

Важные новости — теперь ДетКорпус доступен исследователям не только для онлайн-поиска, но и в качестве датасета

(<https://dataverse.pushdom.ru/dataset.xhtml?persistentId=doi:10.31860/openlit-2021.4-C001>), опубликованного в **Репозитории открытых данных по литературе и фольклору**. Формат и набор данных, включенных в датасет, подобран таким образом, чтобы на этом материале можно было воспроизвести статистические расчеты, сделанные на исходных текстах корпуса, а также проводить новые количественные исследования, опирающиеся на грамматическую и лексическую статистику. Подробное описание состава датасета (2273 прозаических произведения, опубликованных на русском языке в период с 1900-х по 2020-е годы) можно найти в файле README датасета и на сайте <http://detcorpus.ru/>.

Кроме технического обновления, есть и содержательное — новый выпуск Деткорпуса включает в себя более ста произведений 1920-х годов. Тексты предоставлены **Российской государственной детской библиотекой** и подготовлены к публикации: вычитаны и сверены с оригиналами. Благодаря сотрудничеству с РГДБ мы и дальше сможем пополнять Деткорпус произведениями начала XX века, расширяя ретроспективный охват нашей выборки. Среди ранних текстов стоит отдельно отметить подкорпус нон-фикшн 1920-х, в который вошли научно-популярные издания по естественным наукам, а также технике и технологиям. Интерфейс Деткорпуса позволяет познакомиться с фрагментами этих произведений, а иллюстрированные цифровые копии доступны для читателей в **НЭДБ**.

Лаборатория цифровых исследований литературы и фольклора и Центр исследований детской литературы тепло благодарят РГДБ и лично Илью Гавришина, благодаря которому это обновление стало возможным. За помощь с подготовкой текстов к публикации мы признательны главному библиографу ЛОДБ Любове Алейник, а также студенткам филологической программы НИУ ВШЭ СПб Злате Климас, Екатерине Стариковой и Анне Логиновой.