



Санкт-Петербургский
государственный
университет
www.spbu.ru

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ ДЛЯ ОЦЕНКИ ПОКАЗАТЕЛЕЙ ФУНКЦИОНИРОВАНИЯ СИСТЕМ ГОРОДСКОГО ПАССАЖИРСКОГО ТРАНСПОРТА

Графева Н.Г.
доцент кафедры
информационно-
аналитических систем СПбГУ



- 2016 год
- Проект по анализу пассажиропотоков общественного транспорта Санкт-Петербурга
- Заказчик - Организатор перевозок Санкт-Петербурга
- Исполнители: Институт транспортных систем Санкт-Петербурга и Санкт-Петербургский Государственный Университет



- На основании данных системы мониторинга подвижного состава ГПТ и систем электронных платежей сформировать репрезентативный данные высокой точности: о поездках и корреспонденциях, затратах времени пассажиров, о пересадочности, о пассажиропотоках и т.п.



- Данные об оплате на наземном общественном транспорте (система СЭКОП)
- Данные о движении наземного общественного транспорта (система АСУГПТ)
- Данные об оплате метро (АСКОП-М)
- Справочная информация о транспортной системе Санкт-Петербурга (остановки, виды транспорта, маршруты и т.п.)
- Период накопления информации - 2 недели



ОБЪЕМЫ ИСХОДНЫХ ДАННЫХ

- СЭКОП - более 29 000 000 транзакций
- АСКОП-М - около 26 000 000 транзакций
- АСУГПТ - более 14 000 000 треков
- Остановки - более 7 000
- Маршруты - более 800
- Количество электронных проездных документов - около 2 500 000



ПРОБЛЕМЫ С ЗАГРУЗКОЙ ДАННЫХ

- Не удалось получить прямой доступ к исходным данным в системах СЭКОП, АСКОП-М и АСУ ГПТ
- Данные выгружались из информационных систем в промежуточные файлы, а затем загружались в целевую базу (общий объем загружаемых данных - около 10ГБ)
- Следствие такой схемы загрузки данных - несогласованность форматов выгрузки, нечитаемые символы и т.п.
- 90% времени проекта ушло на «очистку» и получение адекватных данных
- Полностью решить проблему сопоставления данных из разных источников не удалось по причине рассогласованности систем.



- СУБД ORACLE 11 (Standard Edition)
- Для хранения базы - SSD диски
- Распараллеливание тяжелых расчетов на 8 потоков
- Индексация ключевых таблиц базы (в среднем по 7 вспомогательных индексов)

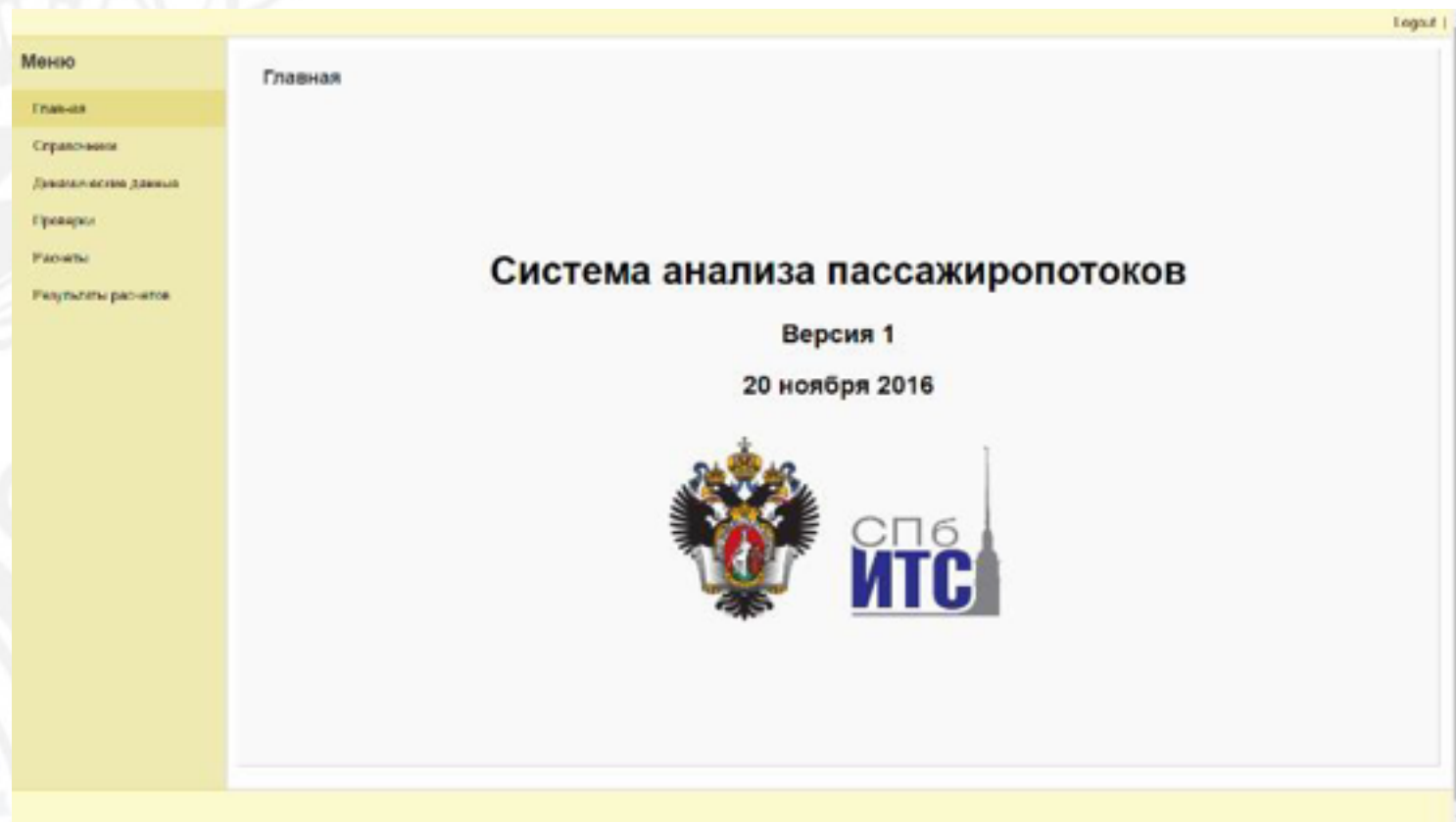


САМЫЕ РЕСУРСООЕМКИЕ ОПЕРАЦИИ

- Самые ресурсоемкие операции - формирование поездок и корреспонденций.
- Главная проблема - информация о реальной поездке пассажира фрагментирована по различным системам с десятками миллионов записей. В одной системе можно найти время оплаты проезда по его проездному документу, в другой - остановку, которую в это время проходил транспорт, в третьей - был ли в это время кондуктор на маршруте (чтобы определить в начале или в конце поездки происходила оплата проезда).
- С учетом разнообразия способов идентификации транспортных средств в различных системах удалось восстановить 92% поездок. Определить оба конца поездки (на какой остановке пассажир вошел и на какой вышел) удалось для 89% поездок.



РАЗРАБОТАННАЯ СИСТЕМА





ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ

- Формирование набора поездок для каждого электронного проездного документа;
- Объединение поездок в корреспонденции (т.е. последовательность поездок, связанных по месту и времени);
- Идентификация регулярных корреспонденций (т.е. корреспонденций, которые совершаются с определенной периодичностью, например, поездки из дома на работу или учебу);
- Формирование матриц корреспонденций с любой степенью детализации (между остановками, транспортными, административными районами и т.д.);



ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ

- Расчет пассажирооборотов для отдельных остановок по маршрутам в разные периоды дня;
- Расчет пассажиропотоков для перегонов всех существующих маршрутов;
- Определение скорости и времени движения транспорта в разные периоды дня по каждому маршруту на каждом перегоне.




КОЛИЧЕСТВЕННЫЕ ХАРАКТЕРИСТИКИ ТРАНСПОРТНОЙ СИСТЕМЫ САНКТ-ПЕТЕРБУРГА

- Среднее количество поездок и корреспонденций в месяц для каждой категории пассажиров (пенсионеры, школьники, студенты, не льготники);
- Затраты времени на трудовые корреспонденции;
- Коэффициент пересадочности для каждой корреспонденции и по системе в целом;
- Городские территории с худшей/лучшей транспортной доступностью.

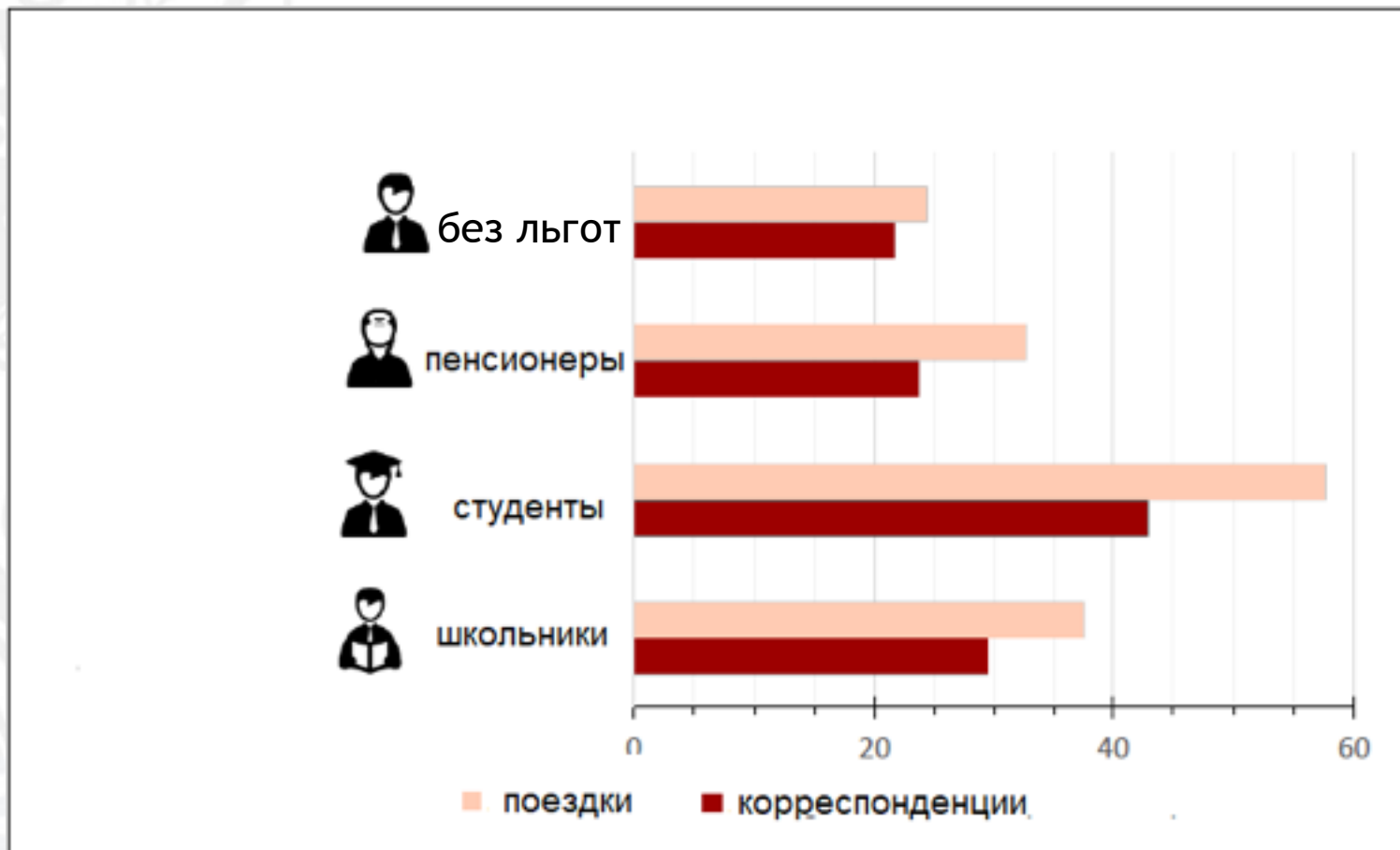


КАТЕГОРИИ ПАССАЖИРОВ ОБЩЕСТВЕННОГО ТРАНСПОРТА

школьники 	ПБ ученический единый
	Школьники с 1-го по 8-ой класс
	ПБ квартальный ученический
студенты 	ПБ студенческий единый
без льгот 	ПБ на 20 поездок и 15 дней
	ПБ многоразовый до 70 поездок по тарифу
	Парковка
	ПБ многоразовый до 60 поездок на 90 дней
	ПБ на 10 поездок и 7 дней
	ПБ на 40 поездок и 30 дней
	Единый на квартал на БСК
	ПБ единый ЛО
	Проездной билет на наземный транспорт
	Количество поездок
Единый Электронный Билет	
пенсионеры 	ПБ единый льготный на БСК
	Пенсионер ЛО

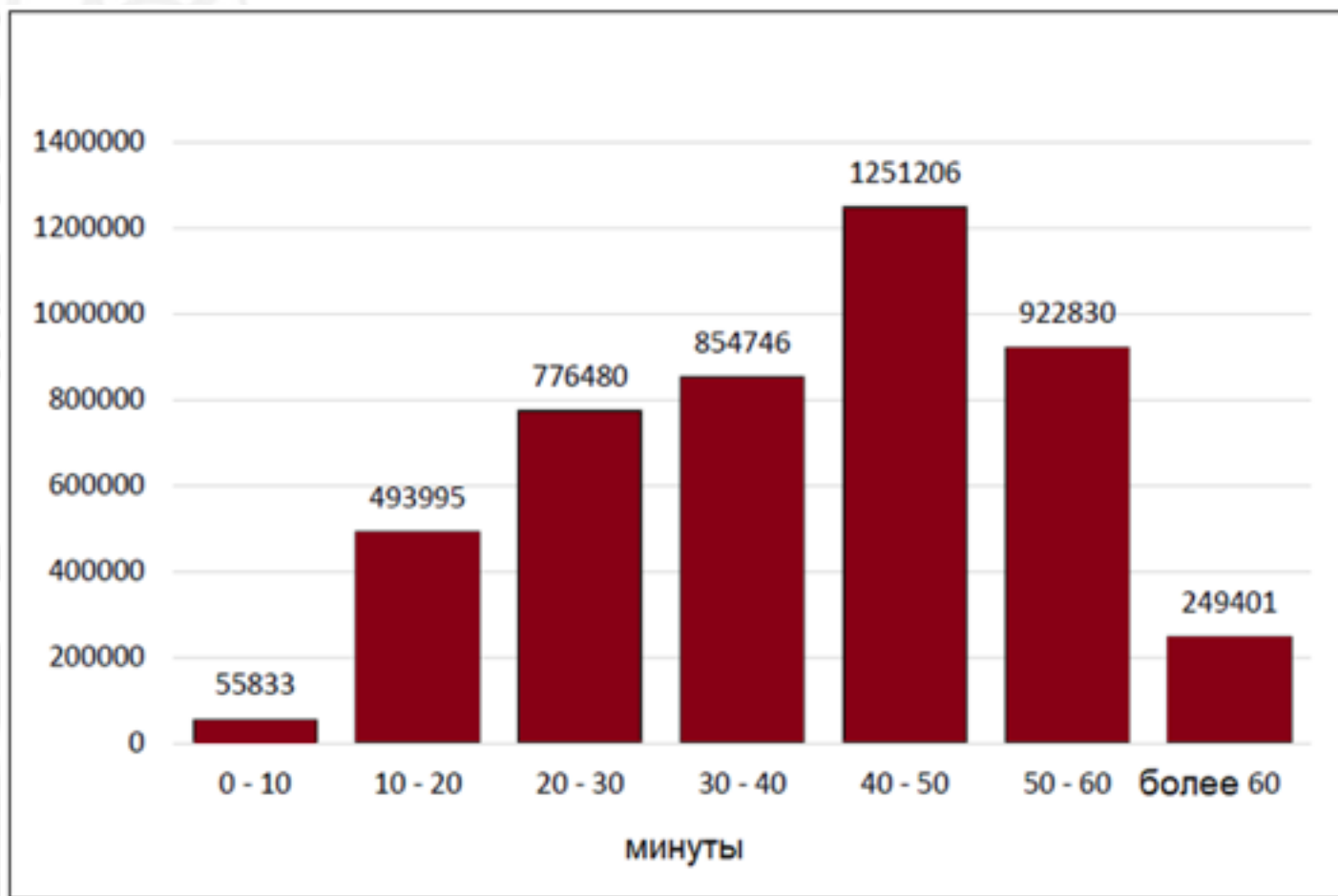


СРЕДНЕЕ КОЛИЧЕСТВО ПОЕЗДОК И КОРРЕСПОНДЕНЦИЙ ЗА МЕСЯЦ





РАСПРЕДЕЛЕНИЕ ТРУДОВЫХ КОРРЕСПОНДЕНЦИЙ (ИЗ ДОМА ДО РАБОТЫ)





КОЭФФИЦИЕНТЫ ПЕРЕСАДОЧНОСТИ

Тип данных	Коэффициент пересадочности
Будни, все корреспонденции	1.24
Выходные, все корреспонденции	1.22
Будни, все устойчивые корреспонденции	1.14
Выходные, все устойчивые корреспонденции	1.21



РЕЙТИНГ РАЙОНОВ ПО УРОВНЮ ТРАНСПОРТНОЙ ДОСТУПНОСТИ

<i>Рейтинг</i>	<i>Район</i>	<i>Средняя продолжительность поездки из дома на работу</i>
3	Адмиралтейский район	28.68
4	Центральный район	29.10
5	Петроградский район	29.52
...
18	Ломоносовский район	37.65
19	Невский район	40.63
20	Красносельский район	44.54



Выявление участков транспортной сети с значительными потерями времени пассажиров;

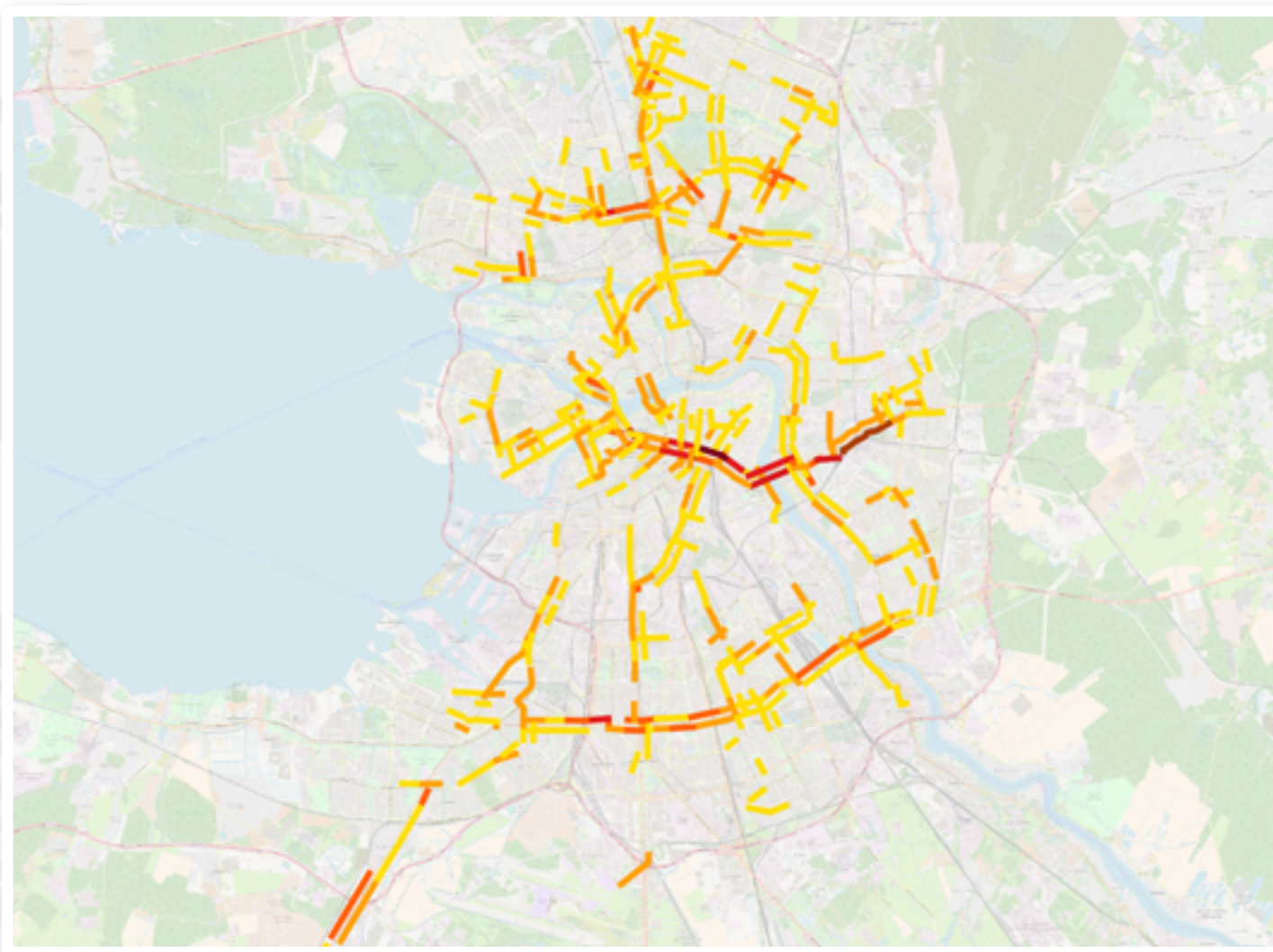
Выявление участков транспортной сети с задержками подвижного состава ГПТ;

Прогноз доходности маршрутов при существенных изменениях маршрутных сетей;

Прогноз потребности в компенсации проезда льготных категорий пассажиров на всех маршрутах.



УЧАСТКИ СЕТИ С НАИБОЛЬШИМИ ПОТЕРЯМИ ТРАНСПОРТНОГО ВРЕМЕНИ





Ваши вопросы?

Санкт-Петербургский
государственный университет
spbu.ru