и каноном ангелу хранителю» печатался на шести станах. Прописные новые слова резал Михаил Осипов. Тираж 2 300 экз.; цена без прибыли 10 алт.; стоимость простого переплета 3 алт. 2 ден. (РГАДА. Ф. 1182. Д. 19. Л. 117 об., 139 об.; Д. 23. Л. 95).

Временем расцвета деятельности Московского Печатного двора являлись 40-е годы XVII в. Книги печатали на нескольких станах в «Деревянных хоробах», «Каменных хоробах», в «Верхней палате», в «Нижней палате». Наиболее яркой фигурой, по документам этого времени, являлся резчик Иван Фофанов, а также выделялись знаменщики Филипп Антипьев, Иван Дмитриев «с товарищи». Эти данные помогают определить наиболее схожие издания по типам начертаний различных букв, символов и украшений. Так, близкими друг другу по системе символов должны быть следующие издания: «Иоанн Златоуст. Маргарит. 1641 г.» – в Перми два экземпляра, «Псалтырь с воследованием. 1642 г.» – в Перми два экземпляра, «Триодь Постная. 1642 г.» – в Перми один экземпляр, «Пролог, первая половина (сентябрь – февраль). 1641 г.» – в Перми один экземпляр, «Пролог, вторая половина (март – август). 1643 г.» – в Перми шесть экземпляров, сборник «Кириллова книга» 1644 г. – в Перми два экземпляра, «Минеи служебные, сентябрь (1644 г.), октябрь (1645 г.), ноябрь (1645 г.), май (1646 г.), июнь (1646 г.), август (1646 г.)» – в Перми 11 экземпляров. С 1647 года появляются новые имена. Так, в описании издания книги «Ефрем Сирий. Поучения. 1647 г.» читаем: «Вышла из дела 31.08.1647 г., 13 и 15.09.1647 г. платят «разборщику Гришке Аврамову, что он знаменил на груше, резцу Федьке Иванову [Попову], что он резал на груше преподобного Ефрема

Сирина». Тираж 1200 экз. ... Печаталась в Деревянных хоромах на четырех станах и в нижней палате «от Никольского монастыря» на четырех станах. Стоимость переплета «по обрезау золотом» экз. на «большой бумаге» 16 алт. 4 ден., простого переплета – 10 алт.» [Кириллические 2003].

Таким образом, благодаря подробному научному описанию старопечатных книг XVI-XVII вв. можно выявить близкие по типу издания, начать их оцифровывать, создавать базу данных электронных текстов и программы для дальнейшего историко-филологического изучения.

Кроме того, на основе выделенных групп изданий может быть создана так называемая обучающая выборка для распознающей системы на базе искусственных нейронных сетей [Аксенов 2004]. Программные продукты, использующие искусственные нейронные сети, показали хорошие результаты при решении задач, связанных с распознаванием образов, в связи с чем возникают предпосылки для перевода старопечатных книг XVI-XVII в электронный текстовый формат. Для того чтобы осуществить такой перевод, необходимо разработать специальную систему кодирования символов, с учетом всех видов дополнительных (например, надстрочных) знаков.

Работа над системой распознавания, основанной на искусственных нейронных сетях, требует выбора топологии нейронной сети. Наиболее широко при решении задач распознавания применяется так называемый персептрон [Минский и др. 1971]. Зависимость количества элементов промежуточного слоя персептрона от количества распознаваемых образов (в нашем случае – различных символов, встречающихся в старопечатных книгах) может быть определена экспериментально.

Опытным путем необходимо также выявить параметры зависимости влияния входных значений на результаты распознавания. Здесь предлагается использовать сигмоид – функцию следующего вида:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Сигмоид позволяет ввести нелинейность в работу нейронной сети и при этом позволяет сократить вычислительную сложность метода обратного распространения ошибки, поскольку его производная легко выражается через саму сигмоидальную функцию.

Таким образом, в результате выполнения описанных выше шагов, может быть построена система, позволяющая обеспечить эффективный поиск информации в электронном представлении старопечатных книг с использованием специальных методов набора символов, используемых в книгах XVI-XVII вв.

Список литературы

Кириллические 2003 – *Кириллические издания XVI-XVII вв. в хранилищах Пермской области.* – Пермь, 2003.

Аксенов 2004 – *Аксенов, О.* Методика формирования обучающих выборок для распознающей системы / О. Аксенов // *Нейроинформатика-2004* : Сб. научн. тр. VI. Всероссийская научно-техн. конф. – Ч. 2. – М. : МИФИ, 2004. – С. 215-222.

Минский и др. 1971 – *Минский, М.* Персептроны / М. Минский, С. Пейперт ; пер. с англ. – М., 1971.

Early printed books of Perm and OCR technology

Yuriy R. Aydarov, Galina P. Volgireva

Perm State University, Perm, Russia

Early printed books from the 16th and 17th centuries from Perm are described in sufficient scholarly detail to allow use of neural net technology to perform OCR on page images.