

Про Диссернет

Павел Айткулов

ajtkulov@gmail.com, Orb Intelligence / Dun and Bradstreet

28 Sep 2022

- data-, ML-engineer
- scala, fp
- спортивное ЧГК, Puzzle, top-250
- шахматы, $Elo = 2050 \dots 2100$, $Elo_{max} = 2118$
- бег, марафон: 3h 57m (2019), $\frac{1}{2}$ марафон: 1h 42m, 10k: 44m 30s (2021)
- русский бильярд

Диссернет <https://www.dissernet.org/>

Что, Где, Когда, Начало

<https://www.dissernet.org/expertise/andriyanovav2011.htm>

Сообщество, но не формальная организация.

Диссернет <https://www.dissernet.org/>

Что, Где, Когда, Начало

<https://www.dissernet.org/expertise/andriyanovav2011.htm>

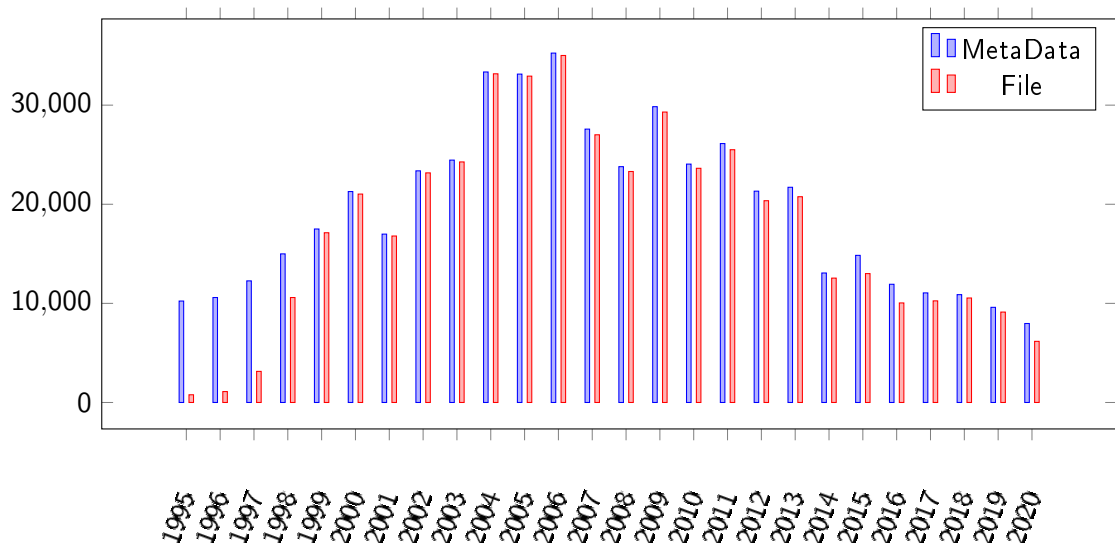
Сообщество, но не формальная организация.

Атаки/обвинения, не политика. Quis custodiet ipsos custodes?

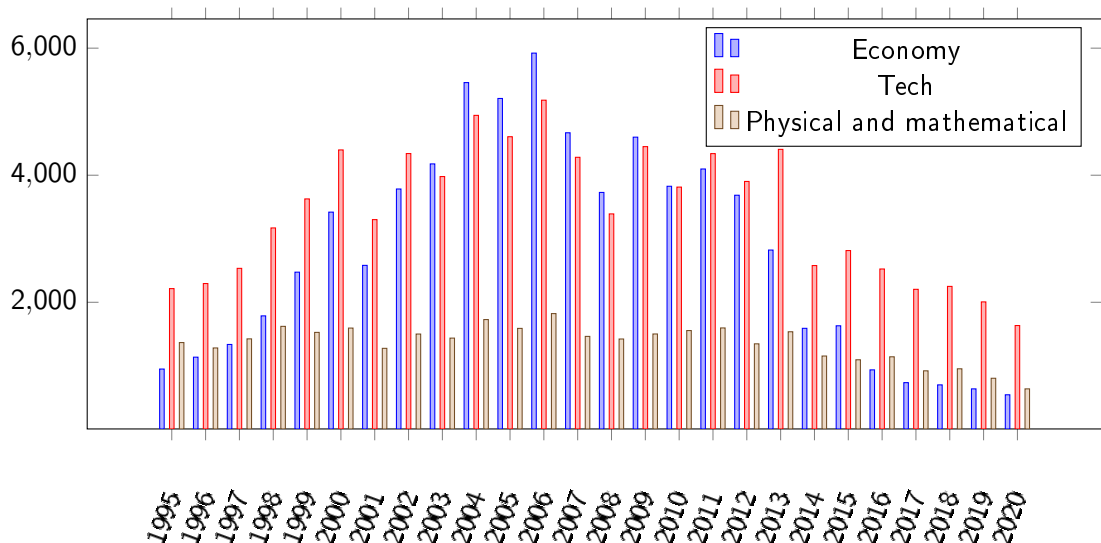
Ошибки, извинения.

https://www.dissernet.org/publications/poselyugina_errata.htm

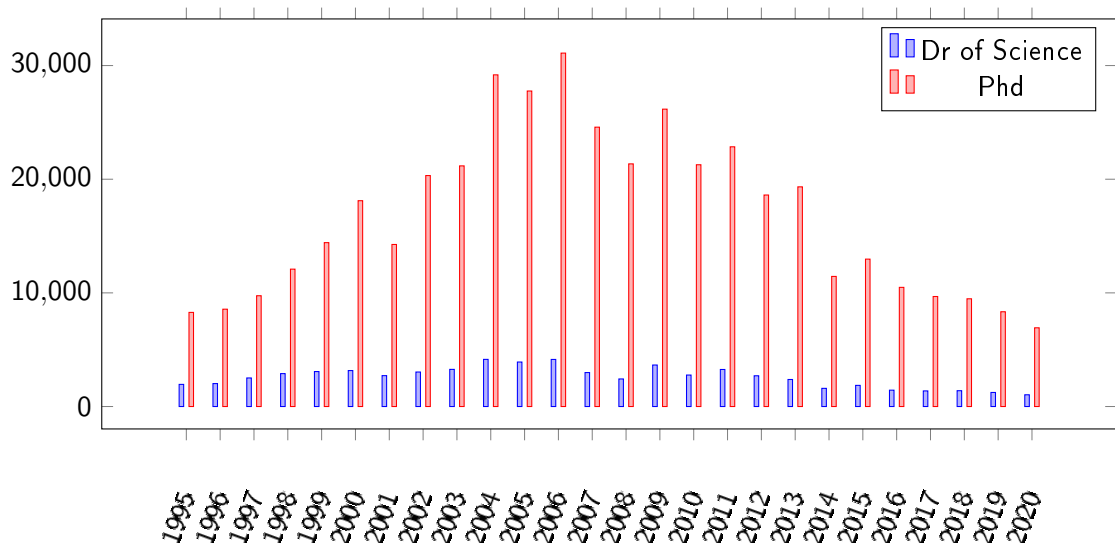
Количество метаданных и файлов диссертаций



Экономисты против технарей



Кандидаты против Докторов



- карбункул, нарезка. <https://www.dissernet.org/expertise/danilinalyu2012.htm>,
https://www.dissernet.org/expertise/ramazanovakk2011_.htm
- ЗОЛУС, новояз (зоДлус)
- ВАК, МинОбр, Диссовет
- фантом
- иностранцы, «таджики». Без коннотаций национализма или расизма
- фабрики, Машина времени, 10 лет для срока давности и после 2010 года

- текстовый
- численный, табличный
- иной

Копирование текста или чисел, но не идей.

Очевидно доказывать, за исключением терВера, «Война и мир» с заменой 1 буквы.

«Барак Обама 18 сентября посетил Токио с рабочим визитом.»

«Президент США прибудет в столицу Японии в начале следующей недели.»

- текстовый
- численный, табличный
- иной

Копирование текста или чисел, но не идей.

Очевидно доказывать, за исключением терВера, «Война и мир» с заменой 1 буквы.

«Барак Обама 18 сентября посетил Токио с рабочим визитом.»

«Президент США прибудет в столицу Японии в начале следующей недели.»

Зачем злодеи это делают? Списывание, торговля, криминал.

Научные статьи, монографии

- Множественные публикации
- Самоплагиат
- К-кругов ада
- Перевод с других языков
- П(Б)латные соавторы. Продажа места в соавторах.

Не только диссертации

Научные статьи, монографии

- Множественные публикации
- Самоплагиат
- К-кругов ада
- Перевод с других языков
- П(Б)латные соавторы. Продажа места в соавторах.

Даже в отписке на ЗОЛУС.

Патенты, судебные решения, судебные экспертизы.

Диссеропедия журналов и вузов.

Ретрагирование статьи. Хищнические журналы.

Структура защиты, научрук, оппоненты, ведущая организация. Статьи в ВАК.

Кандидатская и докторская (совместные работы с учениками).

ЗОЛУС, «родной» vs. «не родной» диссовет.

Диссовет, ВАК, МинОбр (в редких случаях, на отрицательное решение Диссовета не смотрят).

VIP, Экспертный совет ВАК/МинОбр.

Фантомы.

Ректоры/депутаты/остальные. Все слои населения. Структура по остальным специальностям (экономисты, педагоги, юристы, медики, технари)

11.5K сильное подозрение, 3K экспертиз, 1000+ ЗОЛУСов.

250 (из 270) ЗОЛУСов пройдено в 2021 году.

Основной ДатаСет (2М в метаданных, 1М в наличии, 600К диссертаций, 400К авторефератов, 330 Гб текста). О чистоте. Временной Лаг по оцифровке

- E-library(мало в публичном доступе, 14М)
- cyberleninka (2.6М статей, 200Гб текста)
- Web

Личности: Википедия, Декларатор

Аналоги: Антиплагиат (ушли в бизнес), Врониплаг (диссертации, Германия).

Версия 1. длинные «предложения». Хеширование, 64 бита. 500М, 10%

Версия 1. длинные «предложения». Хеширование, 64 бита. 500М, 10%

Версия 2. n-gram. 500Гб хешей. $n = 6$, размер не зависит от n . 20B n-gram.

Отсекаем частотные n-gram, «На правах рукописи, работа выполнена в» ($freq \leq 5$).

Штрафуется серьезнее.

<https://www.facebook.com/andrei.rostovtsev/posts/4305629239498604>

Выводы. Обычно такие главы начинаются с новой страницы. Отсечь библиографию.

Табличный плагиат. После OCR есть разбивка на страницы.

Оставляем только нетривиальные числа. 3-gram, хеширование.

Страница или соседняя страница содержит подстроку «ТАБЛИЦ».

<http://wiki.dissernet.org/wsave/LiYuPN2007.html>

<https://www.dissernet.org/expertise/melnikovaen2011.htm>

Медики.

Бежим по графу текстового копирования. Парно сравниваем таблицы с учетом редакционного расстояния. 2-gram, отдельно для таблиц с \pm .

Сейчас найдено порядка 20 случаев.

¹заменяет последние цифры в результатах исследования везде на единицу

²2-я табл. Величины и ошибки изменяются на 0.1 (большие величины - на 0.2 и 0.3)

³Добавление "0.1" к величинам измерений, погрешности оставлены без изменений.

⁴Все погрешности увеличены/уменьшены ровно на 0.2. Величины измерений в 1-й и 4-й строке изменены ровно на 0.12. Во 2-й и 5-й строке - на 0.09 В 3-й и 6-й строке - 1.2

Бежим по графу текстового копирования. Попарно сравниваем таблицы с учетом редакционного расстояния. 2-gram, отдельно для таблиц с \pm .

Сейчас найдено порядка 20 случаев.

<http://wiki.dissernet.org/wsave/BulatovRN2018.html> ¹

<http://wiki.dissernet.org/wsave/DmitrenkoGD2014.html> ²

<http://wiki.dissernet.org/wsave/ParamonovYuO2019.html> ³

<http://wiki.dissernet.org/wsave/BalikoevaFR2013.html> ⁴

¹заменяет последние цифры в результатах исследования везде на единицу

²2-я табл. Величины и ошибки изменяются на 0.1 (большие величины - на 0.2 и 0.3)

³Добавление "0.1" к величинам измерений, погрешности оставлены без изменений.

⁴Все погрешности увеличены/уменьшены ровно на 0.2. Величины измерений в 1-й и 4-й строке изменены ровно на 0.12. Во 2-й и 5-й строке - на 0.09 В 3-й и 6-й строке - 1.2

Медицинская мафия.

<https://ru.wikipedia.org/wiki/Бингсти>

Версия 1. Фиксированный список «фуфломицинов».

----- [Термин1] ----- <-----> ----- [Термин2] -----

Замена или добавление медицинского термина. «Щелкает» текстовым плагиатом в левом и правом контекстах.

Версия 1. Фиксированный список «фуфломицинов».

----- [Термин1] ----- <-----> ----- [Термин2] -----

Замена или добавление медицинского термина. «Щелкает» текстовым плагиатом в левом и правом контекстах.

Есть граф копирований. Рассматриваем только диссертации биологических, медицинских и фармацевтических наук.

Найдено порядка 100-150 работ.

Версия 2. ML для определения слова-фармакологии.

Логистическая регрессия. Общий список слов vs. аптечный каталог. (ROC AUC=0.97)

Wikipedia - 400K имен, известные и популярные, вся история, пересечения по ФИО, 50% попаданий.

Декларатор - 300K имен, кто обязан сдавать декларации о доходах, в том числе работник районной налоговой инспекции.

Спортсмены, политики, Контр-адмирал, генералы ФСБ, учителя английского, врачи, губернаторы.

Wikipedia - 400К имен, известные и популярные, вся история, пересечения по ФИО, 50% попаданий.

Декларатор - 300К имен, кто обязан сдавать декларации о доходах, в том числе работник районной налоговой инспекции.

Спортсмены, политики, Контр-адмирал, генералы ФСБ, учителя английского, врачи, губернаторы.

<https://www.dissernet.org/expertise/kozlovsv2006.htm>

<https://ru.wikipedia.org/wiki/Контр-адмирал>

https://ru.wikipedia.org/wiki/Козлов,_Сергей_Викторович

Прогоны по кандидатам в ГД РФ.

VIP, quiz

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических

VIP, quiz

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических
- все с одного совета

VIP, quiz

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических
- все с одного совета
- 2010, 69 лет (тот же самый человек), докторская, экономических

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических
- все с одного совета
- 2010, 69 лет (тот же самый человек), докторская, экономических
- в 2021 «научрук» (формально, связь сложнее) стал долларовым миллиардером (\$1.5 млрд) <https://www.svoboda.org/a/31190795.html>

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических
- все с одного совета
- 2010, 69 лет (тот же самый человек), докторская, экономических
- в 2021 «научрук» (формально, связь сложнее) стал долларовым миллиардером (\$1.5 млрд) <https://www.svoboda.org/a/31190795.html>
- все «питерские»

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических
- все с одного совета
- 2010, 69 лет (тот же самый человек), докторская, экономических
- в 2021 «научрук» (формально, связь сложнее) стал долларовым миллиардером (\$1.5 млрд) <https://www.svoboda.org/a/31190795.html>
- все «питерские»

Правильные ответы:

https://ru.wikipedia.org/wiki/Литвиненко,_Владимир_Стефанович

Плагата пока не найдено. На таком уровня вряд ли он есть.

- 1997, 45 лет, кандидатская, экономических
- 1998, 38 лет, кандидатская, экономических
- 2000, 59 лет, кандидатская, экономических
- все с одного совета
- 2010, 69 лет (тот же самый человек), докторская, экономических
- в 2021 «научрук» (формально, связь сложнее) стал долларовым миллиардером (\$1.5 млрд) <https://www.svoboda.org/a/31190795.html>
- все «питерские»

Правильные ответы:

https://ru.wikipedia.org/wiki/Литвиненко,_Владимир_Стефанович

Плагииата пока не найдено. На таком уровня вряд ли он есть.

Либо «дареному коню» ... примеры и контр-примеры.

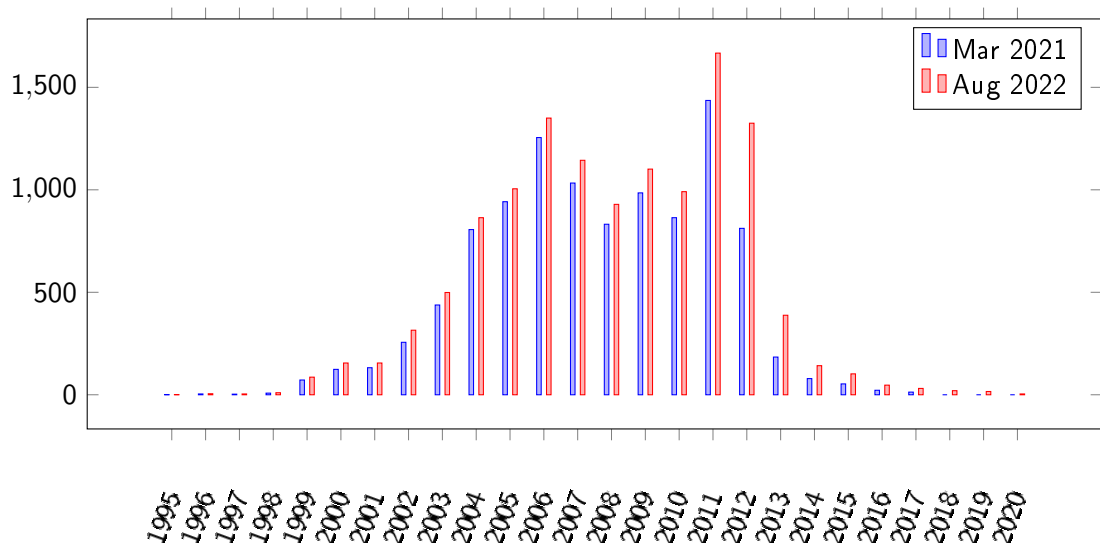
Обновление источников <http://wiki.dissernet.org/wsave/NikogdaYuV2013.html>

Игошенизация («шоколад» → «мясо»)

<https://www.dissernet.org/expertise/igoshinin2004.htm>

https://ru.wikipedia.org/wiki/Игошин,_Игорь_Николаевич

Cases



https://www.facebook.com/serguei.parkhomenko/posts/10225243601346536?comment_id=1022

<https://www.facebook.com/andrei.rostovtsev/posts/4347241862004008>

<https://www.facebook.com/andrei.rostovtsev/posts/4342907015770826>

<https://novayagazeta.ru/articles/2016/11/02/70390-biblioteka-s-privideniyami>

<https://novayagazeta.ru/articles/2021/04/22/dissertatsii-oborotni-v-biblioteke-s-privideniyami>

https://www.dissernet.org/expertise/bogunovss2012_.htm

<https://www.dissernet.org/expertise/229/rozhnikovskijar2007.htm>

<https://www.facebook.com/andrei.rostovtsev/posts/pfbid029QK8S9...>

<https://www.dissernet.org/expertise/murahovskyag2013.htm>

Hijacked journals

Анна Абалкина: Мошеннические сайты, которые используют название, ISSN и другие метаданные оригинальных журналов для обмана потенциальных авторов.

<https://trv-science.ru/2021/06/ostorozhno-zhurnaly-klony/>

https://www.mv.helsinki.fi/home/kopotev/Kutuzov_etc_clustering-comparable-corpora_2016.pdf

https://www.dissernet.org/publications/svoboda_perevodnoy_plagiat.htm

http://wiki.dissernet.org/wsave/IJICC_2019_10_1publ.html

<http://rosvuz.dissernet.org/person/111797>

https://ru.wikipedia.org/wiki/Комиссия_по_борьбе_с_лженаукой

https://ru.wikipedia.org/wiki/Комиссия_по_противодействию_фальсификации_научных_исследований

Увы, у академии мало влияния на государство и процессы. Все в обратную сторону.

Нет доступа до монографий, статей и диссертаций. Текущий датасет - удача, что в пандемию РГБ ненадолго «ошибочно выложили» в открытый доступ. Список МинОбра о существующих защитах не доступен.

Научрук после ЗОЛУСа лишается право занимать должность в Диссовете на 5 лет.

- Текстовый плагиат (n-gram), диссертации, КиберЛенинка.
- Текстовый плагиат, выводы
- Табличный плагиат, точное совпадение
- Табличный плагиат, подмена
- Обновление источников
- Фантомы, подозрения на АлтайГУ 2004-2006, аналог СПбГЭУ 2010х.
- VIP
- Фуфломицины

- Текстовый плагиат (n-gram), диссертации, КиберЛенинка.
- Текстовый плагиат, выводы
- Табличный плагиат, точное совпадение
- Табличный плагиат, подмена
- Обновление источников
- Фантомы, подозрения на АлтайГУ 2004-2006, аналог СПбГЭУ 2010х.
- VIP
- Фуфломицины

На март 2021 года было 9К (за 2013-2021 годы работы Диссернета) подозрений, на август 2022 - 11.5К. Основные фабрики уже были найдены.

2012 год покрыт, 1300+ потенциальных ЗОЛУСОВ. За 2013-2021 подано 1000+ ЗОЛУСов.

Медленно и верно покрываем все, с приоритетом на ЗОЛУСы и срок давности, для 2000-2009 малое покрытие.

<https://www.youtube.com/watch?v=vDKQQSDDj2g>, «Диссернет. Эволюция альтруизма (научно-популярный фильм)»

<https://www.youtube.com/c/DissernetTV>

<https://www.dissernet.org/publications/namedni.htm> НАМЕДНИ-2013

Что дальше?

Судебные решения?

<https://www.dissernet.org/expertise/dudarnn2004.htm>

Что дальше?

Судебные решения?

<https://www.dissernet.org/expertise/dudarnn2004.htm>

Есть датасет.

Другие страны?

Идеи?



