

УДК 004.82

doi: 10.15622/rcai.2025.067

СИСТЕМА ПОДДЕРЖКИ СОЗДАНИЯ КОРПУСА РУКОПИСНОГО НАСЛЕДИЯ ДРЕВНЕЙ РУСИ

Д.В. Демидов (*dvdemidov@mephi.ru*)

Национальный исследовательский ядерный университет «МИФИ»,
Москва

В работе описывается система поддержки разработки корпуса рукописного наследия Древней Руси и технология распознавания рукописей. Представляется концепция сращивания в единую среду средств оптического распознавания символов и машинного обучения с лингвистическим обеспечением и средствами поддержки работы инженеров и филологов. Целью совместной работы коллектива является получение цифрового корпуса рукописей, доступного для проведения палеографических и текстологических исследований.

Ключевые слова: оптическое распознавание; корпусная лингвистика; древние рукописи; эволюционное моделирование.

Введение

Идея создания корпуса рукописного наследия Древней Руси (далее Корпус) возникла в Институте Русского Языка им. В.В. Виноградова РАН, а её реализация началась в 2021 г. в стенах НИЯУ МИФИ совместными усилиями коллектива инженеров и филологов [Демидов и др., 2024а]. Источниками для Корпуса являются рукописные Служебные Миней XI–XVII вв., доступные в виде набора отсканированных разворотов книг. По памятникам этого типа можно проследить изменения в самом языке и многое почерпнуть об истории средневековой Руси. Однако изучение этого массива информации на бумажном носителе или в виде нераспознанных изображений, полученных в результате сканирования, чрезвычайно трудоёмко – поиск, различные виды анализа выполняются вручную.

Целью исследовательской группы является создание инструментария для работы со славянскими рукописями, а также их электронного издания. Это требует разработки идеологии и интерфейса «автоматизированного рабочего места филолога», а также применения системы распознавания. К сожалению, современные коммерческие системы и веб-сервисы оптиче-

ского распознавания не рассчитаны на работу с таким материалом, а специализированные платформы распознавания на основе моделей машинного обучения типа transkribus.org оказываются слабо применимы на практике.

Это связано с такими особенностями рассматриваемых источников как:

- древность самого языка, вышедшего из светского употребления;
- обилие графической информации помимо собственно символов (рис. 1): буквицы, надстрочные и диакритические знаки, вязь, крюковые ноты (межстрочные пометы-указания для певцов);
- сокращения слов, на что указывают титла и буквититла;
- высокая дефектность при ветхом носителе: затемнение бумаги/пергамена, размытие чернил, потёртости, пятна, капли воска, утрата фрагментов, печати, карандашные пометы;
- геометрические искажения (неровные строки, вариация почерков), искажения, возникшие при сканировании (повороты, чёрные рамки).

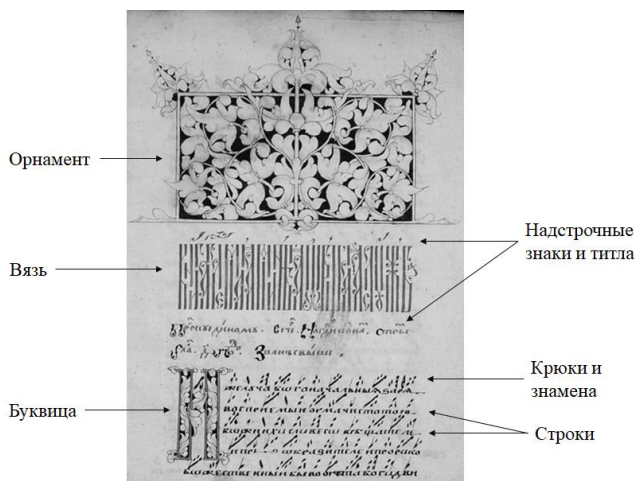


Рис. 1. Пример страницы Служебной Миней

Все эти обстоятельства требуют подготовки системы распознавания славянских рукописных Миней с необходимым лингвистическим и палеографическим обеспечением, выделением отдельных графических элементов букв для распознавания почерков и машинной датировки, каталогизацией источников для текстологических исследований.

Используемый рукописный материал характеризуется широким временным охватом (порядка 700 лет). За этот период существенно менялись все составляющие грамматики языка – азбука, лексика (орфография), морфология, синтаксис. Практически с нуля развилась пунктуация – по-

началу письмо было сплошным, без пробелов. В нотированных рукописях менялась и крюковая нотация (три поколения). В плане оптического распознавания современный русский язык представляется удобным статическим объектом, но древний славянский язык требует учёта фактора времени при обработке рукописей разных эпох: грамматика, крюковая нотация, графические традиции оформления становятся функциями от времени. С одной стороны, это ведёт к усложнению программных средств, а с другой, позволяет учитывать особенности разных периодов и открывает возможности для решения задач датировки рукописей. Встаёт вопрос об адаптивности как программного, так и лингвистического обеспечения для работы с подобным материалом.

Важно отметить, что у филологической группы объект исследования хоть и весьма широк, но определён изначально. К нему относятся:

- фрагменты книг для совершения богослужений: служебник, минея, триодь, паремийник, служебное евангелие, следованная псалтырь и др.;
- лексика, пунктуация, синтаксис церковнославянского языка;
- почерки;
- крюковая нотация.

У инженерной группы ситуация иная. В начале проекта предметом исследования были алгоритмы предобработки изображений текстов: устранение искажений, фильтрация, бинаризация и др. [Васелюк и др., 2024], алгоритмы сегментации текстовых зон, строк, букв и буквиц, надстрочных символов [Воронин и др., 2022], [Демидов и др., 2024b]. Известные подходы, изложенные, например, в [Гонсалес и др., 2012], [Визильтер и др., 2010] исследовались на применимость, шёл отбор наиболее перспективных методов для решения частных задач, таких как обнаружение линий разреза разворотов, текстовая сегментация орнаментов. Исследовалась задача поиска фрагментов текстов: полнотекстовый и нечёткий поиск по выражению [Маннинг и др., 2011], осложнённый наличием надстрочных знаков и титлов. Позднее предметом исследования стали модели классификации символов, модели синтеза размеченных данных, методы аугментации и доразметки. В фокусе этой работы – технология распознавания рукописей, эволюционирующая по мере создания Корпуса рукописного наследия Древней Руси.

1. Модель системы поддержки создания корпуса

Рассмотрим систему поддержки построения корпуса с точки зрения системного подхода. Отметим, что в целевой состав Корпуса входят:

1. Каталог рукописей с текстами.
2. Каталог графических объектов, представляющих отдельный интерес: орнаменты, вязь, отдельные буквы и лигатуры, буквицы.

3. Графические образы разворотов, страниц, строк, символов, букв, надстрочных знаков, крюков, знамен и др., сопоставленные с текстовой и признаковой информацией, используемой для поиска.
4. Морфологический словарь.

Следуя [Волкова и др., 2006], модель системы можно представить кортежем $S = \langle A, R, B, Z \rangle$, где A – множество изображений, включая как сканированные изображения рукописей, так и все производные изображения и фрагменты, продуцируемые средствами распознавания; B – множество текстов рукописей, в т.ч. нотируемых, продуцируемых средствами распознавания или привносимых извне; R – множество отношений между фрагментами изображений и фрагментами текстов, устанавливаемых в ходе работы над Корпусом; Z – множество целей системы, например, стремление обработать все входные изображения, максимизация метрик качества распознавания, установление соответствия между фрагментами изображений и фрагментами текстов и др. В более развитой форме модель системы можно представить кортежем $S = \langle Z, STR, TECH, COND \rangle$, где помимо целей Z имеется множество структур STR , реализующих цели, технология $TECH$ – методы, алгоритмы, модели, конвейеры обработки рукописей и страниц, и $COND$ – условия функционирования системы (внутренние и внешние), например, реализация проекта в академической среде учебного вуза.

1.1. Эволюционный подход к развитию системы поддержки создания корпуса

Первый прототип средств поддержки создания корпуса можно охарактеризовать как наивный: реализовывались классические подходы к распознаванию и апробировались вручную на небольшом количестве образцов. После создания средств каталогизации и первичной обработки страниц стал ясен масштаб Корпуса – сейчас коллектив активно работает с 245 рукописями, содержащими порядка 110 тысяч страниц.

Первичной целью было создание скелетного прототипа системы распознавания и параллельное улучшение отдельных алгоритмов. В табл. 1 приведены некоторые шаги развития алгоритмов и средств поддержки построения корпуса. Можно отметить общую тенденцию к усложнению и ветвлению алгоритмов в ходе активного применения метода проб и ошибок, накопление алгоритмов-кандидатов для повторного использования, иерархизацию конвейера обработки изображений: добавлялись новые этапы, некоторые этапы делились на подэтапы, результаты одного этапа определяли набор последующих этапов. Новые версии компонентов распознавания проверялись на всё большем объёме данных. Если видимые результаты внушали оптимизм, то изменения внедрялись. Однако, автоматическое применение алгоритмов ко всему объёму корпуса никак не облегчало оценку результатов, которая по-прежнему проводилась вруч-

ную и выборочно. К сожалению, новые версии программных средств не всегда давали однозначный результат: на каком-то материале удавалось получить прирост качества, а на ином результаты становились сильно хуже, причём далеко не всегда это было сразу заметно.

Таблица 1

Фаза	Предобработка	Сегментация	Классификация
1	Медианная фильтрация полутонового изображения; Глобальная бинаризация	Разбиение изображения на строки и символы методом профилей	Простая признаковая модель символа (моменты, профили)
2	Контрастирование; Адаптивная бинаризация; Нейросетевая бинаризация	Сегментация букв, вязи; Сегментация строк; Сегментация и аннотация символов внешним средством	Ручная разметка символов; Классификация букв, сегментированных вручную
3	Морфологические преобразования; Разделение методов предобработки в зависимости от цели и потребностей следующего шага конвейера.	Текстурная сегментация орнаментов, вязи; Сегментация крюков и знамен; Полуавтоматическая разрезка разворотов (с валидацией и коррекцией)	Расширенная признаковая модель символов (более 100 признаков); Нейросетевой классификатор крюков и знамен (разметка ручная)
4	Обнаружение геометрических искажений страницы (поворот, трапеция) и редактор для их устранения	Встроенные средства сегментации и аннотирования	Синтез данных для обучения на основе имитации шрифтов и реальных фонов страниц; Нейросетевой классификатор строк

Изначально спроектировать систему «от и до» с учётом всех этих аспектов, не представлялось возможным в силу фундаментальных причин:

1. Зависимость языка от периода времени и необходимость учёта фактора времени в средствах обработки.
2. Неполнота представлений команды разработки об объекте исследования. Ясно, что по мере углубления и уточнения этих представлений, система также должна эволюционировать.
3. Многообразие и неформализуемость языка в силу его естественно-го происхождения (несмотря на осознанное проектирование славянского алфавита под сложившуюся фонетику). Это приводит к сильному сцепле-

нию и взаимозависимости отдельных стадий анализа изображения. Как морфологический анализ нельзя завершить до завершения синтаксического из-за лексической омонимии, так и сегментацию нельзя завершить до завершения классификации из-за омонимии знаков и их фрагментов.

4. Имеющийся объём данных изначально требовал применения статистического подхода наряду со структурным и др., но без размеченных данных это было сильно затруднено. Вручную удалось разметить порядка 15000 образцов символов.

1.2. Критерии оценки успешности функционирования системы

В силу отмеченных фундаментальных причин нет возможности сразу создать идеальную систему, как нет возможности сформулировать единственный критерий успешности её функционирования. Это сложный процесс человеко-машинного взаимодействия, направленный на достижение совокупности неявных характеристик, обуславливающих практическую применимость материалов Корпуса.

Целесообразно изначально выстраивать критерии оценки качества решения каждой из подзадач, изучая взаимное влияние стадий обработки и добиваясь глобальной оптимизации. При этом движение в сторону оптимума возможно как путём изменения настроек компонентов, фиксируя их состав, так и за счёт усложнения внутренней структуры системы. Более того, сами критерии также должны эволюционировать по мере развития алгоритмов и появления новых метрик, вычисление которых ранее не было возможным. В задаче создания Корпуса наблюдались, по крайней мере, следующие фазы эволюции критериев оценки:

1. Качественная оценка оператором визуализированных результатов распознавания страниц.
2. Количественные статистические показатели с разным охватом (по-странично, по всем рукописям) для выявления грубых ошибок и оценки достигнутого прогресса.
3. Количественные показатели с потенциалом использования в качестве обратной связи для подстройки параметров алгоритмов и гиперпараметров моделей и как метрики для машинного обучения.

Критерии оценки можно разделить по подзадачам и автоматизируемым процессам: сегментация, классификация, поиск, обучение, валидация. В этот список можно добавить такие метапроцессы как валидация, оптимизация и перепроектирование (внесение инноваций), которые в ходе эволюции должны будут стать встроенными в систему и направлять дальнейшую эволюцию. Вместо количественных метрик качества можно рассматривать качественные в сочетании с такими подходами как нечёткие алгоритмы и нечёткие модели [Борисов и др., 2007]. Переход от системы, решающей конкретную задачу, к инструментальной системе и затем к интеллектуальной инструментальной системе, демонстрирующей целена-

правленную деятельность, давно известен на примере развития экспертных систем до оболочек и инструментальных комплексов ([Рыбина, 2014]). Конкретно для задачи распознавания изображений страниц древних рукописей уместны следующие критерии оценки, сгруппированные по этапам обработки:

1. Сегментация. Статистическая схожесть результатов сегментации реальных страниц рукописей с результатами сегментации синтезированных страниц или размеченных реальных страниц. Например, тяготение сегментов символов к срединным линиям, небольшой разброс высот символов, сходство текстурных признаков зон текста и пр.

2. Обучение и классификация. Традиционные точность, полнота, F-мера, близость матрицы неточностей к диагональной, непереобученность (устойчивость показателей распознавания на выборках различного масштаба).

3. Валидация. Снижение с течением времени количества ручных корректив в ходе валидации полученных результатов. Уместно для алгоритмов всех этапов от разрезки разворотов и устранения искажений до правок аннотаций или распознанного текста.

Чтобы в ходе развития системы трудоёмкость ручной валидации снижалась, а обратные связи автоматизировались в большей степени, необходимо ввести такие критерии качества, как простота модификации конвейера обработки и системы в целом, а также толерантность к вносимым изменениям (возможность отката в случае неудовлетворительного результата итерации).

2. Технология распознавания рукописей

Помимо средств автоматизации операций создание Корпуса требует системы хранения информации. Здесь уместны и транзакционная СУБД, и объектное хранилище, и документно-ориентированные постреляционные представления. В силу эволюционного характера деятельности по созданию Корпуса в основу поддерживающей эту деятельность технологии целесообразно положить принципы BASE (базовая доступность, мягкое состояние, потенциальная согласованность) [Pritchett, 2008] в противовес принципам ACID транзакционных баз данных (атомарность, согласованность данных, изоляция транзакций, устойчивость) [Gray, 1981]. Здесь под согласованностью данных подразумевается не свойство данных, хранящихся в БД, а согласованность выделенных фрагментов изображений с текстовыми фрагментами. Достижение этой согласованности является одной из целей системы. Для реализации принципа мягкого состояния усложнялись и жизненные циклы рукописей и страниц.

2.1. Жизненный цикл ключевых сущностей

Жизненный цикл рукописи:

1. Рукопись создана – в каталоге создана запись с библиографической информацией о рукописи.
2. Загружены изображения – сканированные изображения разворотов по отдельности или цельным документом.
3. Предложены линии разрезов разворотов на страницы.
4. Линии разреза валидированы – пользователь принял предложения или внёс коррективы линий (подвинул, удалил, добавил).
5. Разрезка осуществляется – автоматический процесс на основе согласованных линий разреза разворотов.
6. Рукопись разрезана на страницы, страницы доступны для просмотра и дальнейшей работы.
7. Запущена постраничная обработка рукописи – массовая операция по предобработке, сегментации и классификации символов.
8. Рукопись обработана, производные результаты доступны для оценки и корректировки.
9. Рукопись опубликована, результаты распознавания доступны для пользователей Корпуса.

Жизненный цикл страницы:

1. Новая страница (загружена или получена в результате разреза).
2. Страница в очереди на обработку.
3. Бинаризация – формируется монохромное изображение страницы для удобства последующих стадий обработки.
4. Сегментация – формируется иерархическое представление геометрической структуры страницы.
5. Классификация – типизация сегментов и классификация символов.
6. Визуализация результатов – сборка текста и генерация выходного изображения для качественной оценки результатов распознавания.
7. Страница обработана, конвейер обработки успешно пройден.
8. Валидация – если в процессе обработки были выявлены ошибки распознавания, пользователь может вручную исправить эти ошибки и перезапустить процесс обработки с учётом исправлений.
9. Страница валидирована, результаты работы могут пополнить обучающие выборки соответствующих подзадач и задействованы в дообучении/повторном обучении моделей.

В случае возникновения ошибок во время обработки, страница возвращается на предыдущий этап для повторной обработки.

2.2. Структура программных средств

Структурно система поддержки создания корпуса включает подсистему управления корпусом, подсистему распознавания, подсистему обучения и инженерную библиотеку (рис. 2). Для работы с корпусом в подсисте-

теме управления предусмотрен веб-портал, доступный в интернет по адресу slavcorp.org.ru. Основная функциональность: каталогизация рукописей, поиск рукописей по каталогу, полнотекстовый поиск фрагмента текста, редактирование морфологического словаря с поисковыми возможностями, валидация разрезов разворотов страниц, валидация сегментации текстовых зон на строки, валидация распознанного текста. Ролевая модель включает три категории: читатель, редактор (филолог или инженер), администратор.

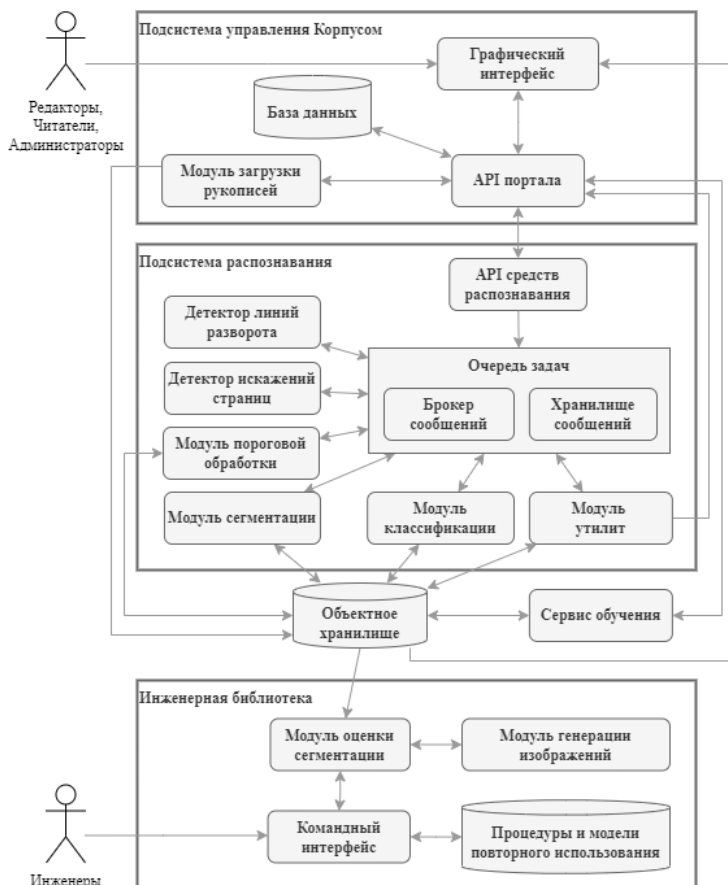


Рис. 2. Архитектура программных средств

Подсистема распознавания включает такие модули как детектор линий разворотов, детектор искажений страниц, модули пороговой обработки, сегментации, классификации и вспомогательных утилит. Доступ к моду-

лям предоставляется через программный интерфейс с применением очереди задач для асинхронной обработки массовых операций. Результаты обработки размещаются в объектном хранилище.

Подсистема обучения отвечает за обучение моделей классификации.

Инженерная библиотека поддерживает поисковую и адаптационную деятельность инженера. На основе материала, производимого подсистемами распознавания и обучения, экспериментально апробируются гипотезы для последующих параметрических, алгоритмических или структурных изменений системы с целью её совершенствования. В библиотеку входят повторно используемые процедуры и модели обработки, оценки, визуализации и др. Доступ к компонентам библиотеки осуществляется разными способами (записные книжки, командный и программный интерфейсы). Это открывает возможности структурного синтеза подпрограмм и сборочного программирования ([Тыгугу, 1984], [Лаврищева, 2025]) и обеспечивает модифицируемость конвейера обработки рукописей.

Входными данными системы являются изображения текстов и в тех случаях, когда для рукописи удаётся найти уже готовый результат, сам текст. Также вводится метаинформация о рукописях, данные для морфологического словаря. Корректирующие и валидирующие действия можно отнести к управляющим воздействиям. Выходными данными являются тексты рукописей, распознанные и валидированные с помощью системы, поисковые выдачи, словарь в машиночитаемом формате.

Ручные и частично автоматизируемые операции: валидация результатов работы различных алгоритмов (зачатки механизма отбора для эволюционного процесса); коррекция промежуточных данных, служащих входными данными для следующих этапов; подстройка параметров алгоритмов/обучение моделей; принятие решений о внесении новых алгоритмов/модификаций существующих алгоритмов.

Пока точкой принятия решений об изменениях системы является человек (инженер или филолог), однако путём добавления оценочных модулей и формализации целевой функции создаются предпосылки делегирования части решений самой системе. Как только циклы обратной связи замыкаются, то более поздние стадии обработки получают возможность влиять на более ранние стадии в следующей итерации. Как отмечается в [Емельянов и др., 2003], обратные связи препятствуют такому действию системы, которое выводит её за установленные пределы. Версионизируемость повторно используемых компонентов гарантирует управляемость процесса эволюции системы. Таким образом наряду с машинным обучением и постепенным усложнением системы становятся применимы подходы эволюционной кибернетики.

3. Перспективы развития

Развитие системы поддержки создания корпуса рукописного наследия Древней Руси можно рассматривать в трёх направлениях.

Корпусная лингвистика. При наполнении Корпуса открываются перспективы построения лингвистических моделей древнего языка и синтаксически-размеченного корпуса с привязкой к периоду времени. Так как на раннем этапе развития русской письменности слова не разделялись графически, то задачи морфологического и синтаксического анализа требуют отдельной нетривиальной процедуры членения, в отличие от современного языка. В терминах трёхмерной стратификационной модели языка [Шаляпина, 2007], где ось членения на языковые единицы рассматривается многоярусной, переход с морфологического на синтаксический ярус мыслится итеративным, а не скачкообразным, при этом лексический ярус является результатом нахождения наиболее непротиворечивого варианта членения по отношению к морфологическому и синтаксическому ярусам. Эта задача не нова для глифowych языков типа китайского, японского или корейского, где нет графического разделения слов, а членение выполняется с опорой на морфологию, и аналогична задаче преобразования непрерывного речевого потока в текст, однако для раннего церковнославянского языка ставится, по всей видимости, впервые.

Обобщение на другие языки. Результатом обобщения разрабатываемой технологии может стать разделение системы поддержки создания корпуса на систему управления корпусом в виде относительно универсальной оболочки и надсистему со средствами визуального проектирования конвейеров обработки изображений текстов древних языков на основе эволюционного моделирования. Прежде всего рассматриваются славянские языки, языки бывшего СССР, например, древнеармянский и среднеармянский (V–XVII), где особый интерес представляет коллекция рукописей в Матенадаране (г. Ереван). В Отделе языков народов Азии и Африки Института Востоковедения РАН (до 2020 г. группа Шаляпиной З.М.) изучают доклассический и постклассический древнекитайский (V до н.э. – V в. н.э.), старояпонский и средневековый японский язык (VIII–XVII) и также испытывают значительные трудности из-за слабой применимости средств оптического распознавания текстов.

Расширение до экосистемы. Предоставление доступа к отдельным программным средствам в виде веб-сервисов для исследователей; создание мобильного приложения с поиском в Корпусе по фотографии фрагмента текста, «узнаванием» рукописи по фото. В образовательных и исследовательских целях интересна разработка средств автоматического перевода с церковнославянского на современный русский язык и наоборот.

Работа выполнена при поддержке программы Приоритет-2030 НИЯУ МИФИ.

Список литературы

- [Борисов и др., 2007] Борисов В.В., Круглов В.В., Федулов А.С. Нечёткие модели и сети. – М.: Горячая линия–Телеком, 2007. – 284 с.: ил.
- [Васелюк и др., 2024] Васелюк А.С., Демидов Д.В., Потемкина А.А. Фильтрация и бинаризация изображений текстов славянских рукописных книг // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во С.-Петербур. ун-та, 2024. – С. 43-50.
- [Визильтер и др., 2010] Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Ососков М.В., Моржин А.В. Обработка и анализ изображений в задачах машинного зрения. – М.: Физматкнига, 2010. – 672 с.
- [Волкова и др., 2006] Волкова В.Н., Денисов А.А. Теория систем: учеб. пособие. – М.: Высшая школа, 2006. – 511 с.: ил.
- [Воронин и др., 2022] Воронин К.А., Демидов Д.В., Плетнёва А.А. Автоматическая сегментация строк в текстах на церковнославянском языке // Гибридные и синергетические интеллектуальные системы: [Электронный ресурс]: научное электронное издание. – Калининград: Издательство БФУ им. И. Канта, 2022. – С. 286-292.
- [Гонсалес и др., 2012] Гонсалес Р., Вудс Р. Цифровая обработка изображений. – 3-е изд. – М.: Техносфера, 2012. – 1104 с.
- [Демидов и др., 2024а] Демидов Д.В., Кравецкий А.Г., Ларионов А.А., Плетнёва А.А. Создание Корпуса рукописного наследия Древней Руси // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во СПбГУ, 2024. – С. 72-81.
- [Демидов и др., 2024б] Демидов Д.В., Миронова В.В. Предобработка изображений букв в славянских рукописных книгах // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во СПбГУ, 2024. – С. 82-88.
- [Емельянов и др., 2003] Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. – М.: Физматлит, 2003. – 432 с.
- [Лаврищева, 2025] Лаврищева Е.М. Программная инженерия и технологии программирования сложных систем. – М.: Изд-во Юрайт, 2025. – 432 с.
- [Маннинг и др., 2011] Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск: пер. с англ. – М.: ООО «И.Д. Вильямс», 2011. – 528 с.
- [Рыбина, 2014] Рыбина Г.В. Основы построения интеллектуальных систем. – М.: Финансы и статистика, 2014. – 432 с.
- [Тыгу, 1984] Тыгу Э.Х. Концептуальное программирование. – М.: Наука. Главная редакция физико-математической литературы, 1984. – 256 с.
- [Шаляпина, 2007] Шаляпина З.М. Трёхмерная стратификационная модель языка и его функционирования: к общей теории лингвистических моделей. – М.: Вост. лит., 2007. – 480 с.
- [Gray, 1981] Gray J. The Transaction Concept: Virtues and Limitations // Very Large Data Bases, 7th International Conference, September 9–11, 1981, Cannes, France, Proceedings. IEEE Computer Society, 1981. – P. 144-154.
- [Pritchett, 2008] Pritchett, D. BASE: An Acid Alternative: In partitioned databases, trading some consistency for availability can lead to dramatic improvements in scalability // Queue. – 2008. – Vol. 6, Iss. 3. – P. 48-55.