

Беляева А.В.
ВИЗУАЛИЗАЦИЯ КАК СПОСОБ ИЗУЧЕНИЯ
ИСТОРИЧЕСКОГО ИСТОЧНИКА

Ключевые слова: визуализация, исторический источник, портрет документа, МАХQDA, США, Китай.

Работа историка зачастую связана с классификацией, систематизацией и анализом массовых письменных источников. Подобная работа трудоемка, отнимает много времени, требует концентрации внимания. В силу этих обстоятельств она не застрахована от ошибок и неточностей, которые, в свою очередь, влияют на конечный результат исследования и сделанные выводы. Историческая наука, и без того критикуемая за субъективность, в данном случае вновь дает повод себя критиковать.

В этой связи актуальным и насущным является поиск путей выхода из сложившихся обстоятельств. Одним из них стало сотрудничество с точными науками. Развитие же новых технологий привело к возникновению новой междисциплинарной области исследования – Digital Humanities, которая объединяет в себе гуманитарные науки и компьютерные технологии. Толчком к сближению столь, казалось бы, далеких друг от друга областей знания послужил целый ряд факторов, но главным из них является появление персонального компьютера и сети Интернет. Прежде всего, у ученых появилась возможность работать и обмениваться информацией из любой точки мира, они перестали быть привязаны к учреждению, где имелась в наличии вычислительная техника. Кроме того, благодаря этому возможность взаимодействовать с техникой появилась у большего числа ученых.

Однако главным преимуществом персонального компьютера является наличие широкого спектра программного обеспечения, которое позволяет использовать компьютер для большого числа задач в русле исторической науки. А сеть Интернет позволяет не только обмениваться мнениями, знаниями и опытом, но и учиться. На сегодняшний день существует огромное количество курсов, дающих возможность, даже бесплатно, обучиться пользованию различными компьютерными программами, которые можно использовать в рамках исторических исследований. Например, можно создавать карты (в том числе интерактивные), 3d-реконструкции зданий и сооружений, создавать изображения исторических личностей по описаниям внешности и т.д. Виртуальное воссоздание реально существовавших объектов по описаниям является трудоемким и сложным процессом, но у ученых имеются привязки к точным данным – географические координаты, размеры и прочее. Иная ситуация с визуализацией письменных источников, которые являются основным типом исторических источников.

Визуализация данных – отдельная область исследований, которая позволяет представить сложные данные в более доступной, наглядной форме. Она делает возможным установление связей между различными документами, событиями или личностями, упомянутыми в этих документах. Она позволяет выявить общие тенденции и/или аномалии, которые могут указать на важные события или процессы в развитии общества, а также отследить изменения в языке и стиле источников.

Визуализация данных письменных источников в исторической науке решает несколько задач.

Во-первых, она оказывает помощь в анализе данных. Сведения из массовых источников, представленные в удобной форме, упрощают процесс исследования и повышают точность полученных результатов. Исследователь задает параметры поиска и форму представления результатов, которые применяются ко всему массиву информации, что исключает возможность упустить что-то из-за невнимательности. Условия поиска

могут быть быстро изменены, что предоставляет широкие исследовательские возможности в короткий срок. Историк получает возможность извлечь из источника разную информацию, а визуализация позволяет объединить ее в одном месте. Разрозненные данные сложнее воспринимать, нежели график или диаграмму. Таким образом, визуализация упрощает и ускоряет работу историка, а также повышает точность результатов за счет исключения ошибок из-за невнимательности.

Во-вторых, визуализация данных может являться дополнительным аргументом в дискуссии: «Что же так важно в изображениях и записях, получением, рисованием, изучением, расчетом и обсуждением которых заняты ученые и инженеры? Это прежде всего уникальное преимущество, которое они дают в риторической или дискуссионной ситуации. “Вы сомневаетесь в том, что я говорю? Я покажу вам”. И, не двигаясь более чем на несколько дюймов, я разворачиваю перед вашими глазами символы, схемы, иллюстрации, тексты, контуры и тотчас же представляю далекие отсюда вещи, с которыми теперь установлена своего рода двухсторонняя связь. Не думаю, что важность этого простого механизма может быть переоценена» [2, с. 117].

К. Кнорр справедливо заметила, что изображение или схема также допускают различия в трактовках, кроме того, оппоненту в дискуссии вообще не обязательно на них смотреть [3, р 205]. Однако, чем больше у одной из сторон спора возможностей защитить свою точку зрения, тем более прочные у нее позиции и более высокие шансы на победу: «Тот, кто плохо визуализирует, проигрывает схватку: его факт не выстоит» [2, с. 124]. В исторической науке много дискуссионных тем и разнообразие мнений по одному и тому же вопросу, что делает данную возможность визуализации весьма актуальной.

При понимании преимущества, даваемого визуализацией данных, возникает вопрос о том, какие данные будут представляться в наглядной форме. Процесс их визуализации достаточно прост: занесение данных в таблицу и построение на ее основе графиков, диаграмм и т.д. Весь процесс, может быть организован с помощью табличного процессора Microsoft Excel, который в наличии практически у любого пользователя персонального компьютера. Но в гуманитарных науках данные многозначны и разнообразны, они могут носить субъективный, противоречивый характер. Однако, как уже было сказано выше, современные технологии значительно расширяют доступный историку арсенал возможностей для работы с историческими источниками, в том числе с письменными.

Современное программное обеспечение для персональных компьютеров позволяет анализировать тексты с помощью метода контент-анализа. Компьютерных программ, позволяющих осуществлять его, достаточно много, однако наиболее полно соответствует проведению исторических исследований программный продукт MAXQDA [1].

Несомненным достоинством данной компьютерной программы является предоставление ею инструментов для визуализации. К функциям визуализации относятся «Карта кода», «Карта документов», «Диаграмма сравнения документов», «Портрет документа», «Кодовая строка».

«Карта кода» представляет собой инструмент, который в отдельном окне, называемом картой, показывает соотношения между кодами: чем чаще коды были заданы одновременно, тем ближе друг к другу расположены они будут на карте. «Карта документов» работает схожим образом, но в отношении документов: чем более схожи два документа с точки зрения распределения кодов, тем ближе друг к другу они будут на карте.

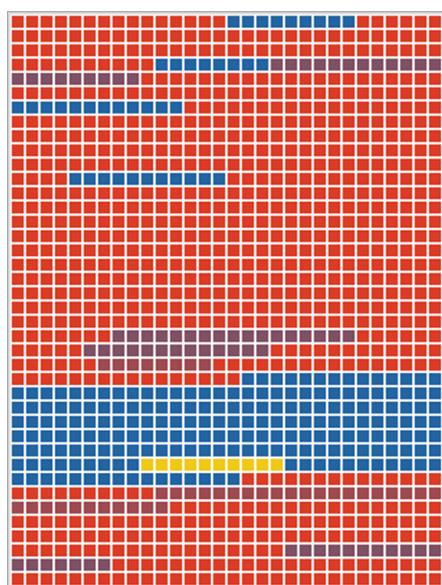
«Портрет документа» преобразовывает кодированные сегменты в стандартизированное количество точек изображения, тем самым позволяет наглядно продемонстрировать встречаемость кодов и их распределение по тексту. «Кодовая

строка» отображает кодированные сегменты в последовательности, позволяет установить, в каком абзаце текста встречается тот или иной код, и выявить незакодированные части текста. «Диаграмма сравнения документов» объединяет функции «Портрета документа» и «Кодовой строки». Это инструмент для сравнения закодированных частей документов.

Функция «Портрет документа» может быть использована для сопоставления различных исторических документов между собой, для выявления ключевых тем и идей текстов. Она позволяет получить представление о документе без необходимости детально читать его, что полезно для сортировки и классификации письменных источников, быстрого поиска необходимой темы в большом объеме данных.

Кроме того, функция «Портрет документа» может быть использована для отражения изменений внимания к конкретной проблеме в определенных хронологических рамках. Такое использование «Портрета документа» уже носит исследовательский характер. Например, данная функция позволяет визуализировать содержание ежегодных выступлений президента США перед Конгрессом «О положении страны» с целью определить степень внимания в них к Китаю. Для этого необходимо было установить соотношение внутри- и внешнеполитических тем в рамках речей «О положении страны», чтобы выявить контекст упоминаний.

Для достижения поставленной цели был выбран индуктивный подход, чтобы двигаться от источника. На основе данных количественного контент-анализа был составлен список кодов, которые были разделены на три категории, каждой из которых был назначен определенный цветовой индикатор – внутренняя политика (красный цвет), внешняя политика (синий цвет), Китай (желтый цвет). Затем было произведено кодирование всех источников.



1994 г.

Красный цвет - внутренняя политика
Синий цвет - внешняя политика
Желтый цвет - упоминание Китая

Рисунок. Пример
«Портрета документа»

можно было представить в виде таблицы или построенного на ее основе графика. Однако встроенные функции визуализации MAXQDA в сочетании с возможностью сохранить полученный «Портрет документа» как изображение ускоряют и упрощают

Цветовые маркеры, которые были присвоены отдельным темам, необходимы для использования «Портрета документа». Он позволяет наглядно продемонстрировать соотношение интересующих пользователя тем в документе. В нашем случае «Портрет документа» позволяет установить соотношение между внутри- и внешнеполитическими проблемами внутри выступлений президента США перед Конгрессом и место Китая в речах. Визуализация позволила установить фрагментарность встречаемости Китая в выступлениях Б. Клинтона и Дж. Буша-мл, усиление интереса в годы президентства Б. Обамы, а в годы пребывания в Белом Доме Д. Трампа китайская проблематика затрагивалась им в выступлениях перед Конгрессом ежегодно. Ввиду многочисленности «портретов» речей «О положении страны» в качестве наглядного примера использован лишь один из них (рис).

Возможности MAXQDA позволяют экспортировать данные в Microsoft Excel, поэтому частоту встречаемости слова «Китай» в документе

процесс работы, ведь в рамках одной компьютерной программы собраны инструменты для анализа и визуализации текстов.

Таким образом, визуализация является мощным инструментом, который обогащает традиционные методы анализа письменных исторических источников. Она позволяет исследователям не только экономить время при работе с большими массивами данных, но и открывает новые перспективы анализа и интерпретации исторических источников.

1. Гарскова, И. М. Новые тенденции в компьютеризованном анализе текстов: концепции, методы, технологии / И. М. Гарскова // История : электронный научно-образовательный журнал. – 2015. – Т. 6. – Выпуск 8(41). – URL: <http://history.jes.su/s207987840001255-9-1> (дата обращения: 18.02.2025).

2. Латур, Б. Визуализация и познание: изображая вещи вместе / Б. Латур // Логос: философско-литературный журнал. – 2017. – № 27(117). – С. 95–156.

3. Knorr, K. *be Manufacture of Knowledge* / K. Knorr. – Oxford : Pergamon Press, 1981. – 189 p.

Косачев Д.П.
АВТОМАТИЗАЦИЯ СБОРА И АНАЛИЗА МЕТАДАНЫХ СТАТЕЙ
ИЗ ЦИФРОВОГО АРХИВА «НЬЮ-ЙОРК ТАЙМС»
С ИСПОЛЬЗОВАНИЕМ PYTHON

Ключевые слова: историческая информатика, Нью-Йорк Таймс, контент-анализ, Великая Отечественная война, образ СССР.

Исследование роли медиа в формировании общественного мнения и их влияния на ключевые политические процессы в период Второй мировой войны представляет собой важный аспект исторического анализа. В условиях глобального конфликта средства массовой информации, такие как газеты, становились не только источниками информации, но и инструментами пропаганды, консолидации общества. Американские ученые отмечают, что «исследования неизменно выявляют более существенные связи между воздействием медиа и общественным мнением в вопросах внешней политики, чем в вопросах внутренней политики. Эта тенденция частично объясняется тем, что граждане имеют прямой личный опыт во многих сферах внутренней политики, но вынуждены в большей степени полагаться на информацию из массмедиа для оценки и анализа вопросов внешней политики» [6, p. 101].

Цель данной работы заключается в сборе и анализе метаданных статей «Нью-Йорк Таймс» за период 1941–1945 гг., упоминающих СССР, с использованием современных методов автоматизированного сбора и обработки данных. Основное внимание уделяется методике получения базы данных, которая в дальнейшем может быть использована как основа для более детализированного контент-анализа содержания публикаций.

Для достижения поставленной цели был разработан скрипт на языке программирования Python, позволяющий автоматизировать сбор метаданных электронного архива «Нью-Йорк Таймс» через API (Application programming interface). Это позволило создать структурированную базу данных, включающую такие поля, как заголовки статьи, даты публикаций, авторы и ключевые теги, присвоенные сотрудниками газеты.

«Нью-Йорк Таймс» является одним из ведущих мировых печатных изданий на протяжении большей части XX в. В период Второй мировой войны «Нью-Йорк Таймс» обладала значительным влиянием, признаваемым членами правительства за её роль в формировании общественного мнения, особенно в вопросах внешней политики. Судья Верховного суда США в то время Феликс Франкфуртер даже рассматривал «Таймс»