

TheyBuyForYou

Использование ИИ и анализа данных в государственном секторе

(CC) 2021

Матей Ковачич

Матей Посинкович



<http://tbfy.ijs.si>

Данная работа опубликована
под лицензией CC BY-NC-SA 4.0

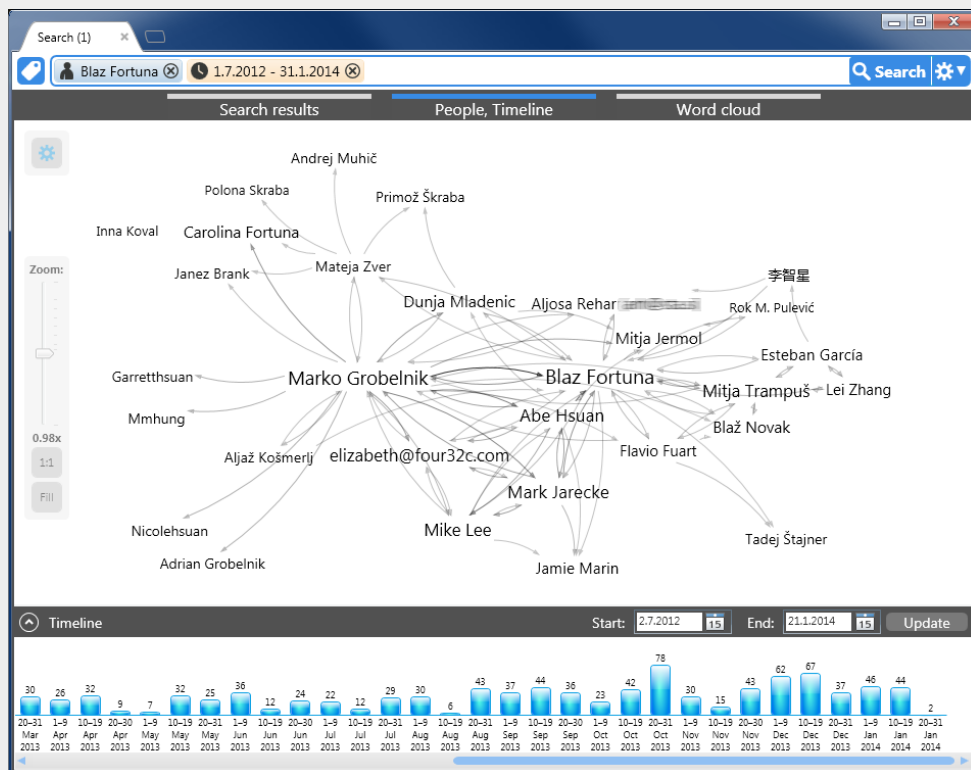
Использование ИИ в государственном секторе

“... разрабатывать более эффективную политику и принимать более обоснованные решения, улучшать общение и взаимодействие с гражданами и жителями, а также повышать скорость и качество государственных услуг.”
(источник: OECD.org).

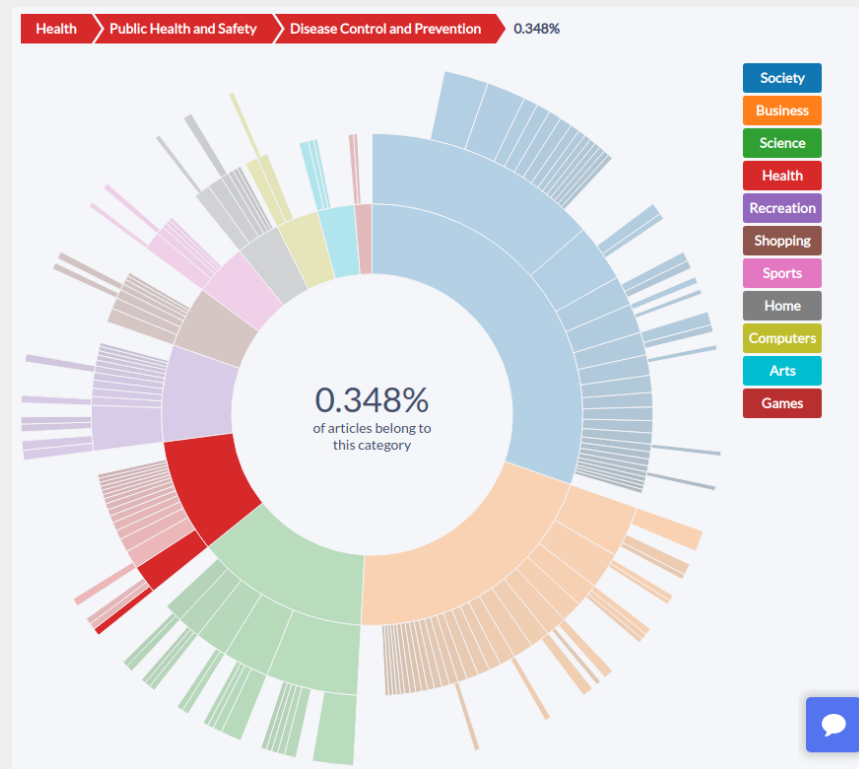
ИИ и большие данные в финансах :

- Выявление скрытых закономерностей и тенденций (повышение прозрачности).
- Прямые (специализированные) расследования.
- Стратегические решения на основе анализа данных.

Примеры анализа данных

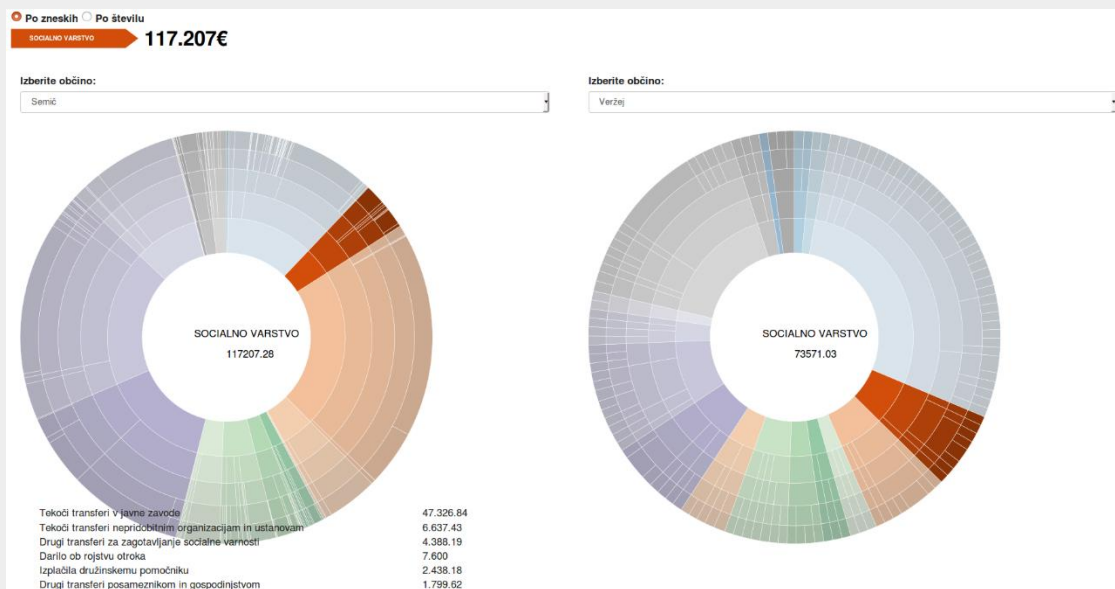


Контекстификация – инструмент для анализа содержания и метаданных электронных писем.

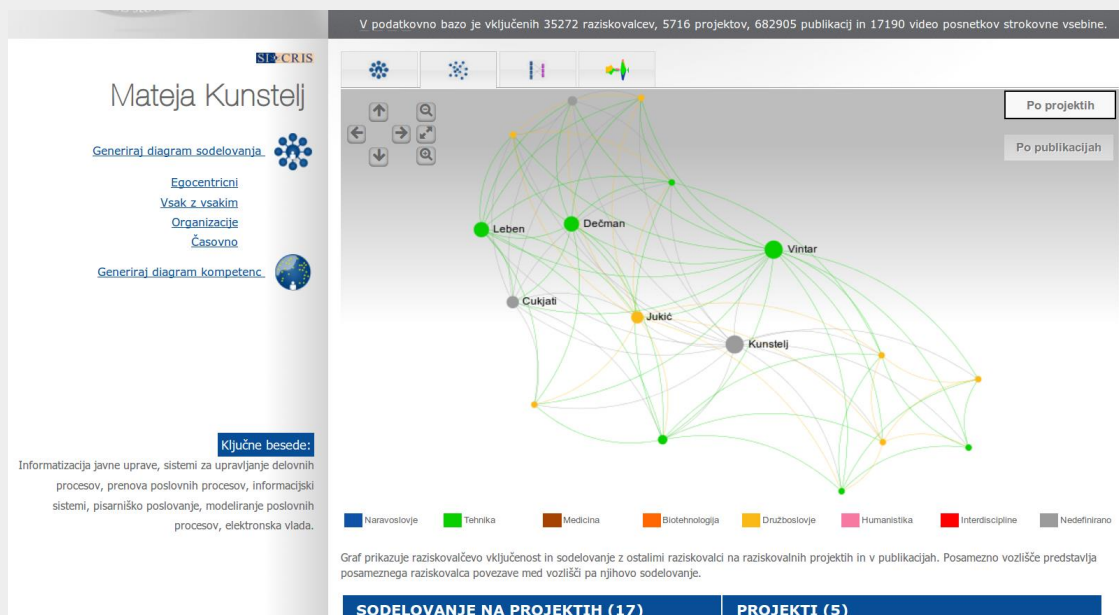


Реестр событий – инструмент для анализа содержания статей СМИ.

Примеры анализа данных

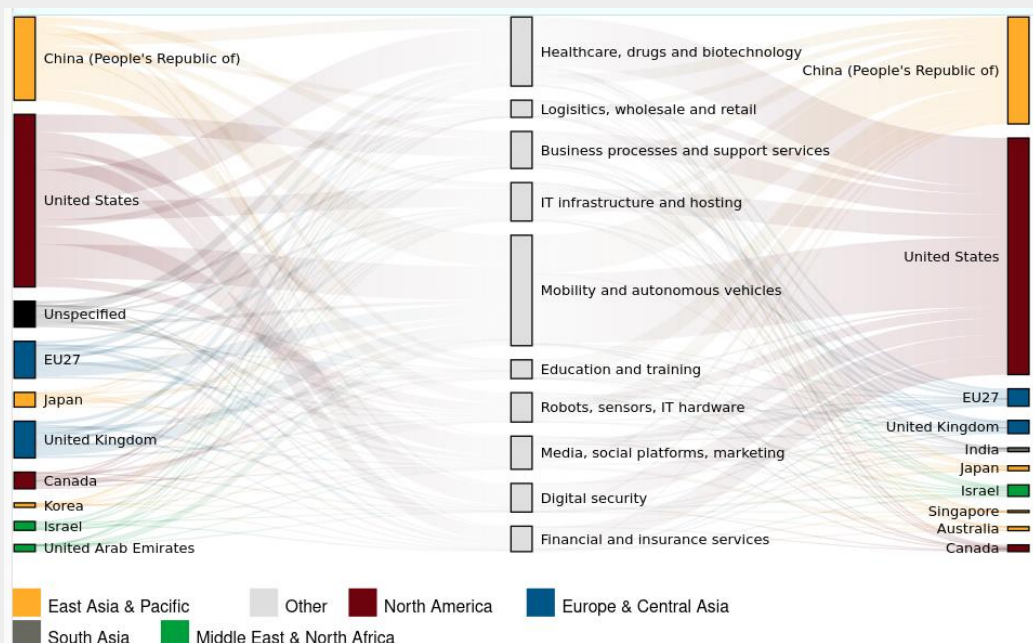


Визуализация и сравнение бюджетов самоуправлений.

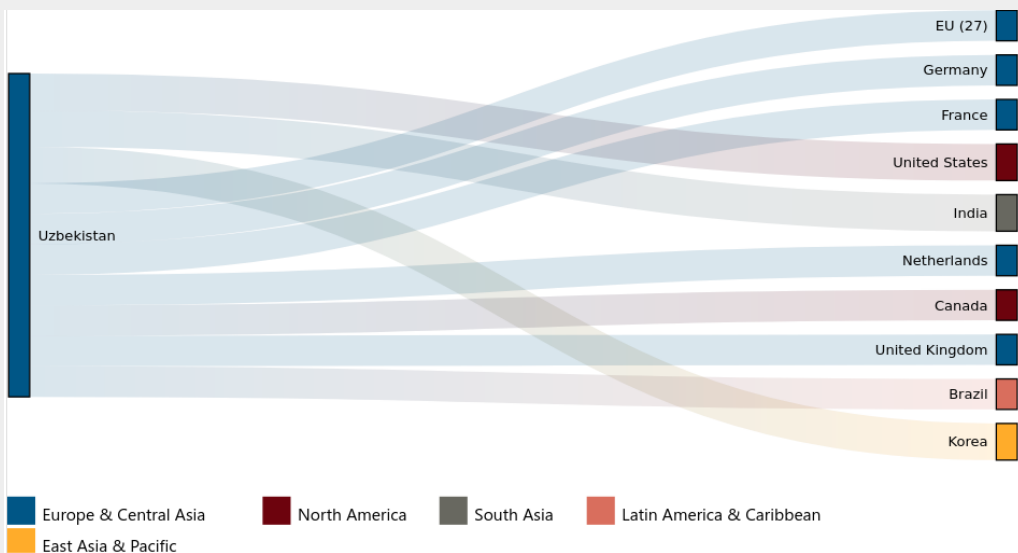


Визуализация и анализ научных публикаций и проектов, отображение графиков компетенций ученых и сотрудничества между ними.

Примеры анализа данных

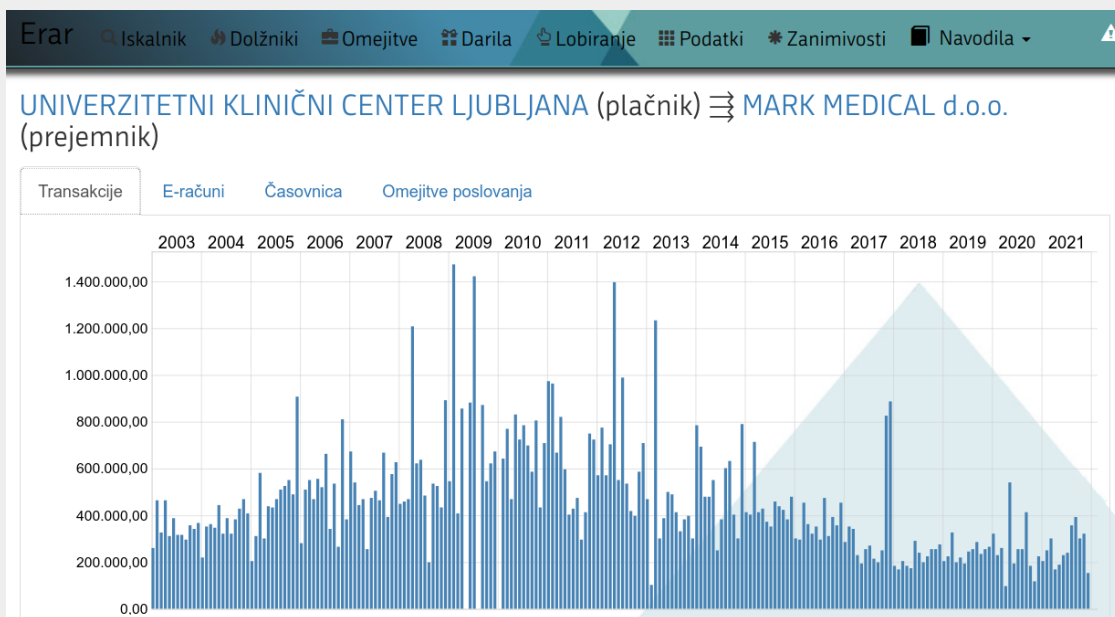


Поток венчурных инвестиций в ИИ (из страны инвестора в отрасль и страну стартапа).

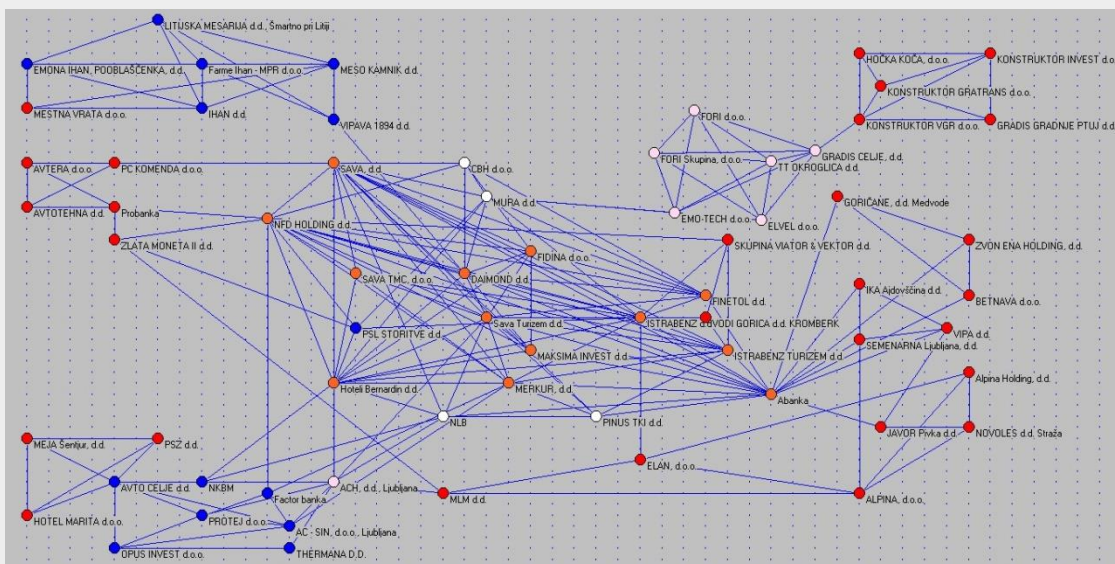


Визуализация отечественного и международного сотрудничества в разработке программного обеспечения ИИ

Примеры анализа данных



Erar.si показывает финансовые транзакции из государственного в частный сектор. Он также содержит данные о хозяйствующих субъектах в Словении, бухгалтерские записи о платежах, данные о налоговых должниках и переводах в офшоры и т. д.



Плохие вложения переданы в компанию по управлению активами банка (КУАБ). Самый большой компонент сети содержит ок. 33% всех компаний перешедших на КУАБ. Этим компаниям было предоставлено ок. 61% проблемных кредитов.

Общая информация о TBFY

Проект TheyBuyForYou стартовал в январе 2018 года и завершился в конце декабря 2020 года.

Консорциум TheyBuyForYou состоял из 10 ведущих компаний, университетов, исследовательских центров, государственных ведомств и местных органов власти в Великобритании, Норвегии, Италии, Испании и Словении.

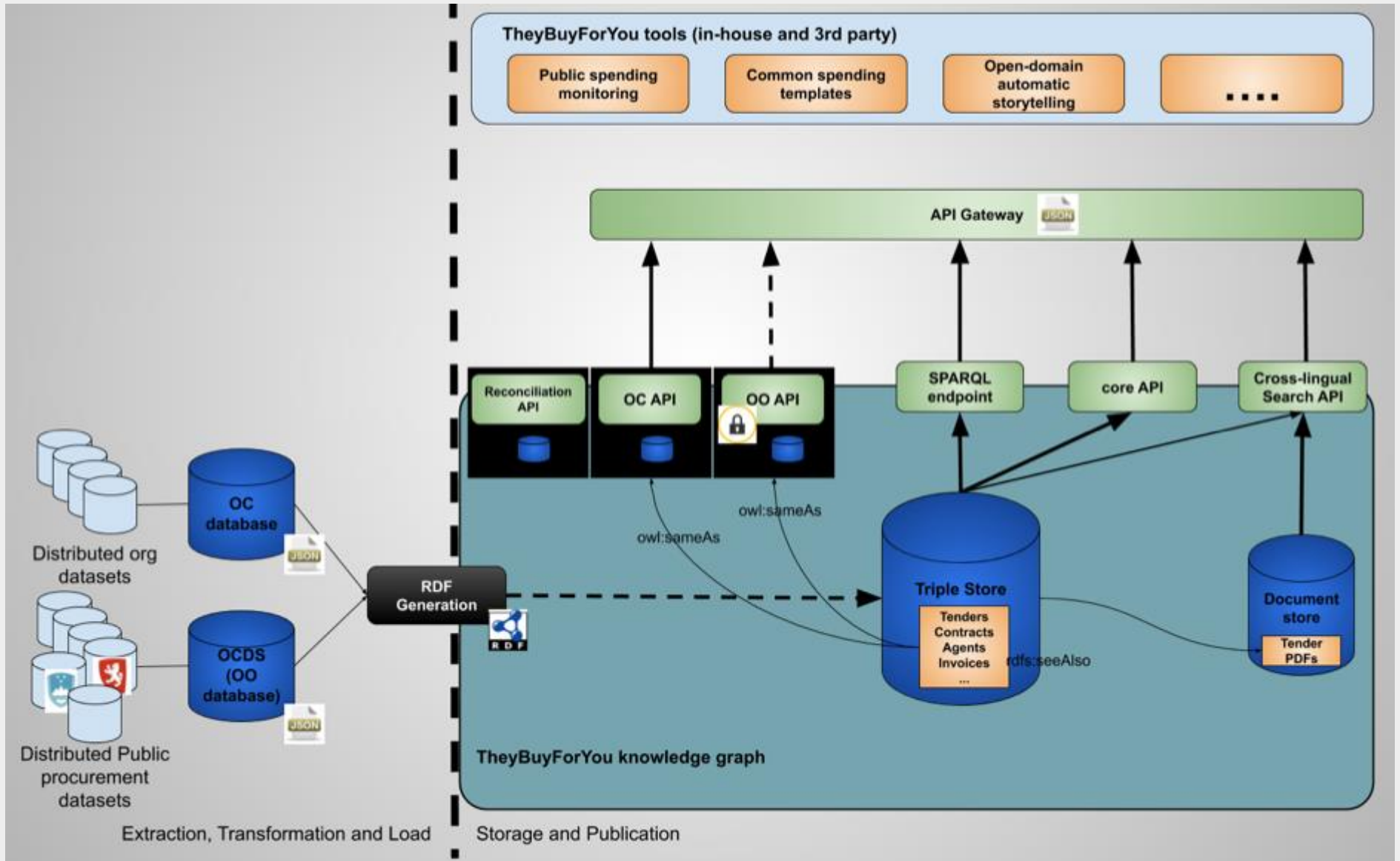
TheyBuyForYou получил финансирование от программы исследований и инноваций Европейского Союза Horizon 2020 в рамках грантового соглашения № 780247.

Цель

Целью проекта – сделать закупки в Европе более доступными. Это включало следующее:

- сбор и нормализация данных о тендерах и наградах;
- создание Европейской сети знаний по закупкам, содержащей миллионы подробных данных о контрактах и тендерах;
- предоставление данных в стандартизованном формате OCDS (Стандарт открытых контрактных данных);
- разработка нескольких инструментов для анализа и визуализации собранных данных;
- тестирование инструментов на нескольких бизнес-кейсах.

Платформа TheyBuyForYou



Платформа TheyBuyForYou

Платформа TheyBuyForYou (<https://tbfy.github.io/platform/>):

- Данные (репозиторий Графика знаний и документов TBFY).
- Схемы (Онтология TBFY, которая импортирует стандартную онтологию открытых данных контрактов для данных о закупках и онтологию euBusinessGraph для данных компании).
- Инструменты (Harvester, сервисы R4R для RESTful, конвейер приема данных Графика Знаний (KG), SPARQL GUI и OptiqueVQS).
- API – Интерфейсы прикладного программирования (набор основных ИПП, созданных или используемых в проекте).
- Услуги с добавленной стоимостью (услуги и инструменты).

График знаний [KG]

График знаний (KG) TheyBuyForYou (TBFY) объединяет данные о закупках и компании. ГЗ охватывает данные с января 2019 года. Новые данные добавляются каждую ночь.

По состоянию на май 2021 года ГЗ насчитывал более 189 миллионов троек и содержал информацию о:

- 1.92 миллионах тендеров,
- 2.51 миллионах наград, и
- 133 тысяч компаний (согласованные поставщики).

График знаний [KG]

Данные KG предоставляются в виде открытых данных в соответствии с лицензией Creative Commons BY-NC-SA 4.0 (использование, совместное пользование и адаптация данных для некоммерческих целей).

Данные (<http://data.tbfy.eu/>) доступны через:

- Конечная точка SPARQL;
- API (Интерфейс прикладного программирования);
- Хранилище данных.

Платформа обнаружения аномалий

THEY BUY FOR YOU

Transactions Tenders

Visualise and analyse public procurement data

An online toolkit exploring public spending and tender data to make public procurement more efficient and transparent.

Behind the toolkit

DATA SCIENCE
Complex analysis of transaction data and corresponding meta data.

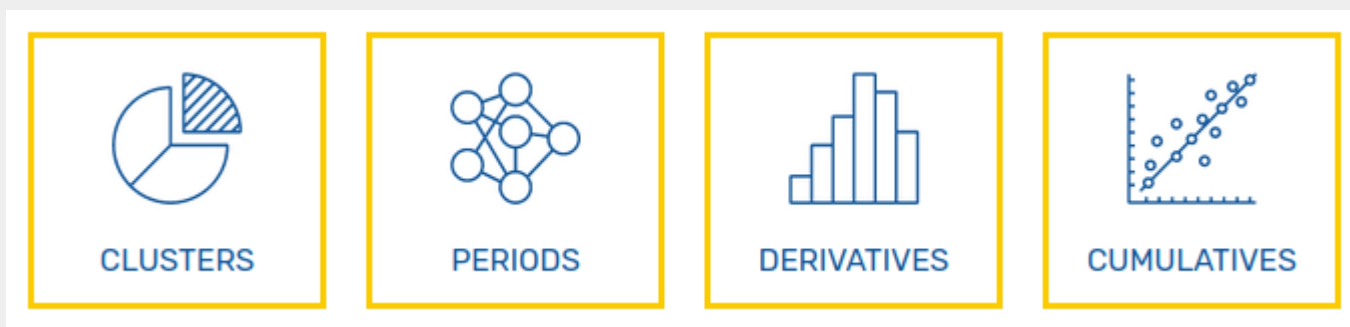
METHODOLOGY
Data models that encode and identify unusual patterns.

INSIGHTS
Anomaly detection in the flow of money and tenders.

Онлайн-инструментарий, разработанный Институтом Йожева Стефана (ИЙС) для изучения данных о государственных расходах и тендерах, чтобы сделать государственные закупки более эффективными и прозрачными. (<http://tbfy.ijs.si/>).

Платформа обнаружения аномалий

Анализ **финансовых операций** между государственными структурами и частными компаниями.



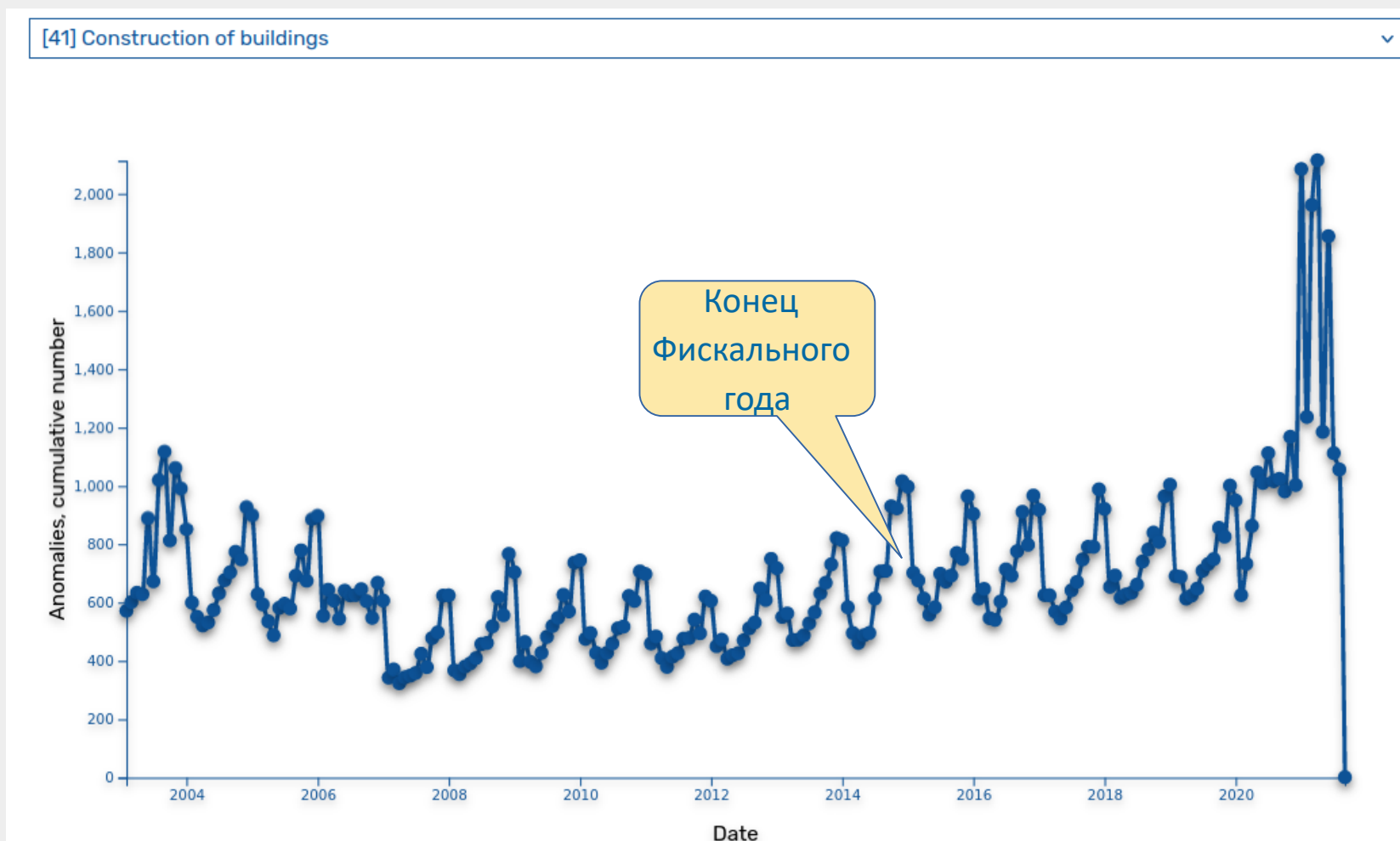
Обнаружение аномалий и их визуализация в данных о расходах могут быть выполнены несколькими методами: аномалия среднего отклонения, естественные разрывы Дженкса, совокупные баллы маржи за период, обнаружение локальных экстремумов и отклонения временных периодов. Аномалии обнаруживаются во всем наборе данных или только в различных отраслях промышленности.

Анализ финансовых транзакций



Финансовая операция определяется как базовое отношение между двумя субъектами (субъектом государственного сектора и субъектом хозяйствования). Мы определяем, когда отношения начались или закончились, и накапливаем начальные / конечные периоды на временной шкале. На основании этого мы выявляем отклонения и перечисляем сущности как часть выявленных экстремумов (метод периодов).

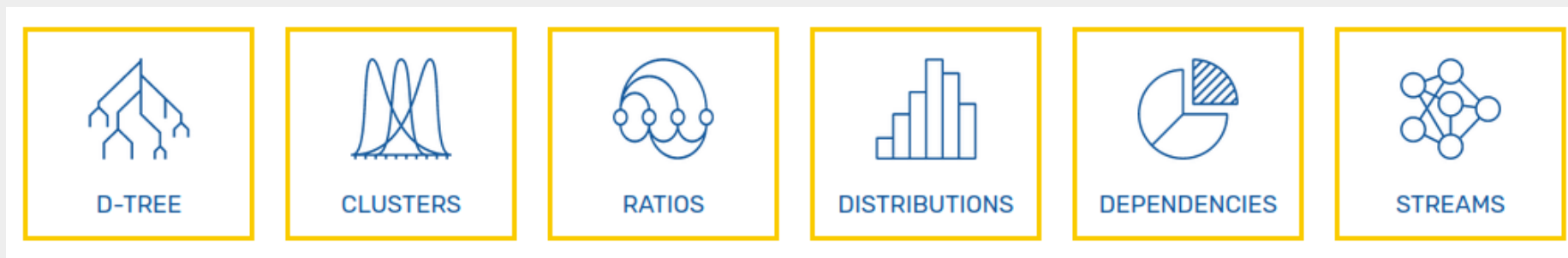
Анализ финансовых операций



Метод производных финансовых инструментов анализирует самые большие изменения в двух организациях и транзакционных отношениях за определенный период. Если изменение определяется как аномалия, оно добавляется в график совокупных аномалий. Цель метода – выявить компании, в которых произошли наибольшие изменения в транзакционных отношениях.

Платформа определения аномалий

Анализ данных о государственных закупках в формате OCDS (СОКД - Стандарт открытых контрактных данных).



Анализ данных о государственных закупках может осуществляться несколькими подходами: контролируемый, неконтролируемый и статистический анализ. Неконтролируемый анализ основан на методе k-средних. Контролируемый анализ основан на анализе дерева решений и используется для получения дополнительной информации о процессе принятия решений о государственных закупках. Статистический анализ проводится для достижения более интуитивно понятных и заранее определенных целей. Инструмент StreamStory используется для раскрытия, визуализации и объяснения внутренней структуры данных.

Анализы госзакупок

COMMON ANOMALY RANK



Common anomaly measure: higher value infers higher anomalous behaviour

Anomaly:	Company Id:	Company Name
1.95	t310	Westinghouse Electric Company LLC
1.95	1526693000	DEMA PLUS, inženiring d.o.o.
1.92	5744954000	PHARMACO Informacijski inženiring d.o.o.
1.91	1888188000	KOLEKTOR KOLING Inženiring, instalacije, proizvodnja d.o.o.

Анализ государственных закупок

PHARMACO Informacijski inženiring d.o.o. Bukovci 127 2281 Markovci Slovenija	ID 5744954000 Tax 86193210 Size 5-9 employees	Tenders won 6 Funds received 487.795,00 € Tenders issued 0 Funds issued 0,00 € Anomalies found 5	External links • Ajpes • Erar
---	---	--	---

Tenders won (6)	Tenders issued (0)	Anomalies (5)
-----------------	--------------------	---------------

Number of competitive offers distribution

PHARMACO Informacijski inženiring d.o.o. is on average winning tenders within a **LESS competitive** environment as an average company does. In comparison to other companies, PHARMACO Informacijski inženiring d.o.o. can be found [here](#) having a score of -93.9.

Current company's competition for won tenders:	Average competition for companies tender winners:	Average competition for tenders winners (CPV = 72):
No competitor 100%	No competitor 45%	No competitor 82%
One competitor 0%	One competitor 19%	One competitor 12%
2+ competitors 0%	2+ competitors 37%	2+ competitors 6%

Веб-платформа позволяет пользователям просматривать обнаруженные аномалии. Аномалии также объясняются на естественном языке. В этом случае конкуренция компании сравнивается со средней конкуренцией и средней конкуренцией в выбранном CPV (контрактном платежном поручении).

Анализ государственных закупок

Contracting authorities significantly dependent on PHARMACO Informacijski inženiring d.o.o.:

Dependence share	Contracting authority name
97.35%	LEKARNA AJDOVŠČINA
13.78%	LEKARNA SEVNICA
10.58%	Zasavske lekarne Trbovlje
6.04%	JAVNI ZDRAVSTVENI ZAVOD MARIBORSKE LEKARNE MARIBOR



LEKARNA AJDOVŠČINA	ID	5053498000	Tenders won	0	External links • Ajpes • Erar
Tovarniška cesta 3E	Tax	71037519	Funds received	0,00 €	
5270 Ajdovščina	Size	10-19 employees	Tenders issued	2	
			Funds issued	44.061,57 €	
			Anomalies found	3	

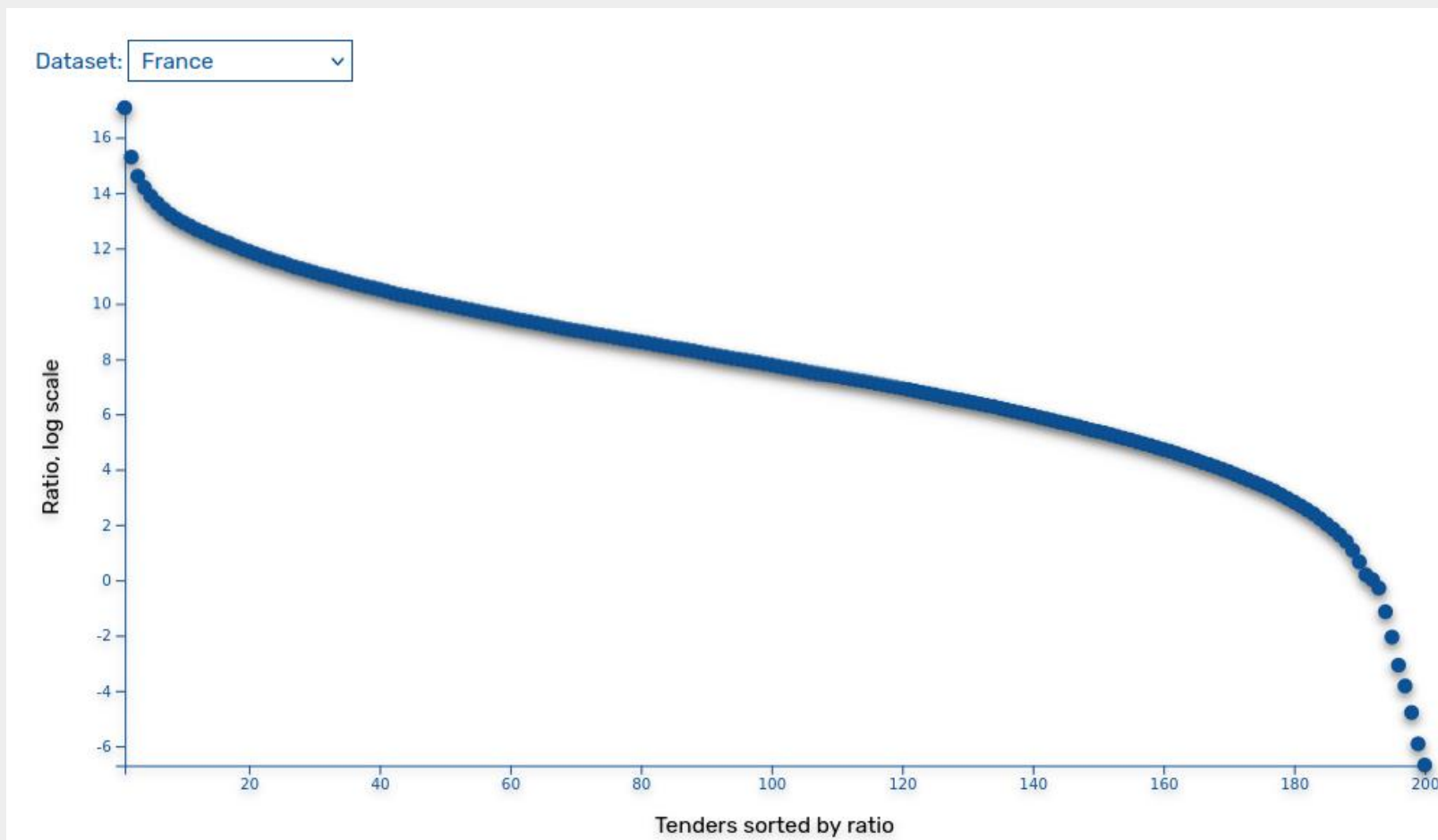
Tenders won (0) Tenders issued (2) Anomalies (3)

[Tenders by value](#) | [Tenders by num](#)



Выявлено несколько типов аномалий. Здесь мы видим зависимость закупающих органов от выбранной компании.

Методы обнаружения аномалий

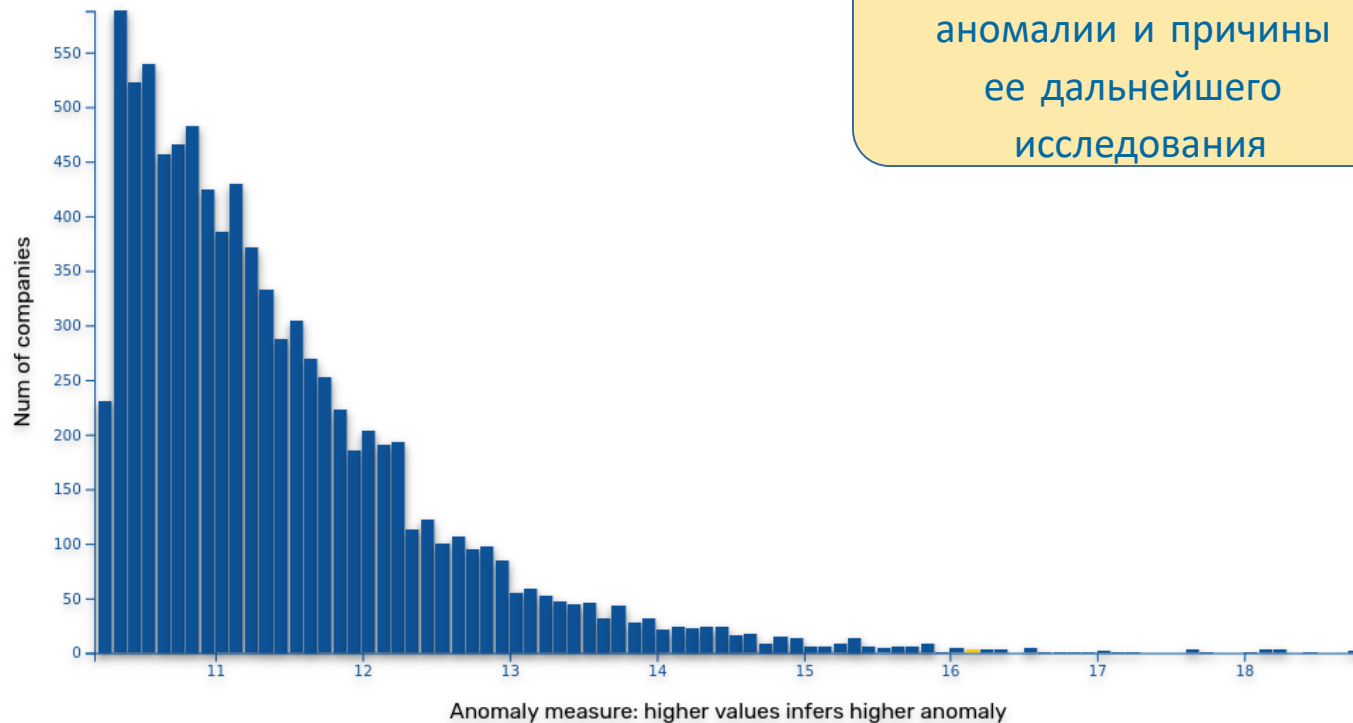


Реализованный метод статистического анализа (метод соотношений) позволяет наглядно представить взаимосвязь между стоимостью тендера и количеством сотрудников экономического оператора. Наш инструмент может анализировать любые данные в формате OCDS (СОКД).

Методы обнаружения аномалий

A high deviation indicates a high deviation between tender size and the number of economic operator employees. It is not common for large tenders to be won by companies with extremely low number of employees. While it is not necessarily in breach of the public procurement legislation, it does warrant a closer look.

Dataset: Sample size: CPV:



Объяснение обнаруженной аномалии и причины ее дальнейшего исследования

Contracting authority:	Economic operator:	Tender amount	Economic operator employees
MINISTRSTVO ZA INFRASTRUKTURU, DIREKCIJA REPUBLIKE SLOVENIJE ZA INFRASTRUKTURU [id:5300177000]	MEGALIT inženirski biro d.o.o. [id:8241767000]	10.614.858,24 €	1

Методы обнаружения аномалий



Визуализация метода кластеров. Данный метод ищет ранее необнаруженные закономерности в данных, обычно те, о которых мы не знаем.

Платформа обнаружения аномалий

Платформа доступна по ссылке: <http://tbfy.ijs.si/>.

Платформа обеспечивает ежедневный экспорт словенских данных о закупках в формате OCDS.

Данные были предоставлены благодаря Министерству государственного управления Словении.

Использует данные о закупках из Графика знаний и данные о словенских финансовых транзакциях из Egar (<https://egar.si/static-data/>).

Исходный код доступен по адресу Github: <https://github.com/TBFY/anomaly-detection-tool/>.



Вопросы ?

Матей Ковачич

Матей Посинкович



TheyBuyForYou получил финансирование в рамках программы исследований и инноваций Европейского Союза Horizon 2020 в соответствии с соглашением о гранте № 780247.

Данная работа опубликована
под лицензией CC BY-NC-SA 4.0