



Кейс

**Разработка платформы с общей
статистикой и возможностью
выгрузки данных**

Автор: Фонд Транспортные
инновации Москвы





СИТУАЦИЯ

Еженедельно аналитическим отделом ТИМ составляется дайджест с новостями и аналитикой в сфере мировых транспортных инноваций.

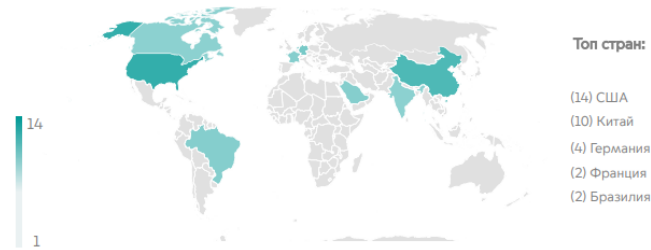
Материалы подготавливаются для руководства всего Московского Транспорта (руководителей Департамента Транспорта, руководителей и заместителей руководителей всех подведомственных организаций, сотрудников инновационных подразделений).



ОБЩИЙ ВЗГЛЯД



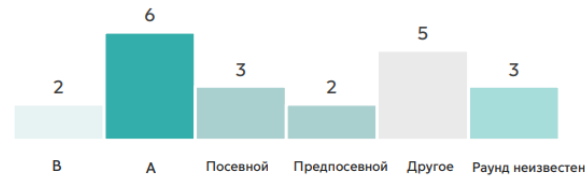
ЛОКАЦИИ КЛЮЧЕВЫХ УЧАСТНИКОВ СДЕЛОК



ТЕМАТИКИ РЕШЕНИЙ СТАРТАПОВ

	Hardware	Software	Service
Железнодорожный пассажирский		1	
Железнодорожный грузовой		2	
Автомобильный грузовой	1	3	6
Автомобильный пассажирский	5	2	1
Водный грузовой		1	
Водный пассажирский		1	
Воздушный грузовой	1		
Воздушный пассажирский	2		1

РАУНДЫ ФИНАНСИРОВАНИЯ



ПРОБЛЕМА

Сейчас при помощи «ручного отсмотра» удастся рассмотреть только стартапы, которые получили финансирование.

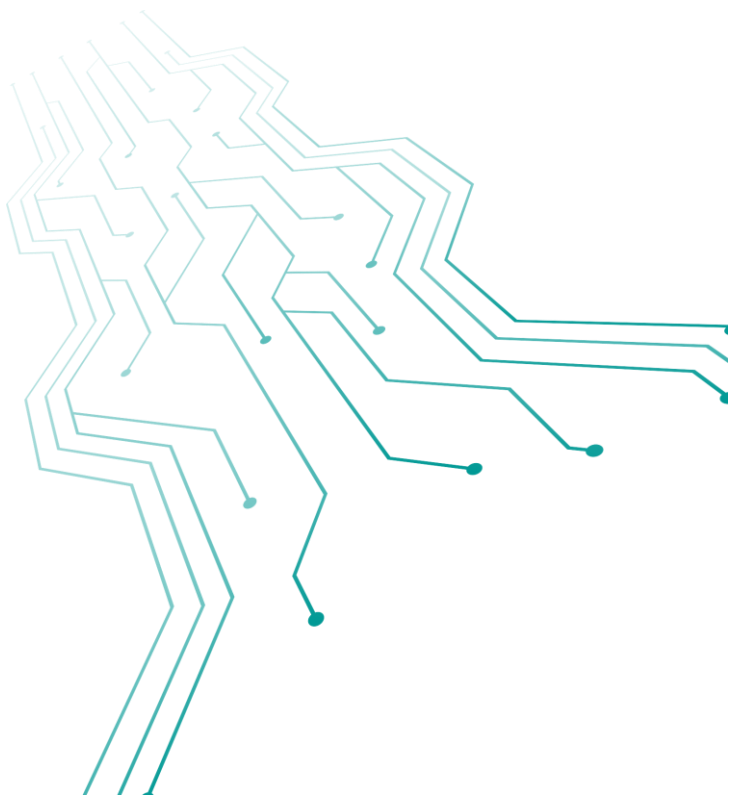
Чтение всех новостей для определения самого часто упоминаемого стартапа занимает необоснованное количество времени, хотя топ из наиболее цитируемых компаний с наглядной статистикой остается востребованной рубрикой для понимания тенденций рынка транспортных инноваций читателями дайджестов.



НАПРАВЛЕНИЕ ПОИСКА РЕШЕНИЯ

Решением проблемы может стать платформа с общей статистикой упоминания конкретных названий стартапов в сфере транспорта в новостных лентах известных изданий и возможностью выгрузки данных за определенный промежуток времени в CSV.

Источниками являются новостные издания с новостями про стартапы. Задача алгоритма — определить, что в тексте речь идет о стартапе транспортной тематики. В таблице должны выгружаться топ из упоминаемых стартапов с метаданными.



НАПРАВЛЕНИЕ ПОИСКА РЕШЕНИЯ

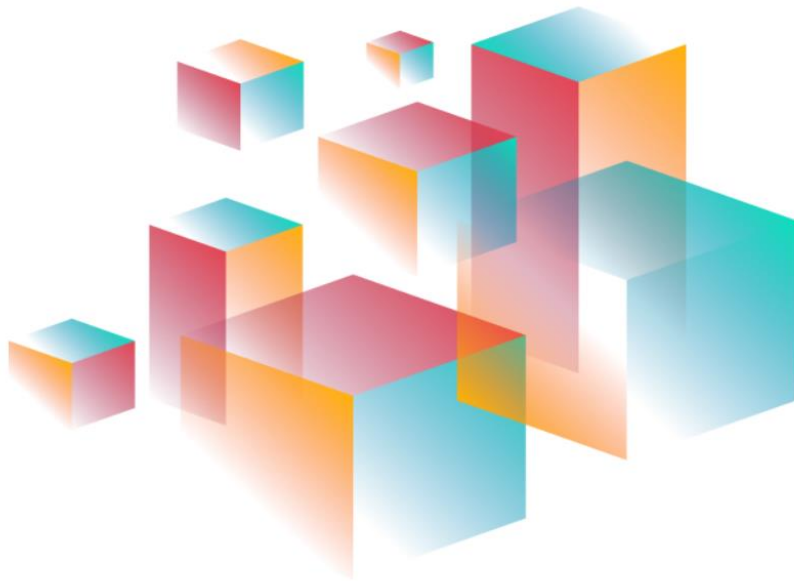
Классический RSS не подходит, так как постит все новости в одну ленту, без фильтрации по теме транспорта.

Также нет возможности сразу получить готовые метаданные упоминаемого стартапа и количество упоминаний каждого стартапа, относящегося к транспортной тематике. Сервисы поиска упоминания (по типу медиалогии) не подходят, так как ищут все упоминания «транспорта», но не стартапов в сфере транспорта в профильных изданиях.

Идентифицировать тематику статьи как транспортную можно, например, по наличию ключевых слов, таких как: «transport», «transportation», «mobility» «future transportation» и других связанных слов, объектов на картинке к новости или по определению названия стартапа (при помощи верификации со всеми существующими транспортными стартапами, если такую базу удастся найти), с помощью уже установленных фильтров в статьях.



ФУНКЦИОНАЛ ПЛАТФОРМЫ



1. Страница с общей статистикой (визуальное изображение количества упоминаний определенных стартапов)
2. Кнопка выгрузки данных в csv за определенный промежуток времени (с субботы по пятницу)

Что должно быть в выгружаемых табличных данных:
Рейтинг из 10 самых упоминаемых стартапов.



ДААННЫЕ ДЛЯ РЕШЕНИЯ



Известные издания с новостными статьями:

<https://techcrunch.com/category/startups/>

<https://www.cnbc.com/startups/>

<https://techstartups.com/category/startups/>

<https://inc42.com/buzz/>

<https://www.eu-startups.com/>

<https://www.techstars.com/newsroom>

<https://www.geekwire.com/startups/>

<https://startupnews.com.au/category/news/>

<https://pitchbook.com/news>

<https://www.cbinsights.com/newsletter/>

(рассылки новостей)

<https://magazine.startus.cc>

<https://www.startupdaily.net>