

УДК 519.688

Интерпретация данных системы контроля и управления доступом

К.И. Киселев

Научный руководитель: старший преподаватель А.Д. Брагин
Национальный исследовательский Томский политехнический университет,
Россия, г. Томск, пр. Ленина, 30, 634050
E-mail: kik15@tpu.ru

Interpretation of physical access control system data

K.I. Kiselev

Scientific Supervisor: Senior lecturer A.D. Bragin
Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050
E-mail: kik15@tpu.ru

Abstract. *In the present study, we address the challenge of interpreting data from the Physical Access Control System (PACS) of Tomsk Polytechnic University (TPU). Given the complexity of the data, which includes inconsistencies and missing information, the research devises a novel interpretation method aimed at identifying actual entries and exits.*

Key words: PACS, ETL, EDA

Введение

Согласно зарубежным экспертам, учебная аналитика становится одним из важнейших технологических трендов, обеспечивающих эффективное использование данных в образовательной среде [1]. Административные государственные организации осознают значимость обработки Big Data с целью принятия управленческих решений, основываясь на data-driven подходе.

Точность и актуальность Business Intelligence & Analytics зависят от качества используемых данных. Данная работа посвящена построению процесса извлечения, преобразования и загрузки (ETL) для задачи учета пребывания в учебных корпусах. Процесс обработки усложняется неоднородным содержанием и проблемами качества данных.

Материалы и методы исследования

Система контроля и управления доступом (СКУД, англ. PACS) Томского политехнического университета (ТПУ) фиксирует проходы в корпуса, осуществляемые по пропуску. В базе данных ТПУ информация хранится в таблице с основными полями: Личность, Дата и время, Направление, Корпус, Событие (таблица 1).

Данные содержат пропуски: направление неизвестно, событие неизвестно. Направление неизвестно наблюдается в учебных корпусах А, Б и К. Для каждого корпуса процент таких записей: А – 20 %, Б – 25 % и К – 1 %.

Встречаются подряд идущие записи проходов одного человека, разница времени между которыми составляет несколько секунд. Необходимо учитывать, что повторные считывания описывают один осуществлённый проход.

Прежде чем переходить к анализу, исходные данные следует интерпретировать. Необходимо понять фактические входы и выходы из корпуса. В качестве решения использовался следующий алгоритм:

- Чтение данных. Сортировка по полям Личность и Дата и время. Отбрасывание записей без направления.
- Рассмотрение проходов по одному корпусу.
- Поиск подряд идущих записей вход-выход. Отбрасывание пар с одинаковым временем.

Преимуществами метода интерпретации являются:

- низкое требование к исходным данным – задействованы поля Личность, Дата и время, Направление, Корпус;
- отсутствие необходимости в доверительном интервале разницы во времени, чтобы отбрасывать повторные считывания.

В качестве недостатков выступают:

- интерпретация данных за прошедшие дни. Метод не рассчитан на работу в режиме реального времени;
- отбрасывание непарных проходов, например, когда после входа в один корпус следует вход в другой корпус. Эти записи можно выделять среди остальных, отбрасываемых в процессе интерпретации.

Таблица 1

Исходные данные и их интерпретация для одного человека за 26 апреля 2023 года

Личность	Дата и время	Направление	Корпус	Событие	Интерпретация
91	2023-04-26 12:38:22	вход	Учебный корпус Д	Зарегистрирован проход	—
91	2023-04-26 14:18:48	—	Учебный корпус Б	—	—
91	2023-04-26 14:18:52	вход	Учебный корпус Б	В доступе отказано	—
91	2023-04-26 14:18:54	вход	Учебный корпус Б	В доступе отказано	—
91	2023-04-26 14:18:58	вход	Учебный корпус Б	Зарегистрирован проход	—
91	2023-04-26 14:19:56	вход	Учебный корпус Б	Проход не начат	<i>вход</i>
91	2023-04-26 16:10:22	выход	Учебный корпус Б	Зарегистрирован проход	<i>выход</i>

Результаты

Интерпретация данных за 2023 год использует 90 % всех записей и отбрасывает 10 %. Точность определения фактических проходов на используемых записях составляет 99 %. Ошибки в интерпретации возникают в случаях:

- когда человек отметил, что выходит, но вернулся в корпус;
 - когда датчики считывания находятся не только на входе в корпус, но возле кабинетов.
- Для некоторых учебных корпусов метод интерпретации не работает:

- когда для записей характерны одновременные входы и выходы. Например, вход в 12:40:15 и выход в 12:40:15;
- когда среди отбрасываемых в процессе интерпретации записей наблюдается большой процент непарных проходов (таблица 2).

В указанных корпусах вместо турникетов, для которых обязательно подтверждать проход по пропуску, установлены терминалы, позволяющие проходить не отмечаясь.

Учебные корпуса, для которых не работает метод интерпретации данных

Корпус	Количество записей	Отброшенные записи, %	Установленные посещения	Непарные проходы, %	Непарных проходов из отброшенных записей, %
Учебный корпус Ж	1 014 040	99.75	1 288	93.98	1.99
Учебный корпус Е	75 164	97.26	1 028	48.24	1.31
Учебный корпус И	22 808	96.87	357	66.57	3.22
Учебный корпус В	49 787	24.72	18 739	38.34	94.66
Учебный корпус Д	78 917	18.76	32 056	31.21	98.20
Учебный корпус З	107 734	17.09	44 661	25.81	84.38
Учебный корпус К	907 262	11.79	400 137	19.54	90.81
Учебный корпус Г	36 305	10.83	16 178	18.91	96.06

Заключение

Результатом работы является ELT процесс по интерпретации данных, полученных с помощью СКУД. Разработанный метод применяется в Центре управления данными ТПУ для задач, задействующих данные посещаемости корпусов.

Были проработаны вопросы, связанные с проблемами качества исходных данных. Результатом является снижение количества повторных считываний с февраля 2024 года. Также обсуждается вопрос актуализации хранимых данных, например, добавление информации о датчике считывания, с помощью которого регистрируются проходы [2].

Список литературы

1. Прохорова М.П., Лебедева Т.Е., Лабазова А.В., Белоусова К.В. Тенденции и перспективы развития он-лайн образования. – Текст: электронный // Инновационная экономика: перспективы развития и совершенствования. – 2019. – № 1 (35). – URL: <https://cyberleninka.ru/article/n/tendentsii-i-perspektivy-razvitiya-on-layn-obrazovaniya> (дата обращения: 15.03.2024).
2. Козлов А.Е. Система контроля и управления доступом на предприятие: понятие, характеристика и основные требования. – Текст: электронный // Вестник ВГТУ. – 2019. – № 1. – URL: <https://cyberleninka.ru/article/n/sistema-kontrolya-i-upravleniya-dostupom-na-predpriyatie-ponyatie-harakteristika-i-osnovnye-trebovaniya> (дата обращения: 15.03.2024).