

Использование программ анализа данных в проведении историко-философского исследования

Я. А. Коховец,
аспирант,
Белорусский государственный университет

Цель статьи – анализ проблемы и представление результатов изучения возможности применения программ анализа данных в историко-философских исследованиях на примере выявления сходств между экзистенциализмом и гностицизмом. Подробно обсуждаются проблемные пункты использования крупных языковых моделей. Рассматриваются программные средства как инструменты для структурирования и анализа текстовых данных, что позволяет выявить ключевые концепты в таких источниках, как труды Мартина Хайдеггера и тексты библиотеки Наг-Хаммади.

Ключевые слова: история философии; data-анализ; концепт-анализ; гностицизм; экзистенциализм; Мартин Хайдеггер; методология; междисциплинарный подход.

The article aims to analyse the problem and present results of the possibility of application data analysis programs in historical and philosophical research, using the example of identifying similarities between existentialism and Gnosticism. It discusses in detail the problematic aspects of using large language models (LLMs). Software tools are considered as instruments for structuring and analyzing textual data, enabling the identification of key concepts in such sources as the works of Martin Heidegger and the texts of the Nag Hammadi library.

Keywords: history of philosophy; data analysis; concept analysis; Gnosticism; existentialism; Martin Heidegger; methodology; interdisciplinary approach.

«Между тем, что я думаю, тем, что я хочу сказать, тем, что я, как мне кажется, говорю, и тем, что вы хотите услышать, тем, что, как вам кажется, вы слышите, тем, что вы слышите, тем, что вы хотите понять, тем, что вы понимаете, стоит десять вариантов возникновения непонимания. Но все-таки давайте попробуем» [1].

Одной из главных герменевтических проблем, с которой сталкивается исследователь, поставивший перед собой конкретную цель исследования, является вопрос о взаимообусловленности интерпретатора и текста, в соответствии с которым интерпретатор привносит в понимание текста свои предположения, которые затем корректируются в процессе интерпретации. Таким образом, процесс понимания зависит от позиции интерпретатора, которая эволюционирует в ходе прохождения герменевтического круга. Это расширяет горизонты восприятия интерпретатора,

делая процесс понимания потенциально бесконечным.

Историко-философские исследования часто требуют работы с обширным текстовым материалом (философские трактаты, письма, дневники и труды различных мыслителей). Современные программные инструменты могут существенно облегчить этот процесс, сделав его более управляемым и эффективным.

Современные технические возможности и программные средства, такие как нейронные сети и программы для анализа данных, предоставляют возможность обрабатывать большие объемы информации с минимальным участием человека на этапе выделения семантических ядер текста. Это позволяет исследователю сосредоточиться на роли оператора программы и конечного интерпретатора полученных результатов. Таким образом, человеку не нужно вручную

анализировать огромные массивы данных, вместо этого он может использовать эти инструменты для автоматизации рутинных задач и фокусироваться на более сложных аспектах интерпретации и анализа.

Машинное обучение как наука появилось еще в 1956 г. с Дартмутского семинара, однако настоящий подъем случился относительно недавно. Его принято связывать с увеличением вычислительных мощностей и доступности данных (развитие сети Интернет), так как для современных моделей требуются огромные объемы данных (45 терабайт на ChatGPT-3) и вычислительных мощностей для тренировки ($3.14E + 23$ флопс), что стало доступно только сейчас [2].

На данный момент возможно выделить два подхода к обработке естественного языка, хотя зачастую они комбинируются:

1. Векторный, или так называемые эмбединги – представление текста в виде векторов единого векторного пространства. Например, векторы для «короля» – «мужчина» + «женщина» должны быть близки к вектору для «королевы».

2. «Традиционный» – основан на наборах правил, зачастую индивидуальных для каждого языка.

Обучение векторной модели, как правило, ведется без «учителя», но требует большого объема данных. Алгоритмом обучения является, например, CBOW: модель предсказывает текущее слово на основе контекста (окружающих слов). Так, в предложении «Собака лает на _____», CBOW попытается угадать недостающее слово (например, «почтальон») на основе контекста.

Векторные модели позволяют решать широкий спектр задач, в том числе создавать единое векторное пространство с изображениями. Однако они не являются ультимативным решением, так как при «понимании» текста не позволяют его анализировать с той же глубиной, что и при «традиционном подходе». Таким образом, векторный подход хорошо себя показывает на таких задачах, как перевод, конспектирование, классификация, вопрос-ответные и диалоговые системы.

«Традиционный» подход решает более точечные задачи: токенизация; определение частей речи и связей между ними в предложении; лемматизация (выделение атомарной леммы из каждого слова); выделение имен собственных.

Как правило, «традиционный» подход интегрируется в алгоритм как составные элементы для решения более сложных и нетипичных задач [3].

В итоге такие LLM, как, например, ChatGPT – вершина развития векторного подхода на данный момент, позволяют эффективно автоматизировать задачи, требующие «понимания» текста, однако все еще предусматривают человеческие усилия для четкой формулировки системного запроса к модели, чтобы определить требования к ее поведению (так называемый промпт), а также интерпретации и проверки полученных результатов. Кроме того, эти модели могут быть нестабильны (например, появление «галлюцинаций» ИИ) и требуют много вычислительных мощностей. Также наблюдается закономерность, при которой «нейронные сети» при анализе и формулировании ответа на конкретный вопрос вторят формулировке оператора, тем самым выстраивая свой ответ таким образом, чтобы он соответствовал желаниям. Нейронные сети выдают ответ в виде определенной вероятности. Это означает, что они могут предоставлять результаты только с некоторой степенью уверенности. Например, если сеть «считает», что ответ сильно вероятен, она может дать его как правильный, тем самым создавая впечатление, что она «угодает» автору вопроса. Приведенные нюансы функционирования LLM на данный момент (до их устранения) ограничивают их использование в общегуманитарном и, в частности, историко-философском исследовании, так как своей «вероятностной» архитектурой добавляют в исследование еще один слой интерпретации.

Компромиссным решением может стать программное обеспечение, нацеленное не на «понимание» и генерацию, а на анализ, извлечение и структурирование информации. Такие программы, в отличие от LLM, применяют не методы машинного обучения, а статистические алгоритмы и методы анализа данных.

Подобные современные средства обработки текста условно можно разделить на две категории: программы лексического (количественного) и качественного анализа данных.

Примеры программ качественного анализа:

- Tropes¹ – программа, разработанная во Франции в 2009 г., представляет собой бесплатное программное обеспечение для семантической

¹ <https://www.tropes.fr/download.htm>.

классификации, извлечения ключевых слов, лингвистического и качественного анализа. Находит применение в различных областях, таких как маркетинговые и медицинские исследования, социологический анализ.

- Leximancer² – программное обеспечение для анализа текстов, автоматически выявляющее высокоуровневые концепции, представляет собой мощный инструмент для обработки и интерпретации больших объемов информации. Анализирует тексты и предоставляет результаты в виде концептуальных карт, что позволяет пользователям получить общее представление о содержании материала и увидеть ключевые концепции, встречающиеся в тексте.

- Catma³ – веб-приложение для разметки и анализа текста, разработанное в университете Гамбурга. Предоставляет возможности не только для анализа данных, но и аннотации текста.

- Atlas.ti⁴ – мощное приложение для качественного анализа разнообразного контента, включающего текстовые, графические и видеоданные. Помогает исследователям выявлять ключевые паттерны, тенденции и зависимости в массиве данных, проверять гипотезы, выдвинутые относительно анализируемых материалов.

Примеры программ количественного анализа текста:

- AntConc⁵ – бесплатный инструмент, создающий на основе анализа текста конкордансы (конкорданс – особый тип словаря, в котором каждое слово или понятие расположено в алфавитном порядке с минимальным контекстом и всеми случаями употребления в некотором тексте; список словоупотреблений с отсылками ко всем контекстам) [4].

- DocFetcher⁶ – инструмент индексирующего поиска, который использует локальную базу данных для быстрого и эффективного поиска информации. Вместо того чтобы каждый раз просматривать все файлы на компьютере, программа сначала индексирует заданные пользователем файлы и каталоги. После этого при выполнении поисковых запросов она использует этот индекс для мгновенного предоставления результатов.

- MultiTextFinder⁷ – осуществляет поиск по содержимому файлов, кроме этого, способна искать сразу несколько слов в документах. В отличие от DocFetcher, данная программа не требует предварительной индексации файлов, но поиск происходит значительно медленнее. Результаты поиска также содержат информацию о количестве совпадений указанного слова в тексте документа и сумму совпадений во всех документах. MultiTextFinder, как и DocFetcher, распространяется бесплатно, однако обе имеют PRO-версии, предоставляющие дополнительные функции за отдельную плату.

В связи с распространением такого рода программного обеспечения в разных областях знаний для нас видится перспективной возможность использования инструментария программ анализа данных в проведении историко-филологического исследования, в частности в контексте выявления предполагаемого сущностного сходства между экзистенциализмом и гностицизмом. Подобные программы позволят нам провести лексический и концептуальный анализ текстов, выявить специфическую лексику посредством поиска наиболее часто повторяющихся лексем, а также соотнести полученные результаты для двух систем мысли.

Для проведения качественного анализа текстов использовалась программа Atlas.ti как десктопный инструмент (в отличие, например, от браузерной Catma), способный к распознаванию текста с PDF-файлов. Источником для проведения анализа в контексте данной работы было выбрано оцифрованное в формате PDF полное издание трудов немецкого философа-экзистенциалиста XX в. М. Хайдеггера.

Труды Мартина Хайдеггера (Martin Heidegger Gesamtausgabe) начали издаваться еще при жизни автора и вплоть до наших дней публикуются в издательстве «Витторио Клостермана» (Vittorio Klostermann). На данный момент собрание насчитывает 102 тома. Очередность их выпуска была установлена самим М. Хайдеггером, что, в свою очередь, является причиной непоследовательной их публикации. Например, в качестве первого изданного тома в 1975 г. был опубликован том 24 «Фундаментальные проблемы феноменологии». Полное собрание трудов разделено на следующие отделы: 1) опубликованные при жизни сочинения 1910–1976 гг.; 2) фрайбургские и марбургские лекции 1919–1944 гг.;

² <https://www.leximancer.com/>.

³ <https://catma.de/>.

⁴ https://atlasti.com/?_gl=1*1hfx635*_up*MQ.*_ga*NzE0MDEyMTUwLjE3MjcwNzcwNTU.*_ga_K459D5HY8F*MTcyNzA3NzA1NC4xLjAuMTcyNzA3NzA1NC4wLjAuMA.

⁵ <https://www.laurenceanthony.net/software.html>.

⁶ <https://docfetcher.sourceforge.io/ru/index.html>.

⁷ <https://interestingsolutions.ru/MultiTextFinder.aspx>.

3) неопубликованные книги и лекции; 4) примечания и записки. План публикации полного собрания сочинений можно найти на страницах семидесятого тома [5, с. 201].

Томы 72 «Die Stege des Anfangs» («Мосты начал»), 92–93 «Ausgewählte Briefe I–II» («Избранные письма 1 и 2») на данный момент еще ожидают публикации, а анализ 14-го тома производится с его английского перевода ввиду недоступности немецкого оригинала.

Гностические труды представлены собранием апокрифических писаний библиотеки Наг-Хаммади. Работа над переводом найденных в Египте гностических текстов началась сразу после их обнаружения в 1945 г., однако первый перевод полного собрания текстов библиотеки Наг-Хаммади на немецкий язык был опубликован в Штутгарте в 1997 г. под названием «Bibel der Häretiker» («Библия еретиков») [6] Гердем Людеманном и Мартиной Янссен.

В качестве исходного материала для анализа свойственных гностическому нарративу семантических реперных точек использовались несколько изданий гностических текстов библиотеки Наг-Хаммади в немецких переводах с коптского. Анализу подверглось не только издание Герда Людеманна и Мартины Янссен, но и переводы Ганса-Мартина Шенке серии «Nag Hammadi deutsch» [7; 8], а также адаптированное пособие для студентов [9].

Во избежание преобразования смыслов во время перевода оба источника исследования приводятся и анализируются на немецком языке, который для трудов М. Хайдеггера является языком оригинала.

Первой сложностью, с которой сталкивается оператор Atlas.ti (помимо огромного тьюториала и отсутствия русской локализации) при работе с такими источниками, как «Nag Hammadi Deutsch» и «Bibel der Häretiker», является то, что помимо оригинальных гностических текстов они снабжены большим количеством вводной информации и пояснений к каждому отдельному тексту. Все эти данные как часть массива текста будут проанализированы программой и проинтерпретированы ею количественно, без оглядки на содержание. Это обстоятельство заставляет исследователя при сравнении результатов выборки принимать во внимание наличие своеобразного «семантического шума» и учитывать его при анализе.

Программа после процесса определения концепций предлагает оператору 100 наиболее часто

встречающихся в источнике концепций и выдает их в виде «облака слов». При нажатии на конкретное слово Atlas.ti предлагает ознакомиться со всеми контекстами его употребления в тексте.

Например, сравним результаты концепт-анализа приведенных источников. Наравне с такими наиболее часто встречающимися понятиями, как Vater (Отец), Mensch (Человек), Kraft (Сила), Geist (Дух), Licht (Свет) и Gott (Бог), мы можем увидеть такие появляющиеся в выдаче результаты, как NHC (аббревиатура, обозначающая кодексы Наг-Хаммади), Vgl (аббревиатура, используемая для указания на необходимость сравнения с другим источником), Nag (часть названия поселения Наг-Хаммади).

Так как основные наиболее часто встречаемые концепты повторяются во всех трех источниках, это позволяет объединить их в одну группу текстов и анализировать уже в совокупности, отсекая «шум». При этом мы получаем следующий топ концептов по убыванию частоты упоминаний в тексте: Vater (Отец), Licht (Свет), Kraft (Сила), Mensch (Человек), Gott (Бог), Geist (Дух), Welt (Мир), Seele (Душа), Name (Имя), Ort (Место).

В отличие от гностических источников, приведенных с комментариями авторов, полное собрание сочинений М. Хайдеггера не содержит сторонних вкраплений «чужого» текста, и его анализ можно проводить сразу.

Наиболее часто встречающимися концептами, согласно Atlas.ti, в произведениях М. Хайдеггера являются: Wesen (Сущность), Mensch (Человек), Frage (Вопрос), Denken (Думать), Wahrheit (Истина), Sinn (Смысл), Seiende (Сущее), Zeit (Время), Dasein (Вот-бытие), Grund (Основа). Но это только топ-10 из ста наиболее часто встречающихся концептов.

Отдельного внимания заслуживает слово Mensch (Человек), которое не просто повторяется, но и входит в топ-10, находясь у М. Хайдеггера на третьем, а у гностиков – на четвертом месте.

Следующий этап анализа – поиск непосредственных отсылок к гностическому в текстах М. Хайдеггера. Для этого нам необходимо выбрать ключевые слова, части слов или словосочетания, по которым будет производиться поиск. В нашем случае в качестве запроса использовались слово Gnosis и корень gnost- как наиболее стабильная, неизменяемая часть слова. В целях большей достоверности и возможности параллельно сравнить результаты работы данный вид

анализа проводился одновременно при помощи программ MultiTextFinder и DocFetcher.

В результате количественного анализа выясняется, что слово Gnosis употребляется М. Хайдеггером один раз в томе 63 полного собрания сочинений в четвертой главе «Онтологии» – «Понятие человек в библейской традиции». Автор отсылает нас непосредственно к «Валентинианскому гнозису», обвиняя М. Шеллера в гностической интерпретации понятия «человек» и приводя его цитаты, описывающие Человека как «Богоискателя», «Того, что между», «Намерение и жест трансценденции» и «Живой образ бога». Продолжая критику философской антропологии М. Шеллера, М. Хайдеггер видит в его словах аналогию с гностической трихотомией «тело/душа/дух», обвиняя философа в подмене философии теологией [10].

Стоит упомянуть, что MultiTextFinder выдал также совпадение в томе 15, распознав частицу gnosis как часть слова Diagnosis. DocFetcher выдал только «Онтологию».

Переходя к поиску по части слова, следует отметить, что обе программы показали идентичный результат. Это свидетельствует о проверяемости данных и надежности метода анализа.

Итак, при помощи лексического анализа полного собрания сочинений нам удалось выявить шесть эпизодов упоминания М. Хайдеггером «гностического». Результаты анализа вместе с результатом для запроса Gnosis представлены в таблице.

Благодаря проведенному концепт- и лексическому анализу сочинений М. Хайдеггера мы можем с уверенностью утверждать, что, хотя философ обращался к гностицизму и «гностическому» в целом в своих произведениях крайне редко, он, несомненно, был достаточно хорошо знаком с этим феноменом.

В это же время гностицизм оказывается настолько же человекоцентричным, как и экзистенциализм. Темы познания, воли, истины, божественного, а главное – бытия и природы человека в этом мире являются не просто точками пересечения для обеих систем мысли, но при ближайшем рассмотрении обнаруживают схожий взгляд на глубинное взаимодействие человека с бытием, которое выходит за рамки рационального познания, открываясь в экзистенциальном опыте раскрытия и сопричастности к таинственному, неизведанному измерению реальности.

При использовании программ качественного и количественного анализа текста в философском исследовании исследователю предоставляется инструмент выделения ключевых слов не только самих по себе, но и в совокупности контекстов, а также возможность аннотирования текста, что, несомненно, облегчает работу по обработке огромных массивов информации. Использование программ анализа текста значительно обогащает и облегчает историко-философские исследования благодаря своей способности систематизировать и количественно оценивать текстовые данные, позволяет более глубоко понимать структуру и содержание текстов, выявлять скрытые паттерны и идеи, которые могут быть неочевидны при традиционном чтении, составлять своеобразную ментальную карту структуры мышления автора, которая прячется за литературными тропами.

Функционал подобного программного обеспечения позволяет исследователю не только производить анализ единичных текстов, но и выйти на уровень мета-анализа. Также инструментарий программ data-анализа способствует проведению «аутоисследования», при котором автор становится объектом собственного

Таблица

Места упоминания М. Хайдеггером гностического в полном собрании сочинений

Место употребления	Страница	Словоформа
Том 17 «Введение в феноменологические исследования»	192	«Gnostikern»
Том 22 «Основные понятия античной философии»	57	«Gnostische»
	230	«Gnostizismus»
Том 33 «Лекции 1923–1944 гг. Аристотель, метафизика θ 1–5»	147	«gnostisch-orientalischen»
Том 63 «Лекции. Онтология (Герменевтика фактичности)»	25	«Gnosis»
Том 73.2 «Неопубликованные лекции – задуманное. О событийном мышлении»	1483	«gnostisch»
Том 97 «Черные тетради 1942–1948. Примечания и записи. Примечания I–V»	144	«paulinisch-gnostisch-römisch-hellenistische»
Том 100 «Черные тетради 1952/1953–1957. Примечания и записи. Vigiliae und Notturmo»	120	«Gnostizismus» (дважды)

изучения, что позволяет глубже понять свои предвзятости, убеждения и циклическое влияние герменевтического круга, ведь в данном случае он может выяснить, как его личный опыт,

восприятие и структура мышления, выраженные в тексте, влияют на процесс сбора и анализа данных, что может улучшить объективность результатов.

Список использованных источников

1. *Вербер, Б.* Империя ангелов / Б. Вербер; пер. А. Кабалкина. – Москва: ООО «Издательство «Эксмо», 2021. – URL: <https://fantasy-worlds.org/lib/id32016/read/> (дата обращения: 16.09.2024).

2. *Language Models are Few-Shot Learners* / T. Brown, B. Mann, N. Ryder [et al.]. – 2022. – URL: <https://arxiv.org/pdf/2005.14165> (дата обращения: 01.10.2024).

3. *Jurafsky, D.* Natural Language Processing: an Introduction / D. Jurafsky, J. Martin. – 2011. – URL: <https://academic.oup.com/jamia/article/18/5/544/829676> (дата обращения: 01.10.2024).

4. *Доминикан, А. И.* Конкорданс как инструмент для изучения значения слова / А. И. Доминикан // Вестник ТвГУ. Серия: Филология. – 2017. – № 2. – URL: <https://core.ac.uk/download/pdf/83100161.pdf> (дата обращения: 13.09.2024).

5. *Heidegger, M.* Gesamtausgabe Band 70: Über den Anfang / M. Heidegger. – Frankfurt am Main: Vittorio Klostermann, 2005. – 1 электрон. опт. диск (CD-ROM).

6. *Lüdemann, G.* Bibel der Häretiker / G. Lüdemann, M. Janßen. – Göttingen: Vandenhoeck & Ruprecht, 1997. — 1 электрон. опт. диск (CD-ROM).

7. *Schenke, H.* Nag Hammadi Deutsch, Band 1: NHC I,1-V,1 / H. Schenke, H. Bethge, U. Kaiser. – Berlin: Walter de Gruyter, 2001. – 1 электрон. опт. диск (CD-ROM).

8. *Schenke, H.* Nag Hammadi Deutsch, Band 2: NHC V,2 – XIII,1, BG 1 und 4 / H. Schenke, H. Bethge, U. Kaiser. – Berlin: Walter de Gruyter, 2003. – 1 электрон. опт. диск (CD-ROM).

9. *Kaiser, U.* Nag Hammadi Deutsch: Studienausgabe NHC I–XIII, Codex Bezae Cantabrigiae 1 und 4, Codex Bezae Cantabrigiae 3 und 4 / U. Kaiser, H. Bethge. – Berlin: Walter de Gruyter, 2013. – 1 электрон. опт. диск (CD-ROM).

10. *Heidegger, M.* Gesamtausgabe Band 63: Ontologie (Hermeneutik der Faktizität) / M. Heidegger. – Frankfurt am Main: Vittorio Klostermann, 1988. – 1 электрон. опт. диск (CD-ROM).

Рукапіс паступіў у рэдакцыю 24.01.2025.

ГУО «Республиканский институт высшей школы» Редакционно-издательский центр предлагает

Под редакцией Э. А. Усовской

КУЛЬТУРНЫЕ ИССЛЕДОВАНИЯ

Допущено Министерством образования Республики Беларусь в качестве учебного пособия для студентов учреждений высшего образования по специальностям культуры и искусства

В учебном пособии рассматриваются сущность и основные направления культурных исследований, выступающих одной из наук о культуре, аналогом принятой в отечественном дискурсе культурологии. Понимание культуры как многогранного явления, способа человеческого существования, охватывающего разные сферы – от повседневности и продуктов массовой культуры до исторической памяти и наследия, – сможет обогатить студентов новыми знаниями и компетенциями.

Предназначено для студентов, обучающихся по специальностям 1-21 04 01 «Культурология (по направлениям)», 6-05-0211-06 «Издательское дело», а также для всех, кто интересуется науками о культуре.

ISBN 978-985-586-892-8



Информацию о реализуемой учебной и методической литературе можно посмотреть на сайте www.nihe.bsu.by.
Заказы принимаются по адресу: 220007, г. Минск, ул. Московская, 15, к. 126, тел./факс 251 06 63.