

Характеристики нормотворческого процесса в России - текстовый корпус

Версия для печати больше не поддерживается и может содержать ошибки обработки. Обновите закладки браузера и используйте вместо этого функцию печати браузера по умолчанию.

Характеристики нормотворческого процесса в России: текстовый корпус проектов НПА с regulation.gov.ru и sozd.duma.gov.ru

Скачать датасет можно в каталоге ИНИД по ссылке



Содержание

- 1 История изменений
- 2 Общая информация о датасете
- 3 Общее описание набора данных
- 4 Структура набора данных (Codebook)
 - 4.1 Тексты НПА на ОРВ ([regulation_texts.csv](#))
 - 4.2 Тексты законопроектов, внесенных на рассмотрение в Госдуму ([duma_texts.csv](#))
- 5 Известные ограничения данных
 - 5.1 Полнота данных
- 6 Внесенные в набор данных изменения по сравнению с источниками
- 7 Источники данных
 - 7.1 Федеральный портал проектов нормативных правовых актов (regulation.gov.ru)
 - 7.2 Система обеспечения законодательной деятельности ГАС «Законотворчество» (sozd.duma.gov.ru)
- 8 Ответственные за набор данных

История изменений

Дата	Версия	Автор	Описание изменений
07.10.2021	1.0	Василевская Мария Михайловна Зорин Семён Юрьевич	Документ создан

Общая информация о датасете

Набор данных состоит из двух частей:

1. тексты проектов нормативно-правовых актов (НПА), размещенных на портале regulation.gov.ru в рамках процедуры оценки регулирующего воздействия (ОРВ) с 2012 по 2021 г. (таблица [regulation_texts.csv](#));
2. тексты законопроектов, внесенных на рассмотрение в Госдуму с 1994 по 2021 г. (таблица [duma_texts.csv](#)).

Единица наблюдения в датасете – текст НПА в его последней редакции и внутренний идентификатор проекта. Внутренние идентификаторы проектов являются общими с соответствующими идентификаторами в другом датасете ИНИД - «Нормотворческий процесс в России: результаты оценки регулирующего воздействия и метаданные проектов НПА с regulation.gov.ru и sozd.duma.gov.ru». (описание датасета на Wiki) На основе этих идентификаторов возможно связать мета-данные проектов НПА с их текстами в последней редакции.

Таблица [regulation_texts.csv](#) содержит 11 384 наблюдения; таблица [duma_texts.csv](#) – более 21 914 наблюдений.

Все таблицы доступны в формате CSV (кодировка: «UTF-8», разделитель: «;»).

Общее описание набора данных

Атрибут	Значение
Общие сведения	
Полное наименование набора данных	Характеристики нормотворческого процесса в России: текстовый корпус проектов НПА с regulation.gov.ru и sozd.duma.gov.ru
Сокращенное наименование набора данных	Характеристики нормотворческого процесса в России за 1994-2021 гг. – текстовый корпус
Краткое описание набора данных	<p>Набор данных состоит из двух частей:</p> <ol style="list-style-type: none"> 1. тексты проектов нормативно-правовых актов (НПА), размещенных на портале regulation.gov.ru в рамках процедуры оценки регулирующего воздействия (ОРВ) с 2012 по 2021 г. (таблица regulation_texts.csv); 2. тексты законопроектов, внесенных на рассмотрение в Госдуму с 1994 по 2021 г. (таблица duma_texts.csv). <p>Единица наблюдения в датасете – текст НПА в его последней редакции и внутренний идентификатор проекта. Внутренние идентификаторы проектов являются общими с соответствующими идентификаторами в другом датасете ИНИД – «Нормотворческий процесс в России: результаты оценки регулирующего воздействия и метаданные проектов НПА с regulation.gov.ru и sozd.duma.gov.ru». (описание датасета на Wiki) На основе этих идентификаторов возможно связать мета-данные проектов НПА с их текстами в последней редакции.</p> <p>Таблица regulation_texts.csv содержит 11 451 наблюдений; таблица duma_texts.csv – более 22 тыс. наблюдений.</p>
Краткое описание источника набора данных	<p>С началом разработки нового проекта нормативно-правового акта (НПА) ведомство-разработчик обязано разместить тексты готовящихся проектов на портале Минэкономразвития regulation.gov.ru в открытом доступе.</p> <p>С внесением законопроекта на рассмотрение в Государственную Думу тексты проектов публикуются на сайте sozd.duma.gov.ru.</p>
Покрываемый временной период	Для проектов НПА на ОРВ: 2012-2021 гг.
Доступные форматы	CSV (кодировка: «UTF-8», разделитель: «;»)
Размер набора данных	953 Мбайт
Тематика набора данных	Право, Нормотворческий процесс
Режим доступа	Открытый (у Вас есть возможность работать с данными на своем личном ПК, в том числе и скачивать данные)
	Периодичность публикации и обновления
Дата размещения в каталоге	07.10.2021
Периодичность обновления	Ежегодно
Дата последнего обновления набора данных	07.10.2021
Характер последнего обновления	--
Дата следующего обновления набора данных	07.10.2022
	Дополнительные сведения
Цитирование набора данных на русском языке	<p>Характеристики нормотворческого процесса в России: текстовый корпус проектов с regulation.gov.ru и sozd.duma.gov.ru // Минэкономразвития России, Госдума России; обработка: Василевская М.М., Зорин С.Ю., Инфраструктура научно-исследовательских данных, АНО «ЦПУР», 2021. Доступ: Лицензия CC BY-SA. Размещено: 28.09.2021. URL: https://data.rcsi.science/data-catalog/datasets/178</p>
For references (English)	<p>Rule-making process in Russia: text corpora of regulations drafts from regulation.gov.ru & sozd.duma.gov.ru // Ministry of Economic Development of Russian Federation, Federal State Duma; data processing: Vasilevskaia M., Zorin S., The Research Development Infrastructure (RDI), CAG, 2021. Access: License CC BY-SA. Posted: 28.09.2021. URL: https://data.rcsi.science/data-catalog/datasets/178</p>
Ссылка на открытый репозиторий	https://github.com/CAG-ru/cag-public/tree/master/projects/ria
Геоданные (пространственная привязка нет набора данных)	
Лицензия, под которой публикуется набор данных	<p>CC BY-SA (Creative Commons)</p> <p>https://creativecommons.org/licenses/by-sa/4.0/</p> <p>Допускается использование, копирование и распространение данных в научно-исследовательских, коммерческих и любых иных целях. При преобразовании материала разрешается распространять переделанные части материала на условиях этой же лицензии — CC BY-SA</p>
Стандарт публикации	http://opendata.gosmonitor.ru/standard/3.0
Параметры запросов API	Интеграция через API не предусмотрена

Структура набора данных (Codebook)

Датасет представлен в виде двух плоских таблиц в формате CSV (кодировка: «UTF-8», разделитель: «;»):

- Таблица «Тексты проектов НПА на ОРВ» ([regulation_texts.csv](#)): 11 тыс. наблюдения за 2012-2021 гг.
- Таблица «Тексты законопроектов, внесенных на рассмотрение в Госдуму» ([duma_texts.csv](#)): 22 тыс. наблюдений за 1994 - 2021 гг.

Тексты НПА на ОРВ ([regulation_texts.csv](#))

Атрибут	Описание	Пояснение	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
regulation_project_id	Идентификатор проекта на regulation.gov.ru	Соответствует идентификаторам проектов в таблицах regulation_projects.csv , regulaton_duma_matches.csv датасета «Нормотворческий процесс в России: результаты оценки регулирующего воздействия и метаданные проектов НПА с regulation.gov.ru и sozd.duma.gov.ru»	0			string
text	Полный текст проекта НПА		0			string

Тексты законопроектов, внесенных на рассмотрение в Госдуму ([duma_texts.csv](#))

Атрибут	Описание	Пояснение	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
duma_project_id	Идентификатор проекта на regulation.gov.ru	Соответствует идентификаторам проектов в таблицах duma_projects.csv , regulaton_duma_matches.csv датасета «Нормотворческий процесс в России: результаты оценки регулирующего воздействия и метаданные проектов НПА с regulation.gov.ru и sozd.duma.gov.ru»	0			string
text	Полный текст проекта		0			string

Известные ограничения данных

Полнота данных

Набор содержит все тексты проектов, размещённых на regulation.gov.ru и sozd.duma.gov.ru. В случае regulation.gov.ru это касается всех тех проектов, которые достигли стадии «Размещение проекта акта».

Внесенные в набор данных изменения по сравнению с источниками

- Тексты проектов НПА на ОРВ и тексты законопроектов, внесённых в Госдуму Правительством, размещаются в открытом доступе на сайтах regulation.gov.ru и sozd.duma.gov.ru, соответственно. Как правило, они приводятся в виде отсканированного документа в формате pdf, реже - в форматах doc, rtf, zip.
- Для того, чтобы собрать тексты документов, был разработан скрипт, осуществляющий постраничный обход порталов, а затем документы были приведены в текстовый формат с использованием распознавания образов, реализованного в свободно распространяемой библиотеке [textract](http://textract.com). В процессе распознавания ввиду ошибок форматов и/или некорректного размещения на порталах были утеряны 212 текстов законопроектов и 67 текстов проектов на ОРВ.

Источники данных

Федеральный портал проектов нормативных правовых актов (regulation.gov.ru)

Атрибут	Значение
Полное наименование источника данных	Федеральный портал проектов нормативных правовых актов regulation.gov.ru
Сокращенное наименование источника данных	regulation.gov.ru
Владелец источника данных	Министерство экономического развития РФ
Краткое описание источника данных	Интернет-портал, содержащий документацию и данные о нормативно-правовых актах
Ссылка на источник данных	https://regulation.gov.ru/
Перечень используемых классификаторов	нет
Понятия, используемые в источнике	<p>Нормативный правовой акт (НПА) – это правовой акт, который содержит правовые нормы, общие установления, рассчитанные, как правило, на длительное применение и распространяющиеся на неопределенный круг лиц. Нормативный акт, являясь предписанием общего значения, распространяет свое действие не на одно конкретное отношение, а на тот или иной вид общественных отношений.</p> <p>Оценка регулирующего воздействия (ОРВ) – специальная процедура, сопровождающая разработку НПА, регулирующей экономические отношения, в рамках которой дается четкое определение проблемы, которую предполагается устранить регуляторными методами, рассматриваются альтернативные способы ее решения, оценивается экономическая целесообразность и обоснованность вводимых в законодательство изменений, и проводятся общественные обсуждения. В ряде случаев, результатом проведения ОРВ может стать отказ от введения нового регулирования или от изменения действующего, поскольку наиболее оптимальным способом может быть признано решение предоставить рыночным механизмам скомпенсировать проблему самостоятельно.</p> <p>Согласно п. 3 Правил раскрытия ФОИВами информации, утвержденных Постановлением Правительства РФ от 25 августа 2012 г. N 851, при разработке проектов нормативно-правовых актов федерального значения ведомства-разработчики обязаны размещать тексты НПА и ряд дополнительной информации по ним на Федеральном портале проектов нормативно-правовых актов, regulation.gov.ru.</p>
Методология и изменения методологии для источника данных	<p>На текущий момент ОРВ проводится в отношении следующих видов НПА:</p> <ul style="list-style-type: none"> • законопроекты (некоторое время портал regulation.gov.ru отдельно учитывал проекты новых федеральных законов и проекты поправок к уже существующим); • проекты указов Президента; • проекты постановлений Правительства; • проекты ведомственных актов; • проекты решений ЕЭК. <p>Необходимость публикации проектов НПА закреплена в п. 3 Правил раскрытия ФОИВами информации, утвержденных Постановлением Правительства РФ от 25 августа 2012 г. N 851</p>
Ссылки на методологию	Спектр НПА, проекты которых должны проходить ОРВ, описан в п. 60.1 Регламента Правительства РФ и в ст. 1 от 31 июля 2020 г. N 247-ФЗ «Об обязательных требованиях». В этих же документах, а также в Постановлении Правительства от 17 декабря 2012 г. № 1318 «О порядке проведения федеральными органами исполнительной власти оценки регулирующего воздействия», предусмотрен список исключений.

Система обеспечения законодательной деятельности ГАС «Законотворчество» (sozd.duma.gov.ru)

Атрибут	Значение
Полное наименование источника данных	Система обеспечения законодательной деятельности Государственной автоматизированной системы «Законотворчество»
Сокращенное наименование источника данных	СОЗД ГАС «Законотворчество»
Владелец источника данных	Государственная Дума Федерального Собрания Российской Федерации
Краткое описание источника данных	Реестр электронных карт законопроектов, рассматриваемых Государственной Думой
Ссылка на источник данных	https://sozd.duma.gov.ru/
Перечень используемых классификаторов	нет
Понятия, используемые в источнике	Подготовленный к внесению в Государственную Думу законопроект и материалы к нему, предусмотренные статьей 105 настоящего Регламента, направляются субъектом (субъектами) права законодательной инициативы на имя Председателя Государственной Думы. Поступивший законопроект регистрируется в Управлении документационного обеспечения Аппарата Государственной Думы в Системе автоматизированного делопроизводства и документооборота Государственной Думы.
Методология и изменения методологии для источника данных	Ему присваивается регистрационный номер, который указывается вместе с наименованием законопроекта в течение всего периода прохождения законопроекта в Государственной Думе. Одновременно в Системе обеспечения законодательной деятельности на законопроект заводится электронная регистрационная карта, в которой фиксируются дата и время поступления законопроекта в Государственную Думу, состав субъектов права законодательной инициативы, внесших законопроект в Государственную Думу, сведения о прохождении законопроекта в Государственной Думе, об одобрении Государственной Думой федерального конституционного закона или о принятии федерального закона и о рассмотрении соответствующего закона Советом Федерации и Президентом Российской Федерации
Ссылки на методологию	П. 1 ст. 107 Регламента Государственной Думы

Ответственные за набор данных

№	ФИО ответственного лица	Электронная почта ответственного лица
1.	Зорин Семён Юрьевич	s.zorin@cpur.ru
2.	Василевская Мария Михайловна	m.vasilevskaia@cpur.ru

Источник —

https://wiki-data.rcsi.science/index.php?title=Характеристики_нормотворческого_процесса_в_Рос

сии_-_текстовый_корпус&oldid=977