науки о культуре в перспективе «digital humanities» : Материалы Международной конференции 3–5 октября 2013 г., Санкт-Петербург / Под ред.

Л.В. Никифоровой, Н.В. Никифоровой. — СПб. : Астерион, 2013. — 600 с.

Historizing topic models: A distant reading of topic modeling texts within historical studies

Topic modeling (TM) - это метод, используемый в новой "цифровой истории", который представляет собой методологию, основанную на данных, которая может быть ближе всего к выполнению обещания историка литературы Франко Моретти сделать возможным "дистанционное чтение" больших объемов текста. Вдохновленный этим обещанием, TM используется для исторических исследований с начала 2000-х годов, и это исследование представляет собой обзор современного состояния TM среди исторических исследований, давая историческое и методологическое введение в использование TM в исторических исследованиях. TM's впервые была разработана для интеллектуального анализа данных в рамках обработки естественного языка и машинного обучения в 1990-х годах и имела своим подавляющим преимуществом способность охватывать значительно больший объем данных по сравнению с традиционными методами. Основной используемой моделью темы является скрытое распределение Дирихле, которое позволяет использовать TM в качестве функции поиска, количественной проверки интуиции или в качестве инструмента обобщения для больших корпусов текстов. Имея много конкурирующих теорий и предположений, которые постоянно оспариваются и развиваются, TM сам по себе в настоящее время представляет собой очень активную область исследований в области компьютерных наук. Обзор исторических текстов берет свою отправную точку в качестве первой рецензируемой исторической статьи в 2006 году и конечной точки публикации исследовательской монографии firs в 2013 году и выявил 23 исторических исследования, использующих TM. Чтобы дать общий обзор области, исследования были рассмотрены с использованием количественного подхода дистанционного чтения и проанализированы в соответствии с академическим образованием авторов, полом, академическим стажем и страной академического учреждения; типом корпусов, языком, хронологией и географической направленностью. Результаты показали, что большинство авторов были младшими незанятыми исследователями мужского пола, в основном связанными с американскими университетами, а тексты состояли из значительного количества нестандартных онлайн-текстов. Несмотря на применение в исторических исследованиях, ТМ по-прежнему воспринимается как технологический подход, большинство авторов которого имеют опыт работы в технических дисциплинах. Корпуса, где основное внимание уделяется английским текстам с американской или глобальной направленностью и с акцентом на новейшую историю. В целом, ТМ, по-видимому, является возникающей, а не установленной исторической методологией.

Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts

https://zfdg.de/2017\_005