Редактирование статьи

|  |  |
| --- | --- |
| **Основные данные** | |
| Название статьи: | "Отчего же и не написать чего-нибудь?" Тематическая модель дневника гимназиста XIX века |
| Аннотация: | В статье рассматривается опыт построения тематической модели на основе корпуса текстов личного происхождения. Тематические модели позволяют выявить в корпусе текстов скрытые темы. Объектом исследования является дневник гимназиста, а позднее студента Императорской Медико-хирургической академии В.И. Чемезова (5.04.1845 - 7.05.1911), который он вел в период с осени 1860 по 1878 год. Впервые отрывки дневника В.И. Чемезова были опубликованы Н.Я. Эйдельманом в 1975 году, полный текст в машиночитаемом формате опубликован лишь недавно на сайте проекта "Прожито". Дневник содержит уникальную информацию о повседневной жизни гимназиста середины XIX века, истории школы и педагогики в России, описание отмены крепостного права, студенческих волнений и многое другое. Особенно интересно нам было рассмотреть изменение интересов и переживаний автора на протяжении времени, для чего мы применили к тексту источника метод тематического моделирования. Новизна исследования заключается в том, что это первая попытка применения метода тематического моделирования к данному источнику, да и, наверное, в целом к русскоязычному дневнику XIX века. Анализ текста проводился на языке программирования R. В качестве конкретного алгоритма моделирования была избрана модель латентного размещения Дирихле. В ходе работы было предложено решение сразу нескольких конкретных задач: предварительной подготовки текста, выбора оптимальных параметров моделирования, визуализации тематической модели и интерпретация тем. В итоге было получено наглядное представление тем дневника и их распределение во времени по дневниковым записям, что на наш взгляд наилучшим образом отражает изменение интересов и переживаний автора. |
| Журнал: | Историческая информатика |
| Рубрика: | Новые методы и технологии обработки исторических источников |
| Научное направление: | Исторические науки |
| Шифр специальности: | 07.00.02, 07.00.09 |
| Основной язык статьи: | русский |
| Я не автор этой статьи: | нет |
| Перевод на английский: | нет |

[Изменить](http://author.nbpublish.com/author_edit_main_data.php?id=36641)

|  |  |
| --- | --- |
| **Англоязычные данные** | |
| Название статьи: | "Why not write something?" Topic model of a 19th century gymnasium student's diary |
| Аннотация: | The article discusses the experience of implementations a topic model based on a corpus of texts of personal origin. Topic models investigate the hidden topics in the text corpus. The object of the research is the diary of a gymnasium student and later a student of the Imperial Medical-Surgical Academy V.I. Chemezov (04/05/1845 - 05/07/1911), which he wrote in the period from autumn 1860 to 1878. For the first time, fragments of Chemezov's diary were published by N.Ya. Eidel’man in 1975, the text in machine-readable format was published recently on the «Prozhito» project website. The diary contains unique information about the daily life of a gymnasium student in the mid-19th century, the history of schools and pedagogy in Russia, a description of emancipation reform of 1861, student riots, and much more. It was especially interesting for us to analyze the evolution of topics over time. The application of topic modelling to a Russian-language diary of the 19th century is proposed for the first time. The analysis of the text was performed with the R programming language. As a method for topic modeling was chosen the probabilistic topic model of Latent Dirichlet Allocation. As the significant parts of model implementation several tasks have solved: text preprocessing, parameters of model tuning and visualization of topics. As a result, the visualization was obtained that demonstrate the topics and the changes of topics over time. |

[Изменить](http://author.nbpublish.com/author_edit_english_data.php?id=36641)

|  |  |
| --- | --- |
| **Необязательные данные:** | |
| Символика университета: | Вы можете добавить логотип и адрес университета в начало статьи. Для добавления перейдите по [ссылке](http://author.nbpublish.com/author_make_order.php?id=3&id_article=36641). |
| Научный руководитель: | не указан |
| Пояснительная записка для редактора: | отсутствует |
| Информация о грантах: | отсутствует |

[Изменить](http://author.nbpublish.com/author_edit_optional_data.php?id=36641)

|  |  |
| --- | --- |
| **Авторы** | |
| Автор: | Кузнецов Алексей Валерьевич | [Вниз](http://author.nbpublish.com/author_change_position.php?id=36641&id_author=-35835&offset=1) | [Вверх](http://author.nbpublish.com/author_change_position.php?id=36641&id_author=-35835&offset=-1) [| Добавить список публикаций](http://author.nbpublish.com/author_make_order.php?id=48&id_article=0) |

[Добавить соавтора](http://author.nbpublish.com/author_edit_coauthor.php?id=36641)

|  |  |
| --- | --- |
| **Ключевые слова** | |
| Ключевое слово: | компьютерный анализ текстов (computational text analysis) |
| Ключевое слово: | тематическое моделирование (topic modelling) |
| Ключевое слово: | латентное размещение Дирихле (Latent Dirichlet Allocation) |
| Ключевое слово: | эго-документ (ego-document) |
| Ключевое слово: | дневник (diary) |
| Ключевое слово: | источники личного происхождения (sources of private origin) |
| Ключевое слово: | обработка естественного языка (natural language processing) |
| Ключевое слово: | интеллектуальный анализ текста (text mining) |
| Ключевое слово: | история России (Russian history) |
| Ключевое слово: | личная история (personal history) |

[Изменить ключевые слова](http://author.nbpublish.com/author_edit_keywords.php?id=36641).