

# АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

---

УДК 81'322.2'37:316.7

Л.А. Мосунова, Е.В. Митягина, П.В. Ананин

## Автоматическое извлечение ценностно-смысловых компонентов культурного кода из корпуса читательских отзывов

*Исследуется проблема автоматического анализа больших корпусов гуманитарных текстов. Отзыв о книге рассматривается как пространство культурного кода, содержащее сведения о ценностных доминантах в сознании современного читателя. Создана база данных «Отзывы о художественных произведениях, содержащие сведения о культурном коде», отличающаяся от аналогов процедурой формирования и тематикой. Описан процесс автоматического поиска и представлены результаты эксперимента по извлечению ценностно-смысловых компонентов культурного кода из корпуса читательских отзывов (8278 текстов). Сделан вывод об эффективности разработанной методики, основанной на статистических методах и программах текстового анализа. Автоматический анализ текстов отзывов подтвердил идеи учёных об изменении культурного кода в условиях глобализации: при его определённой стабильности происходит замена некоторых традиционных ценностей на ценности иных культур.*

**Ключевые слова:** культурный код, компоненты культурного кода, читательский отзыв, ценностно-смысловая сфера, идентичность, автоматическое извлечение данных, текстовый анализ данных, библиотека электронных книг ЛитРес

**DOI:** 10.36535/0548-0027-2022-05-3

### ВВЕДЕНИЕ

В мировом научном сообществе с начала XXI в. становится всё более популярной концепция глобального общества (*global society*), с точки зрения которой все люди Земли являются гражданами единого планетарного общества. Так, по О.С. Сабдену, «с изменением мировой цивилизации переход на новую, седьмую формацию может осуществиться на основе построения интегрального общества, конвергентного пути развития мира» [1, с. 116]. Рост международного общения и повсеместное распространение Интернета ускоряют процессы глобализации, которые ведут к культурной интеграции и унификации. Исследования показывают, что одна из важнейших проблем современного человечества – опасность растворения национальных культур в глобальной культуре [2, с. 291]. Популярны международные культурные явления нередко вытесняют национальные или превращают их в интернациональные [3, с. 33]. Учёные расценивают это как утрату национальных культурных ценностей, утрату идентичности и ратуют за возрождение национальной культуры [4, с. 25–28].

Не случайно широко обсуждаемой в науке становится тема культурного кода, который обуславливает потребность человека в самоидентификации [5–7]. Будучи явлением коллективным («коллективное программирование» сознания, которое складывается из поколения в поколение), культурный код отличает членов одной группы или категории от других, делает нас особенными. Культурный код одновременно и стабилен, и способен «обрастать» новыми культурными доминантами путём приумножения этих доминант через их репродуцирование в новой исторической ситуации и через новые интерпретации старого культурного массива под влиянием уникального личного опыта [8, 9]. Иными словами, культурный код соединяет в себе социальное и общее (результат общественной практики) с индивидуальным и уникальным (результат субъективного опыта). Это соединение личного опыта индивида с коллективным опытом социальной общности составляет ядро ценностно-смысловой сферы отдельного человека [10].

Читательский отзыв, являясь сугубо личным высказыванием о социокультурном явлении, о книге, –

это своего рода отражение актуального состояния культурного кода. Изучение мнений читателей о книгах современных авторов и проблемах, поднимаемых в них, поможет расширить наши представления о культурном коде и увидеть его изменение в сегодняшних условиях. Мы предположили, что если извлечь из эмоциональных, спонтанных, нередко сумбурных высказываний читателей компоненты культурного кода, составляющие его основу, а именно ценностно-смысловые доминанты, то можно определить, какие культурные ценности остаются стабильными, а какие изменились под влиянием глобализации и модернизации.

Возникает исследовательская задача: каковы пути и средства автоматического получения фактов для интерпретации, поскольку убедительные обобщения можно сделать лишь при большой выборке. Возможности для решения этой задачи открывают цифровые технологии, помогающие обрабатывать и обобщать большие данные. Массивы читательских отзывов существуют и постоянно пополняются в открытом доступе в сети Интернет, они позволяют анализировать текущее состояние ценностно-смысловой сферы и идентификации читающей публики.

Однако в настоящее время отсутствуют методики, которые могут осуществлять автоматический анализ текстов в рамках темы извлечения ценностно-смысловых компонентов культурного кода из корпуса читательских отзывов и применять статистические методы для анализа данных о компонентах культурного кода. Необходимость создания методов автоматического сбора и обработки соответствующих данных на основе текстового анализа очевидна. В настоящей статье предлагается методика извлечения ценностно-смысловых компонентов культурного кода из корпуса читательских отзывов, включающая составление поисковых запросов, описание процесса подготовки данных и критериев для анализа, а также оценку полученных данных.

## ИСТОЧНИКИ ДАННЫХ И ВЫДЕЛЕНИЕ РЕЛЕВАНТНЫХ ТЕКСТОВ

Проведённое нами исследование в качестве объекта имело корпус читательских отзывов – 8278 текстов, размещённых на платформе библиотеки электронных книг ЛитРес<sup>1</sup>. Предметом анализа стали отклики читателей на книги лауреатов российских литературных премий с 2000 по 2020 гг.: «Большая книга», «Русский Букер», «Национальный бестселлер», «Ясная поляна», «Большая литературная премия России», «Писатель года», «Патриаршая литературная премия», «Государственная премия в области литературы и искусства».

Релевантность сообщений – первичная характеристика, отражающая степень соответствия контента теме исследования, – определялась по двум критериям: качественному и количественному. Пилотное исследование в виде ручной разметки и анализа трёхсот

отзывов выявило статистически значимые закономерности. Так, было обнаружено, что не релевантными являются отзывы, содержащие менее 4-х предложений, менее 40 слов или менее 200 знаков. В силу лаконичности в них отсутствовало содержание, позволяющее проводить дальнейший анализ.

Качественный критерий заключался в наличии фактов, выражении эмоций или смыслов, указывающих хотя бы на один компонент ценностно-мотивационной сферы его автора (его ценности, интересы или мотивы). Выявить соответствие данному критерию можно, создав кодификатор ключевых слов ценностно-смысловой сферы, а также определив так называемые «стоп-слова», которые напрямую сигнализируют о нерелевантности отзыва.

По Д.А. Леонтьеву, «личностные ценности являются генетически производными от ценностей социальных групп и общностей разного масштаба», они «выступают как внутренние носители социальной регуляции, укоренённые в структуре личности» [10, с. 231]. Ценности самой масштабной общности – народа, нации – сконцентрированы в государственных документах, в том числе в Указе Президента РФ от 2 июля 2021 г. № 400 «О Стратегии национальной безопасности Российской Федерации», где к традиционным российским духовно-нравственным ценностям относятся «жизнь, достоинство, права и свободы человека, патриотизм, гражданственность, служение Отечеству и ответственность за его судьбу, высокие нравственные идеалы, крепкая семья, созидательный труд, приоритет духовного над материальным, гуманизм, милосердие, справедливость, коллективизм, взаимопомощь и взаимоуважение, историческая память и преемственность поколений, единство народов России» (IV, ст. 91)<sup>2</sup>.

Использование этих понятий в качестве ключевых позволило выявить их в читательских отзывах. Для расширения зоны смыслового поиска к каждому понятию был подобран синонимический ряд, содержащий от 2-х до 22-х близких по смыслу слов. Приведём примеры таких синонимических рядов.

**Милосердие** – *добросердечие, доброта, сочувствие, благожелательность, благодать, благоволение, доброжелательность, благосклонность, любезность, великодушие, жалость, сострадание, ласка, любовь, снисхождение, добро, сердоболие, умение прощать.*

**Справедливость** – *праведность, честность, беспристрастность, беспристрастие, заслуженность, нелицеприятность, непредубеждённость, непредвзятость, объективность, правда, правильность, верность, законность, достоверность, истинность, правосудие.*

**Единство (народов России)** – *целостность, цельность, монолитность, целость, общность, соборность, идентичность, слитность, тождественность, сплочённость, солидарность, единомыслие, единогласие, спаянность, гармония, гармоничность, мир, равенство, тождество, согласие, сходство.*

<sup>1</sup> Крупнейший представитель рынка электронных и аудиокниг в России и странах СНГ, занимает две трети рынка легальных электронных и аудиокниг России, сегодня имеет 166 тыс. читателей.

<sup>2</sup> [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_389271/1e2b5c5fc29c839c457c3d876e9cc7b475bc7d45/](http://www.consultant.ru/document/cons_doc_LAW_389271/1e2b5c5fc29c839c457c3d876e9cc7b475bc7d45/)

## Методика исследования

Процесс автоматического поиска включал четыре этапа.

На первом этапе, «подготовительном», осуществлялся сбор и подготовка данных для последующей обработки. Для составления корпуса читательских отзывов использовался интерфейс прикладного программирования API каталога интернет-магазина «ЛитРес»<sup>3</sup>. В ходе специальных регистрационных процедур был получен *Access token* разработчика и *ID* приложения. С помощью соответствующего запроса была выгружена информация по отзывам читателей о книгах лауреатов российских литературных премий за 2000–2020 гг.

Эта информация содержала общие сведения: идентификатор отзыва, дату написания отзыва, имя (или псевдоним) автора отзыва, идентификатор пользователя, являющегося автором отзыва, текст отзыва, число «положительных» и «отрицательных» оценок. Вся собранная информация была преобразована в базу данных в табличном виде и сохранена в формате \*.xlsx. В базу данных вошла 7961 запись об отзывах читателей – из первоначально выгруженных 8278 текстов были исключены 317 дублирующие, а также технически повреждённые высказывания, не позволяющие выполнять последующий анализ (рис. 1).

На втором этапе, «текстуальном», были отдельно подготовлены тексты отзывов для дальнейшей настройки автоматического поиска. На этом и последующих этапах обработка корпуса отзывов осуществлялась с помощью различных функциональных узлов программы PolyAnalyst™, предоставленной ООО «Мегапьютер Интеллидженс». Специальная подготовка текстовых данных включала в себя два основных момента.

1. Индексация текстов отзывов. К содержанию текста отзыва применялся узел «Индекс». Выходом узла стали исходные данные, к которым добавился индекс текста, т. е. информация о его количественном и лексическом составе. В каждом отзыве были определены:

- а) количество слов;
- б) часть речи, к которой относится каждое слово отзыва;
- в) связи между словами определённой предметной области.

Для установления связей между словами использовались логические понятия: холонимы – тип отношения «целое – часть» (в отзывах, например, это *жизнь – прошлое*); меронимы – тип отношения «часть – целое» (*служение отечеству – отечество*); гиперонимы – тип отношения «общее – частное» (*право – право на жизнь, права человека*); гипонимы – тип отношения «частное – общее» (*память поколения, историческая память – память*). Это действие позволило в дальнейшем осуществлять поиск не только по словам, но и по синтаксической связи между словами, а также учитывать при поиске морфологическое изменение слов с учётом рода, времени, падежа.

2. Проверка орфографии. Важным шагом в обработке текстовой информации корпуса читатель-

ских отзывов стала проверка орфографии, а также автоматическое исправление типичных ошибок. С помощью узла «проверка орфографии» в исследуемом массиве отзывов удалось обнаружить 1257 ошибок и установить порог достоверности в 50%, что позволило автоматически исправить 553 слова (43,99 %). Таким образом были исправлены наиболее явные ошибки (*аффтар – автор; безисходность – безысходность*) и при этом отвергнуты сомнительные, по мнению программы, варианты исправлений (*пропиарены – пропарены*), что позволило увеличить перечень слов для корректного поиска.

На третьем этапе, «понятийном», выполнена настройка автоматического поиска отзывов, содержащих упоминание о традиционных российских духовно-нравственных ценностях. Для этого к каждому понятию был подобран словарь синонимов (рис. 2). Всего составлено 20 различных словарей.

Для совершенствования результатов поиска к некоторым словарям синонимов составлялся словарь стоп-слов, в сочетании с которыми результаты поиска были не релевантными (например, стоп-слово «день» для словаря ценности «жизнь»; это исключало из поиска варианты «день жизни», «несколько дней жизни» и т. д.).

Далее словари использовались для настройки узла «Таксономия», который выполнял поиск в текстах отзывов слов, словосочетаний, связей между словами из словаря синонимических рядов каждой категории, обозначающей духовно-нравственную ценность. Поиск был настроен при помощи языка PDL<sup>4</sup>. Это позволило включить в результаты поиска отзывы, где искомые слова или словосочетания в пределах предположения были разделены одним или несколькими словами (например, по запросу «*семейный круг*» были найдены результаты «*семейный тесный круг*»).

Настройка узла «Таксономия» соответствовала следующей логике:

- сначала текст анализировался на наличие слов или словосочетаний из списка синонимов каждой ценности;
- затем определялась значимость встречаемости отдельных слов и словосочетаний в текстах. В случае, если в одном или нескольких предложениях отзыва встречалось необходимое слово (в любом времени, падеже и роде) или словосочетание (вне зависимости от того, сколько слов разделяло слова в этом словосочетании), то для каждой ценности засчитывалось одно упоминание. Исключение составили лишь комбинации слов со стоп-словами, которые не учитывались при подсчёте;
- наконец, численные данные интерпретировались в выводы о значимости отдельных понятий: чем чаще в текстах отзывов упоминались синонимы или синонимичные словосочетания из списка соответствующей ценности, тем большее количество упоминаний засчитывалось этой ценности. В табл. 1 представлены результаты работы узла.

<sup>3</sup> <https://www.litres.ru/>

<sup>4</sup> PDL – язык для поиска лингвистических конструкций в неструктурированных текстах

Языком Остера и отчасти Хармса о жизни человека, прообразом которого является шизофреник и блогер из Оренбурга Сергей Шишкин – Сергей «Такой день».

1	Название книги	Автор	Ссылка на литрес	Рейтинг	Количество оценок	Количество отзывов	Имя читателя	Оценка книги	Отзыв	Лайки на отзывы	Дислайки на отзывы
7955	Убить Бобрыкина. Ист	Александра Николае	https://www	3,5	97	23	Konstantin Sazonov	3	Языком Остера и отчасти Хармса о жизни человека, прообразом которого является шизофреник и блогер из Оренбурга Сергей Шишкин – Сергей «Такой день».	3	3
7956	Дети мои	Гузель Яхина	https://www	4,4	8032	702	Евгения Бабунашви	5	Яркий, эмоциональный исторический роман. Увлекательно и красочно изложена история и судьба	0	1
7957	Патологии	Захар Прилепин	https://www	4,3	307	29	Sphinks-pro	4	Ярко, эмоционально, мощно, убедительно, но несколько коряво. Пролог и вовсе не к месту.	6	1
7958	Зулейха открывает гла	Гузель Яхина	https://www	4,6	24719	2103	Татьяна Коблова	5	Яркое впечатление. Отличная книга. Хочется перечитать ещё раз. Написана красивым литературным языком.	0	1
7959	Дети мои	Гузель Яхина	https://www	4,4	8032	702	Victoria ADAWY	5	Яхина очень талантливыи писатель. Ее стиль легкий и каждому дано увидеть то, что было задумано. Книга	0	0
7960	Дети мои	Гузель Яхина	https://www	4,6	1921	180	rehau-systems	5	Яхина талантливая писательница. ее книга всегда трогает самые дальние уголки сердца и вызывает интерес. эта	0	0
7961	Дети мои	Гузель Яхина	https://www	4,6	1921	180	kompstroy	5	Яхтная и этой книгой поразила и проникла в самое сердце. Красивый язык обволакивает и сейчас немного	0	0
7962											
7963											
7964											

Рис. 1. Фрагмент исходной базы данных для последующей обработки

Сохранить Импорттировать записи Экспортировать записи

Жизнь

Фильтр: Включить все Существительное Глагол Наречие Прилагательное Частица Местоимение Числительное Другое Любая

Класс

- Взаимопомощь
- Гражданственность
- гуманизм
- Достоинство
- духовный
- Единство народов России
- Жизнь**
- Идеал
- Коллективизм
- милосердие
- Ответственность
- Патриотизм
- Право
- Преемственность поколений
- Свобода
- Семья
- Служение
- справедливость
- Труд
- Уважение
- ценности

Леммы:

- 1 биография
- 2 Бытие
- 3 бытование
- 4 действительность
- 5 дни
- 6 жизненный процесс
- 7 жизненный путь
- 8 Жизнь
- 9 житейское море
- 10 Житье
- 11 Существование
- 12 уклад

Лемма: биогра...

Часть речи

- Включить все
- Существительное
- Глагол
- Наречие
- Прилагательное
- Частица
- Местоимение
- Числительное
- Другое
- Любая

Найти

Описание:

Количество: 21

Рис. 2. Словарь синонимов к слову «жизнь»

**Количество и иерархия ценностно-смысловых компонентов  
в корпусе читательских отзывов о книгах**

№	Корневой узел	8278
1	Жизнь	1601
2	Милосердие	673
3	Труд	576
4	Достоинство	474
5	Единство народов России	463
6	Духовность	340
7	Семья	310
8	Справедливость	281
9	Уважение	208
10	Служение	201
11	Идеал	168
12	Право	142
13	Свобода	135
14	Патриотизм	84
15	Преимственность поколений	50
16	Ответственность	49
17	Гуманизм	34
18	Взаимопомощь	4
19	Коллективизм	3
20	Гражданственность	1
	Прочие	4715

The screenshot shows the PolyAnalyst web report interface. On the left, a taxonomic tree (таксономия) is displayed, listing root nodes and their counts: Жизнь (1601), Достоинство (474), Право (142), Свобода (135), Патриотизм (84), Гражданственность (1), Служение (201), Ответственность (49), Идеал (168), Семья (310), Труд (576), Духовный (340), Гуманизм (34), Милосердие (673), and Справедливость (281). The 'Гуманизм' node is selected. The main area shows a detailed view of a text review. The text reads: "Текст" – первое моё знакомство с Дмитрием Глуховским. Абсолютно ничего не зная ни об авторе, ни о самом произведении, взялась за чтение и с первых страниц просто захватило, поглотило! "Текст" очень легко читается, простой язык, но тем не менее поднимаются серьёзные темы. Будьте готовы к нецензурным выражениям, жестокости, в каких-то моментах пошлости, но несмотря на все, это не отталкивает – **человечностью** пропитан весь роман. Привязываешься к... Below the text is a table with columns: #, F, R, Отзыв, Re..., Название кн..., and Автор. The table lists 14 entries, with the first entry highlighted: #1, F: 70, R: 3,71, Отзыв: "Текст" – первое моё знакомство с Дмитрием Глуховским, Re...: 70, Название кн...: Текст, Автор: Дмитрий Глуховский. The interface also includes navigation controls and a search bar.

Рис. 3. Фрагмент веб-отчёта о таксономии текстов по результатам автоматического поиска

На завершающем четвёртом этапе, «оформительском», таксономия текстов была конвертирована в табличный формат в виде веб-отчёта (рис. 3) и файла \*.xlsx для дальнейшего анализа результатов автоматического поиска и его реконфигурации<sup>5</sup>.

<sup>5</sup> <https://pa.data.tsu.ru:5043/polyanalyst/static/paclient/publication-view.html?reportUUID=44c6631c-63b7-46e9-ae94-ab1970784a81&guest=1>

## Результаты исследования

Основным итогом исследования стала база данных «Отзывы о художественных произведениях, содержащие сведения о культурном коде», которая отличается от аналогов не только процедурой формирования (дифференциация отзывов по типу сообщения – характеристике, показывающей, описываются ли в отзыве факты, выражаются эмоции или транслируются смыслы), но и тематикой, позволяющей определять

типичные образы и культурные доминанты у читателей современной литературы.

На этом этапе исследования мы провели анализ наименьшей части отзывов, а именно тех, которые имеют смысловой тип сообщения. Автоматический поиск позволил определить, что таковые составляют 20% всего корпуса анализируемых текстов (51% читателей описывают эмоции от восприятия книги, 29% излагают содержащиеся в книге факты).

Понимая смыслы как форму субъективного отношения человека к действительности, отражающую пристрастность сознания в суждениях, принимая во внимание их обусловленность уникальным личным опытом и биографией, мы включили в эту группу те отзывы, которые с очевидностью вписаны в личный культурный контекст читателя и показывают, что произведение «прожито», стало частью личной биографии.

Приведём примеры: «...Столько жизней сплелись воедино... В моей семье со стороны обоих родителей были раскулаченные, враги народа. Я понимала, что это боль... но то, как это описано в этой книге... как явно описаны и прожиты все чувства, эмоции, которые пережили наши предки тогда...» (отзыв о книге Г. Яхиной «Зулейха открывает глаза»); «Сколько волнений и переживаний вызывает его долгая, изнурительная, драматическая борьба за правду. Радость и гордость вызывает победа человека высокого духа. Пройдут годы и десятилетия, спадёт мутная литературная пена, а бойцы Чусовой, которых обессмертил Алексей Иванов, будут стоять вечно» (отзыв о книге А. Иванова «Золото бунта»).

Показательно, что именно смысловой тип сообщений проявлял такой ключевой компонент культурного кода, как идентификация. Книга помогала читателю осознать свою идентичность – этническую, национальную, религиозную, региональную. Вот один из примеров: «Для русского человека в этой

книге всё понятно. Я согласен с отзывами о двойственности, которое оставляет это произведение... но это и есть мы – попытка найти свою правду, попытка найти свою Родину, любовь к ней и одновременно ложь, несправедливость, равнодушие друг к другу и извечный вопрос – что делать? Каждый сам решает для себя – что делать? Книга заставляет задуматься» (отзыв о книге Захара Прилепина «Санька»).

На основе автоматического поиска нами было определено, что в отзывах с воспроизведением смыслов встречаются следующие традиционные российские духовно-нравственные ценности (табл. 2). Их иерархия несколько отличается от таковой по общему корпусу текстов, но в целом тенденция сходна.

Анализ данных по всему корпусу отзывов (см. табл. 1) и его сопоставление с результатами анализа «смыслового» блока (см. табл. 2) показывает, что основа культурного кода – передаваемые из поколения в поколение традиции и ценности – остаётся сохранённой. В релевантных отзывах читателей доминируют мысли о связи поколений и о важности исторической памяти. По-прежнему сохраняются такие ценностно-смысловые компоненты культурного кода, как традиционная семья и труд.

Вместе с тем,кратно менее в сознании читателей представлены такие ценности, как служение отечеству, родина, права и свободы человека, достоинство, благородство, человечность, справедливость и др. Практически не упоминаются в отзывах о книгах некогда базовые ценности: патриотизм, взаимопомощь и взаимовы уважение, коллективизм, братство, гражданственность. Возможно, это свидетельствует о том, что в условиях глобализации происходит трансформация культурного кода и замена традиционных российских ценностей, например коллективизма, ценностями иных культур: индивидуализм, конкурентность, космополитизм, приоритет материального над духовным.

Таблица 2

**Количество и иерархия духовно-нравственных ценностей в отзывах читателей со смысловым содержанием**

1	Память (86) и поколения (89)	175
2	Семья	160
3	Труд	142
4	Служение отечеству (0) / Родина (71), отечество (7), отчизна (1)	79
5	Права и свободы (1) / Права человека (26) / Свобода и независимость (34)	61
6	Достоинство (50) / Благородство (9)	59
7	Милосердие (6) / Сострадание (29) / Жалость (23)	58
8	Гуманизм (3) / человеколюбие (2) /человечность (29)	34
9	Справедливость (25) / Честность (12) / Верность (11) / Правосудие (2)	25
10	Жизнь (право на жизнь) (19)	19
11	Приоритет духовного над материальным / духовность (11), материальность (6)	17
12	Высокие нравственные идеалы / Идеалы (10), нравственность (3)	13
13	Патриотизм (8)	8
14	Взаимопомощь (2) и взаимовы уважение (0) / взаимовыручка (2)	4
15	Коллективизм (0), Братство (1)	1
16	Гражданственность	0

Побочным, но весьма важным результатом нашей работы стал составленный на основе отзывов со смысловым содержанием список современных книг, которые послужили импульсом для смыслопорождения и понимания собственной идентичности в сознании наиболее думающих читателей:

Гузель Яхина	«Зулейха открывает глаза»
Людмила Улицкая	«Даниэль Штайн, переводчик»
Евгений Водолазкин	«Лавр»
Леонид Юзефович	«Зимняя дорога. Генерал А. Н. Пепеляев и анархист И. Я. Строд в Якутии. 1922–1923»
Евгений Водолазкин	«Авиатор»
Даниил Гранин	«Блокадная книга. Часть первая»
Алексей Иванов	«Золото бунта»
Григорий Служитель	«Дни Савелия»
Валентин Распутин	«Прощание с Матёрой. Пожар (сборник)»
Виктор Пелевин	«Искусство легких касаний»
Захар Прилепин	«Санька»
Александр Чудаков	«Ложится мгла на старые ступени»
Михаил Елизаров	«Земля»

Этот список может быть использован в качестве рекомендательного в обучении молодого поколения по разнообразным образовательным программам [11] для организации дискуссий, круглых столов, семинаров, а также для создания собственных текстов, отражающих личный опыт и позицию автора, способствующих осознанию своей идентичности.

## ЗАКЛЮЧЕНИЕ

В научной литературе описаны разнообразные подходы к изучению культурного кода: социокультурный, психоаналитический, аксиологический, семиотический, лингвокультурологический и др. [12–14]. Тщательно и добросовестно выполненные исследования учёных создают прочную методологическую базу для дальнейшего изучения культурного кода в условиях стремительно меняющейся жизни. Ускорение темпов общественного развития вносит коррективы в культурный код, особенно в среде молодёжи, поэтому сегодня актуально изучение масштабных процессов и явлений, происходящих в этой сфере, важен поиск путей сохранения и укрепления своей культурной идентичности.

В процессе исследования мы пришли к выводу, что в настоящее время обновление устоявшихся гипотез и концепций относительно культурного кода невозможно без автоматического получения новых больших данных.

Можно с уверенностью полагать, что предложенная в настоящей статье методика извлечения ценностно-смысловых компонентов культурного кода из корпуса читательских отзывов о книгах, включающая составление поисковых запросов, детальное описание процесса подготовки данных и критериев для анализа, а также оценку полученных фактов, может

быть верифицирована при исследовании других массивов данных [15]. В процессе работы над проблемой мы не только использовали программы прикладного программирования (PolyAnalyst™), но и поставили задачу доказать их эффективность в изучении таких тонких и сложных гуманитарных феноменов, как культурный код. Нами был проведён эксперимент по изучению актуального состояния культурного кода россиян, определены векторы его трансформации и сделан однозначный вывод о возможности дальнейшего и более широкого применения описанной нами методики.

В перспективе очевидна необходимость создания методов автоматической обработки соответствующих данных на основе машинного обучения. Мы предполагаем, что оптимальной моделью машинного обучения может быть градиентный бустинг из библиотеки машинного обучения *scikitlearn* для классификации аннотированных сообщений как релевантных или не релевантных теме культурного кода. Предварительная оценка модели путём 5-кратной перекрёстной проверки показала точность: 0,765 ( $\pm 0,023$ ) [0,743; 0,776; 0,788; 0,749; 0,770], что приближается к эталонным результатам.

Таким образом наше исследование восполняет нехватку методик, позволяющих осуществлять автоматический анализ текстов в рамках изучаемой темы и применять для анализа данных о компонентах культурного кода статистические методы, а также методы машинного обучения.

## СПИСОК ЛИТЕРАТУРЫ

1. Сабден О.С. Мировая цивилизация и модернизация общественного развития // Век глобализации. – 2021. – № 1. – С. 116–128. DOI: <https://doi.org/10.30884/vglob/2021.01.10>
2. Балашова К.И. Национальные культуры в контексте всемирной глобализации // Молодой учёный. – 2021. – № 27(369). – С. 291–294. – URL: <https://moluch.ru/archive/369/83035/> (дата обращения: 10.03.2022).
3. Бирюкова М.А. Глобализация: интеграция и дифференциация культур // Философские науки. – 2000. – № 4. — С. 33–42.
4. Захарова Я.М., Зиновьева Е.Г. Культура России как фактор формирования имиджа страны // Традиционные национально-культурные и духовные ценности как фундамент инновационного развития России. – 2019. – № 2 (16). – С. 25–28.
5. Нарочницкая Н.А. Русский код развития. — М.: Книжный мир, 2013. — 352 с.
6. Cipriani G. The Ethics of Relation in the New Modernity of Digital Art and Culture // International Journal of cultural research. – 2021. – № 3 (44). – P. 6–17.
7. Spivak D. L. Intangible Heritage in Cultural Strategy of UNESCO // International Journal of cultural research. – 2021. – № 3 (44). – P. 114–130. DOI: [https://doi.org/10.52173/2079-1100\\_2021\\_3\\_114](https://doi.org/10.52173/2079-1100_2021_3_114)
8. Бенин В.Л. Основы духовно-нравственной культуры народов России: третье пришествие культурологии в современное российское образо-

- вание // Культурный код. – 2020. – № 1. – С. 7–16. DOI: <https://doi.org/10.36945/2658-3852-2020-1-7-16>
9. Шапинская Е.Н., Лисенкова А.А. Гуманитарное образование в цифровую эпоху: трансформация или кризис? // Культурный код. – 2020. – №1. – С. 7–16. – С. 42–54.
  10. Леонтьев Д.А. Психология смысла: природа, строение и динамика смысловой реальности. 2-е, испр. изд. – М.: Смысл, 2003. – 487 с.
  11. Мосунова Л.А., Борзова Т.В. Развитие национальной культурной идентичности студентов гуманитарных специальностей в процессе обучения продуцированию нарративных текстов // Перспективы науки и образования. – 2021. – № 6(54). – С. 141–154. DOI: <https://doi.org/10.32744/pse.2021.6.10>
  12. Гревнев В.М. Социокультурные коды как мировоззренческие и ментальные матрицы региональной идентичности // Вестник Кемеровского государственного университета культуры и искусств. – 2017. – № 41. – С. 58–63. – URL: <https://elibrary.ru/item.asp?id=32205692> (дата обращения: 22.02.2022).
  13. Степанова Н.И. Коды культуры: семиотический и культурологический аспекты // Идеи и идеалы. – 2012. – № 1(11). – С. 130–135. – URL: <https://cyberleninka.ru/article/n/kody-kultury-semioticheskiy-i-kulturologicheskiy-aspektu> (дата обращения: 23.01.2022).
  14. Симбирцева Н.А. «Код культуры» как культурологическая категория // Знание. Понимание. Умение. – 2016. – № 1. – С. 157–167. – URL: <https://cyberleninka.ru/article/n/kod-kultury-kak-kulturologicheskaya-kategoriya/viewer> (дата обращения: 24.12.2021).
  15. Митягина Е.В., Ашихмина Т.Я., Котельников Е.В. Методы анализа социальных медиа (на примере социальной сети ВКонтакте) // Advanced Science. – 2017. – № 1(5). – С. 30.

*Материал поступил в редакцию 29.03.22.*

#### **Сведения об авторах**

**МОСУНОВА Людмила Александровна** – доктор психологических наук, доцент, профессор кафедры журналистики и интегрированных коммуникаций Вятского государственного университета  
e-mail: [imosunova@hotmail.com](mailto:imosunova@hotmail.com) 8-912-828-48-91

**МИТЯГИНА Екатерина Владимировна** – доктор социологических наук, доцент, проректор по развитию на основе анализа данных Вятского государственного университета  
e-mail: [mityagina2004@yandex.ru](mailto:mityagina2004@yandex.ru)

**АНАНИН Павел Валерьевич** – старший преподаватель кафедры социальной работы и молодежной политики Вятского государственного университета  
e-mail: [vp014@mail.ru](mailto:vp014@mail.ru)