



Piotr Łapiński

Analiza ścieżek osób zwiedzających muzeum
Brama Poznania na bazie logów audioprze-
wodników

Analysis of paths taken by visitors of the Porta
Posnania museum based on audioguide logs

Praca licencjacka

Promotor: dr Maciej Beręsewicz

Prace przyjęto dnia:

Podpis promotora

Kierunek: Informatyka i Ekonometria

Poznań 2021

Spis treści

| | |
|--|-----------|
| Wstęp | 1 |
| 1 Internetu Rzeczy i Bramą Poznania | 3 |
| 1.1 Definicja i historia Internetu Rzeczy | 3 |
| 1.1.1 Pochodzenie nazwy oraz definicja | 3 |
| 1.1.2 Rozwój Internetu Rzeczy | 4 |
| 1.1.3 Internet Rzeczy i Big Data | 6 |
| 1.2 Smart Cities i Smart Tourism | 7 |
| 1.2.1 Koncept Inteligentnych Miast | 7 |
| 1.2.2 Inteligentna Turystyka jako element budowy Inteligentnego Miasta . . | 9 |
| 1.2.3 Implementacja i użycie audioprzewodników w muzeach w Polsce i na świecie | 10 |
| 1.3 Rozwiązania użyte w Bramie Poznania | 12 |
| 1.4 Podsumowanie | 15 |
| 2 Teoria grafów w badaniu ścieżek | 16 |
| 2.1 Analiza | 16 |
| 2.2 Teoria grafów | 16 |
| 2.2.1 Definicja Grafu | 17 |
| 2.2.2 Trawersowanie grafów | 17 |
| 2.3 Analiza sieci społecznościowych | 18 |
| 2.4 Sieci temporalne | 19 |
| 2.5 Implementacja w języku Python | 20 |
| 2.5.1 NetworkX Graph | 20 |
| 2.5.2 Pathpy2 | 20 |

| | |
|---|-----------|
| 2.5.3 Pandas | 22 |
| 2.6 Podsumowanie | 24 |
| 3 Analiza ścieżek odwiedzających Bramę Poznania | 25 |
| 3.1 Pozyskanie i analiza danych z audioprzewodników | 25 |
| 3.2 Proces czyszczenia danych | 27 |
| 3.2.1 Ograniczenie okresu czasowego | 27 |
| 3.2.2 Segmentacja sesji działania urządzeń | 28 |
| 3.2.3 Segmentacja sesji użytkowników | 31 |
| 3.3 Statystyki zwiedzających (średni czas trwania, opis zwiedzających osób) | 32 |
| 3.3.1 Czas zwiedzania | 32 |
| 3.3.2 Charakterystyka grup zwiedzających muzeum | 36 |
| 3.3.3 Stopień wykorzystywania audioprzewodników | 37 |
| 3.4 Przygotowanie struktury grafu atrakcji muzealnych | 38 |
| 3.4.1 Zbiór przejść - krawędzie grafu skierowanego ważonego | 38 |
| 3.4.2 Zbiór stref – węzły grafu | 40 |
| 3.5 Wizualizacja ścieżek obieranych przez zwiedzających | 41 |
| 3.6 Wnioski | 46 |
| Podsumowanie | 47 |
| Bibliografia | 50 |
| Spis rysunków | 52 |
| Spis programów w języku Python | 53 |

Wstęp

Często na myśl o przyszłości pierwszym obrazem, który widzimy jest wizja technologii zintegrowanej z naszym codziennym życiem. Pomimo małego prawdopodobieństwa unormalizowania się latających samochodów, ogromnych statków kosmicznych czy innych wytworów rodem z powieści i filmów science fiction zaawansowane rozwiązania technologiczne pojawiają się w naszym otoczeniu niebywale szybko, choć często dzieje się to w miejscowościach, których zwyczajnie nieauważamy (Ahlgren i in. 2016).

Wraz z rozwojem "inteligentnych" urządzeń pojawiła się fala wcześniej niedostępnych nam danych. Przykładami takich zmian mogą być skanery RFID pojawiające się w miejscowościach zatrudnienia, zmniejszające potrzebę manualnej weryfikacji tożsamości i czasu pracy lub zastępujące kasjerów w modelach samoobsługowych zakupów, systemy inteligentnych domów umożliwiające kontrolę urządzeń elektronicznych za pomocą aplikacji mobilnych lub systemów asystentów domowych takich jak Google Home czy Amazon Alexa.

Tak zaprojektowane sieci urządzeń wchodzących ze sobą i ich użytkownikiem w bliskie interakcje umożliwiają nam dostęp do informacji o ludzkich potrzebach i zachowaniach, które wcześniej dostępne były tylko poprzez kosztowne i wymagające wielu pracowników manualne rozwiązania (Chui i in. 2019).

Muzeum *Brama Poznania* jest jednym z niewielu muzeów w Polsce (Naramski 2020), które zaadaptowało system inteligentnych audioprzewodników pozwalający na płynną narrację podczas zwiedzania i zapewniające dużo bardziej immersywne doświadczenia oraz dostęp do większej ilości angażujących treści. Przez połączenie przewodników z różnorakimi częściami ekspozycji i ich integrację z jej interaktywnymi elementami takimi jak audiowizualne ekrany, ruchome makiety czy stoły z progresywnymi mapami, obrazującymi rozwój zabudowań na terenie wyspy, muzeum Bramy Poznania nie tylko uczyniło zwiedzanie dużo bardziej interesującym ale również umożliwiło zbieranie danych dotyczących wszystkich ich atrakcji.

Celem tej pracy jest oczyszczenie i interpretacja danych zebranych przez audioprzewodniki w okresie między rokiem 2016 a 2018 w celu przeprowadzenia analizy ścieżek, która mogłaby zapewnić wgląd w preferencje zwiedzających i pozwolić Bramie Poznania spojrzeć na ich ekspozycje z nowej perspektywy.

Wszystkie dane wykorzystane w tym procesie zostały dostarczone przez pracowników muzeum w postaci surowych logów z urządzeń zawierających informacje o akcjach podjętych w określonych momentach, oraz odwiedzanych przez zwiedzających strefach i uruchamianych przez nich nagraniach. Dane zostaną oczyszczone po czym wyodrębnione zostaną z nich szczegółowe informacje dotyczące sesji konkretnych zwiedzających, zachowując pełną ich anonimowość, co umożliwi ich dalszą analizę, której rezultatem będzie graficzna reprezentacja najczęstszych ścieżek obieranych przez zwiedzających, obecne odchylenia od domyślnego nurtu zwiedzania oraz poziom zaangażowania w dodatkowe atrakcje.

Praca złożona jest z trzech rozdziałów. Pierwszy obejmował będzie spojrzenie na koncepcję Internetu Rzeczy, jego historię, definicję i wykorzystanie w modelach inteligentnych miast i intelligentnego zwiedzania muzeów takich jak Brama Poznania. Drugi rozdział skupiony będzie na procesie wymaganym do uzyskania przydatnych informacji oraz oczyszczenia danych otrzymanych z audioprzewodników, jak również opis bibliotek i metod wykorzystanych podczas pracy.

Ostatni rozdział zawierał będzie opis przeprowadzonej analizy wraz z kodem i prezentacją wyników w postaci mapy cieplnej dodatkowych atrakcji muzeum, wizualizacji ścieżek obieranych przez użytkowników oraz stopień w jakim korzystają oni z audioprzewodników w trakcie zwiedzania.

Rozdział 1

Internetu Rzeczy i Brama Poznania

1.1 Definicja i historia Internetu Rzeczy

1.1.1 Pochodzenie nazwy oraz definicja

Jako pierwsze wystąpienie określenia *Internet Rzeczy* (z ang. *Internet of Things*) uważa się tak zatytuowaną prezentację członka społeczności RFID Kevina Ashton w Procter & Gamble w roku 1999, podczas której omawiał możliwości integracji łańcucha dostaw P&G z technologią RFID i preżnie rozwijającym się u schyłku milenium Internetem (Ashton 2009).

Dwadzieścia lat później wraz z popularyzacją Internetu, rosnącą mocą obliczeniową i malejącymi cenami urządzeń elektronicznych termin Internetu Rzeczy przekształcił się z pomysłu młodego przedsiębiorcy w nieodłączną część funkcjonowania wielu firm i codziennego życia wielu ludzi.

Definicja Internetu Rzeczy kształtowała się wraz z jego rozwojem. Począwszy od ogólnego rozwinięcia definicji Internetu będącego siecią komputerów, do sieci tworzonej między innymi urządzeniami (Evans 2011). Dzisiaj, jako najdokładniejszą definicję mianuje się definicję IERC określoną w raporcie World Wide Web Consortium (Barnaghi i in. 2011):

Internet Rzeczy jest dynamiczną globalną infrastrukturą sieciową ze zdolnością samokonfiguracji bazującą na standardowych i interoperacyjnych protokołach komunikacyjnych gdzie rzeczy fizyczne i wirtualne posiadają tożsame, atrybuty fizyczne i osobowości wirtualne oraz używają inteligentnych interfejsów, będąc płynnie zintegrowane w sieć informacyjną

1.1.2 Rozwój Internetu Rzeczy

Powstanie Internetu Rzeczy wywołane było rosnącą ilością urządzeń podłączonych do światowej sieci oraz ułatwionym dostępem do coraz bardziej przenośnych technologii. Potrzeba komunikacji między nimi, oraz rosnący potencjał informacji z nich pozyskiwanych doprowadził do ewolucji narzędzia wcześniejszej służącego głównie do wymiany informacji między ludźmi. Za okres odpowiadający "narodzinom" Internetu rzeczy przyjmowany jest przełom roku 2008 i 2009, kiedy to ilość podłączonych do Internetu urządzeń przerosła liczbę żyjących na Ziemi ludzi.

Początkowo Internet Rzeczy składał się z odrębnych sieci a każda z nich była wyspecjalizowana w jednej z wielu dziedzin, gdzie mogły znaleźć zastosowanie. Powstawały sieci mające na celu usprawnić transport towarów realizujące założenia pierwszych konceptów Kevina Ashton, sieci biznesowe, ułatwiające funkcjonowanie zakładów pracy oraz systemy mające na celu pobieranie opłat za przejazdy bez konieczności zatrzymywania pojazdów.

Z czasem powstające między sieciami połączenia umożliwiły lepsze zarządzanie nimi, oraz podniesienie poziomu ich bezpieczeństwa. W tym samym czasie przez dostęp do pozyskiwanych z nich danych i ich analizę otwarte zostały drzwi do nowych poziomów optymalizacji efektywności produkcji, dostaw oraz wielu innych.

Niewątpliwie stymulantem dla rozwoju tej technologii okazała się rosnąca popularność smartfonów, tabletów i coraz wydajniejszych przenośnych komputerów. Wraz z wzrastającą ilością tych urządzeń na rynku pojawiły się nowe możliwości. Każdy posiadacz smartfona posiadał w swoim rękę kontroler będący w stanie obsługiwać inne urządzenia. Narodziły się w ten sposób systemy inteligentnych domów umożliwiające zdalną kontrolę zamków, termostatów, oświetlenia i systemów alarmowych. Naprzeciw ich rosnącej popularności wyszło wiele firm, z których warto wspomnieć o globalnie znany Amazonie.

W czerwcu 2014 roku na rynku ukazał się Amazon Fire Phone, pierwszy smartfon wyprodukowany przez firmę Jeffa Bezosa. Popularność tego projektu okazała się niewystarczająca, jednak elementy jego oprogramowania przetrwały do dziś. Prezes Amazon chcąc wykorzystać technologię rozpoznawania mowy obecną w tym urządzeniu rozwinał ją, a następnie wprowadził dla niej odrębny produkt (Duhigg 2019).

Do pierwotnej funkcjonalności umożliwiającej odtwarzanie dowolnej piosenki o którą poprosił użytkownik dodane zostały informacje pogodowe, możliwość wyszukiwania głosowego oraz integracji z systemami inteligentnych domów. W ten sposób nowy produkt nazwany Ama-

zon Echo, osiągnął wysokie wyniki sprzedażowe oraz dostarczył amerykańskiej firmie nowych możliwości rozwoju. Dziś, powszechnie znany jako Alexa, domowy asystent Amazonu obecny jest w prawie 25% amerykańskich domów (Marketing Land 2020).

Wielorakie zastosowania znalazła również technologia RFID (z ang. *radio-frequency identification*), której popularność rozwijała się w latach 90 XX wieku a perspektywa jej rozwoju była tematem dyskusji nawet w roku 2015. (Landt 2005).

Wraz z rosnącą precyją wykonania nadajników i możliwością umieszczenia ich w elastycznych naklejkach, plastikowych kartach czy nawet przedmiotach codziennego użytku rozwijały się zaczątko rozwiązania, które szybko ustabilizowały swoją pozycję w wielu sektorach gospodarki.

Przykładem adaptacji technologii RFID może być rozwój płatności kartą. Pierwotnie, karty płatnicze posiadające pasek magnetyczny należało umieścić w terminalu, lub przesunąć paśkiem po czytniku. Poprzez umieszczenie we wnętrzu karty cewki aktywowanej przez pole elektromagnetyczne generowane w terminalu skróceniu uległ czas transakcji, oraz wzrosło bezpieczeństwo danych osobowych. Pomimo zasadniczych różnic między zasięgiem standardowego nadajnika a tak zwanych kart zbliżeniowych, zdefiniowanie oraz utrzymywanie standardów bezdotykowych płatności można w dużej mierze przypisać współpracy Visa oraz Mastercard, dwóch największych dostawców usług pośrednictwa finansowego, z Secure Tech Alliance¹ (Smart Card Alliance 2007).

Możliwość płatności za zakupione produkty jednym ruchem ręki szybko stała się standardem, co doprowadziło do kolejnych innowacji. Popularna sieć sklepów z odzieżą i artykułami sportowymi Decathlon od dawna używała w swoim łańcuchu dostaw technologii RFID, jednak w 2014 roku rozpoczęły się projekty, które doprowadziły do opatrzenia tagiem RDIF wszystkich produktów na każdym etapie ich produkcji. W artykule na polskiej stronie medialnej, Decathlon tak ocenił zalety wprowadzonego rozwiązania:

Decathlon (2019): *Technologia radiowego przesyłania danych pomaga sieci usprawniać procesy kasowe, pozwalając wygospodarować oszczędności w czasie pracy zespołów. (...) Wprowadzenie technologii RFID umożliwia również bieżącą kontrolę stanów magazynowych i identyfikację konkretnych produktów na stanowiskach kasowych, co przekłada się na eliminację kradzieży. Największą korzyścią jest jednak przyspieszenie procesu inwentaryzacji – teraz dzieje się ona 5 razy szybciej niż kiedyś.*

¹Dawniej: Smart Card Alliance.

Opisane przykłady są jednymi z wielu adopcji zasad funkcjonowania Internetu Rzeczy na światowym rynku. Z roku na rok liczba nowych urządzeń połączonych z siecią rośnie (Prognozowane wartości osiągają nawet 16 miliardów w roku 2025, Ivudria (2021)) a wraz z nimi, ilość dostępnych danych oraz stopień kontroli własnego otoczenia przez człowieka a rosnące możliwości komunikacji nie tylko między ludźmi, ale również między martwymi obiektami stały się źródłem wielomiliardowych zysków rzędu 51 miliardów dolarów w roku 2021 z prognozowanym 76% wzrostem do roku 2025 (Ivudria 2021).

1.1.3 Internet Rzeczy i Big Data

Rosnąca liczba urządzeń połączonych ze sobą i światową siecią internetową niosła za sobą możliwość pozyskiwania nowych danych, zarówno o samych urządzeniach, takich jak ich stan, lokacja etc. jak i danych generowanych lub zbieranych przez połączone ze sobą w sieci rzeczy. Wynikiem tego było uzyskanie dwóch typów danych umożliwiających odpowiednio optymalizację i analizę systemów oraz infrastruktury z których były pozyskiwane jak i wgląd w interakcje między użytkownikiem a urządzeniem (Tsai i in. 2014). Pojawienie się tych nowych danych przyniosło za sobą możliwość przetworzenia ich w informacje będące źródłem dużej ilości wiedzy.

Ze względu na swoją charakterystykę oraz ogromną ilość, dane pozyskiwane z urządzeń, skanerów RFID i innych technologii wchodzących w skład Internetu Rzeczy, można zakwalifikować do zbiorczego terminu Big Data. Ze względu na dużą ilość i niskie usystematyzowanie danych pozyskiwanych w ten sposób powstało zapotrzebowanie na określenie sposobów efektywnego ich pozyskiwania, oczyszczania i analizy. Przegląd oferowanych i użytkowanych sposobów "kopania danych" z Internetu Rzeczy zaprezentował w swoim artykule Tsai i in. (2014). Nieusystematyzowany styl uzyskiwanych danych oraz różne możliwości ich zastosowania skłoniły autora do określenia uczących się algorytmów oraz algorytmów inteligencji roju jako potencjalnie najefektywniejszych metod pozyskiwania i analizowania danych z szerokiego zakresu urządzeń.

W dzisiejszym świecie powstają systemy zaprojektowane w sposób łączący względnie nieograniczone zasoby Big Data, z komunikatywnością i responsywnością rozwiązań Internetu Rzeczy. Zaproponowany przez Chui i in. (2019) system monitorowania zachowań pacjentów, mający umożliwić przewidywanie i zapobieganie chorobom układu krążenia oraz urazów wywołanych przez upadki obrazuje przyszłościowe ukierunkowanie opieki zdrowotnej, skupiającej się na wszystkich aspektach pobytu chorego w szpitalu. Tego typu rozwiązania automatyzacyjne,

mimo napotykanych jeszcze problemów w bezpieczeństwem oraz zaufaniem pomogą odciążyć pracowników opieki zdrowotnych i podnieść jakość oferowanych przez szpitale usług.

Obecnie, preferowanym narzędziem do analizy i przetwarzania danych w czasie rzeczywistym są rozwiązania chmurowe, umożliwiające szybkie wykonywanie zadań wymagających dużej mocy obliczeniowej w połączeniu z minimalizacją kosztów. Tego typu rozwiązanie sprawdza się idealnie w przypadku, gdy bezpośredni oraz niezakłócony dostęp do chmury jest możliwy. W przypadku urządzeń wchodzących w bezpośrednią interakcję z ich użytkownikiem, takie warunki nie zawsze są osiągalne. Aby możliwie jak najlepiej wykorzystać możliwości Internetu Rzeczy oraz przetwarzania w chmurze, zaproponowane zostały rozwiązania nazywane potocznie "mgłą Internetu Rzeczy" (z ang. *Fog IoT*).

Opis oraz charakterystykę przetwarzania w "mgłę" zaprezentował Oteafy i Hassanein (2018) obejmując zakres wad i zalet obu technologii oraz powstające między nimi możliwości rozwoju. Poprzez skupienie jak największej ilości drobnych, niewymagających dużej mocy obliczeniowej operacji jak najbliżej użytkownika, poprzez wykonywanie ich przez podłączone do sieci urządzenia lub urządzenie agregujące i kontrolujące ich operacje, powstała możliwość uzyskania szybszego czasu reakcji oraz zmniejszenia stopnia polegania na infrastrukturze chmurowej. W ten sposób w czasie, gdy wykonywana jest operacja wymagająca udziału "chmury", równolegle wykonywane są wszystkie, które tego nie wymagają zachowując pełną funkcjonalność pierwotnego systemu.

Przetwarzanie danych uzyskanych z połączonych ze sobą urządzeń za pomocą uczących się, lub przystosowanych do rozproszonej natury Internetu Rzeczy algorytmów (Mahdavinejad i in. 2018) oraz integracja zewnętrznych rozwiązań takich jak przetwarzanie w chmurze z ograniczeniem możliwości błędów lub utraty funkcjonowania systemów jest więc ważnymi krokami na drodze do stworzenia inteligentnej sieci monitorowania, zarządzania i pełnego wykorzystywania istniejących już połączeń między inteligentnymi urządzeniami oraz ich użytkownikami.

1.2 Smart Cities i Smart Tourism

1.2.1 Koncept Inteligentnych Miast

Wraz ze wzrostem ilości ludności oraz postępującym rozwojem technologicznym nowoczesne miasta coraz częściej zmuszane są do dostosowywania się względem ulegającym zmianom warunkom w celu zachowania wydajnego stylu funkcjonowania. Koncepcja Inteligent-

nych Miast (z ang. *Smart Cities*) powstała w celu ułatwienia monitorowania miejskiego zużycia zasobów, podniesienia poziomu życia mieszkańców oraz umożliwienia podniesienia jakości oferowanych przez miasto udogodnień i usług. Pomimo rosnącej popularności koncept ten nie posiada jeszcze ustrukturyzowanej definicji. Aby ułatwić formalizację tego terminu, obejmującego zakres funkcjonowania całego miasta, wyszczególnione zostały wymiary określające dziedziny jego zastosowania (Caragliu i in. 2011; Naramski 2020; Vanolo 2014):

- **Smart Mobility:** Wykorzystanie infrastruktury sieciowej do usprawnienia wydajności gospodarczej i politycznej jest cechą charakterystyczną tego wymiaru. Przez wdrożenie rozwiązań zwiększających łączność między strefami rozwoju miejskiego, społecznego i kulturalnego propagowany jest rozwój miasta w ramach jego inteligentności.
- **Smart Environment:** Głównym założeniem wymiaru środowiskowego jest racjonalna gospodarka zasobami miasta, zarówno pod względem ograniczonych zasobów naturalnych jak i kulturalnych. Identyfikacja źródeł przychodu oraz dziedzictwa kulturalnego miasta jest ważnym punktem podczas jego rozwoju ponieważ umożliwia rozwój jego osobowości i kształtowanie jego środowiska.
- **Smart Economy:** Aby stawać się atrakcyjnymi, miasta przyjmują elastyczne i innowacyjne podejście do własnego rozwoju, stawiając na rozwój wysokich technologii i kreatywnych dziedzin przedsiębiorczości. Przez ukierunkowanie rozwoju gospodarczego oraz rozwój tak zwanej *Infrastruktury Miękkiej*, na którą składają się sieci wiedzy, środowiska wolne od przestępstw czy organizacje charytatywne miasto staje się bardziej zachęcającym miejscem dla wysoko wykwalifikowanych i kreatywnych pracowników.
- **Smart Governance:** Inteligentne zarządzanie miastem wymaga transparentności oraz udziału jego mieszkańców w podejmowaniu najważniejszych dla niego decyzji. Znaczącą rolę w tym procesie odgrywa dostępność usług publicznych, ich obsługa oraz strategia polityczna miasta.
- **Smart Living:** Kluczowym elementem rozwoju inteligentnego miasta jest stopień w jakim wprowadzane zmiany wpływają na życie mieszkańców. Składa się na to dostępność usług edukacyjnych, kulturalnych, poczucie bezpieczeństwa, dostępność atrakcji turystycznych oraz usprawnień takich jak karty miejskie czy systemy monitorujące aktywność transportu publicznego w czasie rzeczywistym.
- **Smart People:** Rozwój inteligentnego miasta powinien prowadzić do wzrostu jakości kapitału ludzkiego, budząc pokłady ciekawości i kreatywności. Często ten termin realizo-

wany jest przez prowadzenie neoliberalnej polityki, wysoką elastyczność miast w stosunku do potrzeb jego mieszkańców oraz prowadzenie polityki mającej na celu zwiększenie ilości firm na jego terenie.

Powyższe wymiary mogą jednocześnie stanowić kryterium, na bazie którego można wyróżniać inteligentne miasta. Caragliu i in. (2011) o określaniu miasta mianem inteligentnego pisze (w tłumaczeniu autora): *Uważamy miasto za inteligentne, gdy inwestycje w kapitał ludzki i społeczny oraz gdy tradycyjna (transportowa) i nowoczesna (Informatyczna) infrastruktura komunikacyjna zasila zrównoważony wzrost gospodarczy i wysoką jakość życia, z mądrym zarządzaniem zasobami naturalnymi poprzez sprawowanie partycypacyjnego zarządzania miastem.*

Przykładem integracji pomiędzy rozwiązaniami Internetu Rzeczy oraz konceptem Inteligentnego Miasta może być opublikowany przez pracowników szwedzkiego uniwersytetu w Uppsalii artykuł, w którym jako studium przypadku przedstawione jest projekt GreenIoT (Ahlgren i in. 2016). Głównym celem projektu było wyposażenie obszarów miejskich w czujniki umożliwiające wgląd w poziom zanieczyszczenia powietrza i planowania ruchu drogowego w czwartym największym mieście w Szwecji. Największym problemem napotkanym podczas jego realizacji było pozyskiwanie, agregacja i oczyszczenie danych pozyskanych z inteligentnych urządzeń w sposób umożliwiający ich integrację z systemami wspomagającymi wydajność energetyczną miasta i udostępnienie danych w ramach otwartego dostępu dla każdego, kto chciałby zaprezentować swoje rozwiązania w zakresie ich analizy i nowych produktów wykrywających.

1.2.2 Inteligentna Turystyka jako element budowy Inteligentnego Miasta

Turystyka Inteligentna (z ang. *Smart Tourism*) podobnie jak Inteligentne Miasta nie posiada jeszcze dokładnie ukształtowanej definicji. W swoim artykule Naramski (2020) wyróżnia użytkowanie Internetu oraz infrastruktury sieciowej oraz wprowadzanie polityki ukierunkowanej na potrzeby społeczne i kulturalne w rozwoju miejskim jako charakterystykę tego konceptu.

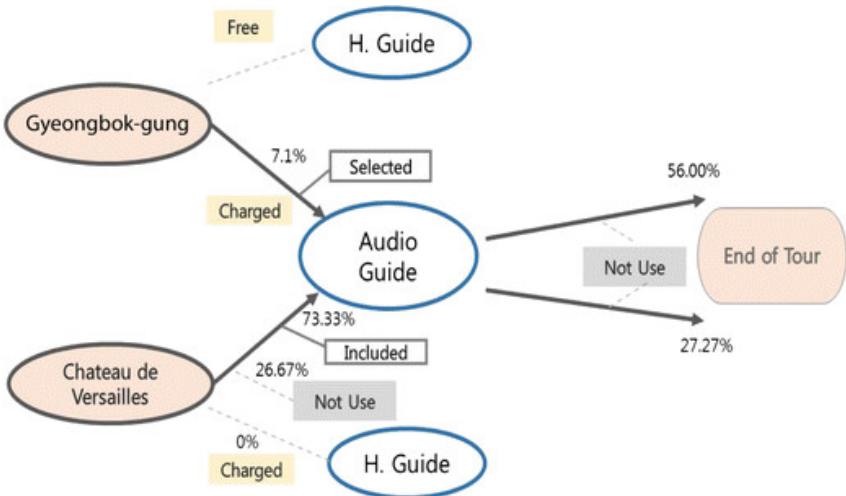
Kiatkawsin i in. (2020) twierdzi, że Inteligentna Turystyka nie jest nowym konceptem, a jedynie ewolucją cyfrowego turysty. Zarówno E-Turystyka jak i Inteligentna Turystyka w swoich założeniach kształtują się przez rozwój technologii informacyjnych oraz komunikacyjnych jednak Smart Tourism pozwala i zachęca do gromadzenia, integracji i analizy danych uzyskiwanych z urządzeń i zastosowanych systemów w celu usprawnienia zwiedzania oraz zaoferowania bardziej immersywnych, interesujących sposobów poznawania ekspozycji i historii zabytków.

Do tego celu wykorzystywanych jest wiele technologii, z których warto wyróżnić Internet Rzeczy, Wirtualną Rzeczywistość, aplikacje mobilne oraz Big Data. Każda z nich posiada swój własny, unikalny wpływ na rozwój doświadczeń turystów i usprawnienia ekspozycji muzealnych. Zastosowanie rozwiązań Internetu Rzeczy umożliwia stworzenie interaktywnego, realnego środowiska w którym połączone z siecią urządzenia przyjmują role narratorów, uczestników lub źródeł wiedzy dla zwiedzających. Uwzględnienie Wirtualnej Rzeczywistości umożliwia przeniesienie turystów w zupełnie inny, wirtualny świat. Przykładowo, przez zastosowanie Wirtualnej Rzeczywistości turyści zwiedzający ruiny rzymskiego Forum Romanum mogliby oglądać je w pełnej świetności w taki sposób, jakby obserwowali je dwa tysiące lat wcześniej. W ten sposób atrakcje turystyczne nabierają nowego wymiaru i potencjału. Zastosowanie mobilnych aplikacji podczas zwiedzania umożliwia oszczędność na systemach audioprzewodnikowych oraz przewodnikach dla małych grup. Dzięki temu rozwiązaniu można również dostarczyć materiały w wielu wersjach językowych bez potrzeby ustawiania nadmiernej ilości tablic, lub pokrywania atrakcji opisami. Przykładem takiego zastosowania może być słynna Sagrada Familia położona w Barcelonie. Po wejściu zwiedzający mają możliwość pobrania aplikacji, zawierającej nagrania i opisy tekstowe atrakcji w ich ojczystym języku. Dostępność telefonów komórkowych sprawia, że większość osób decyduje się podejmować zwiedzanie katedry w ten sposób jednak aplikacja wymaga dużego udziału ze strony użytkownika.

1.2.3 Implementacja i użycie audioprzewodników w muzeach w Polsce i na świecie

Zastosowanie nowych technologii przynosi instytucjom takim jak muzea wiele korzyści. Mimo tego, zastosowanie audioprzewodników w muzeach jest stosunkowo niskie. Lee (2017) podczas wywiadów przeprowadzanych ze zwiedzającymi francuskiego Pałacu Wersalskiego wyróżnił największe przeszkody, wpływające na stopień implementacji tej technologii jako:

- Środowisko o wysokim poziomie hałasu i ludzi mogących rozpraszać zwiedzających.
- Trudności ze zrozumieniem funkcjonowania urządzenia, prowadzące do frustracji i porzucenia użytkowania audioprzewodników przez część zwiedzających.
- Izolacja społeczna wywołana pozbawieniem doświadczenia zwiedzania kontaktu z innymi zwiedzającymi i przewodnikiem.



Rysunek 1.1. Retencja użytkowania audioprzewodników wykazana przez Lee (2017)

Źródło: Lee (2017).

- Brak poczucia indywidualności i elastyczności narracji dostarczanej przez audioprzewodnik, w porównaniu do umiejętności dostosowania się do zwiedzającej grupy przez tradycyjnego przewodnika

W swoim artykule Seung Jae Lee przedstawił również ciekawą zależność między rzeczywistym użytkowaniem audioprzewodników na początku i na końcu zwiedzania, poprzez porównanie Pałacu Wersalskiego oraz Pałacu Gyeongbok. Podczas gdy w Pałacu Wersalskim, audioprzewodniki są wydawane bezpłatnie, w koreańskim Pałacu Gyeongbok są one płatne. W wyniku tej różnicy zobrazowane zostały różnice, między wykorzystaniem darmowych oraz płatnych audioprzewodników, widoczne na Rysunku 1.1.

W przypadku oferowanych bezpłatnie audioprzewodników, ponad jedna czwarta zwiedzających zadeklarowała, że w ogóle ich nie używała, a kolejna jedna czwarta przyznała, że pod koniec zwiedzania przestała korzystać z pomocy audioprzewodnika. W przypadku dobrowolnego, płatnego wyboru audioprzewodnika, mimo, że tylko 7,1% zwiedzających zdecydowało się na jego użycie, w końcowym etapie zrezygnowało z niego aż 56% zwiedzających.

W badaniu 218 polskich muzeów Mateusz Naramski (Naramski 2020) przedstawił stopień zaznajomienia biorących w nim udział muzeów z terminem Inteligentnej Turystyki oraz zastosowane przez nie rozwiązania. Zaledwie 0,9% muzeów twierdzi, że zrealizowały już założenia konceptu Smart Tourism a 17,9% tego typu instytucji planuje wprowadzenie rozwiązań z tego zakresu. Niestety, aż 72% badanych nie było zaznajomionych z tym pojęciem, a 9,2% zadekla-

rował brak planów wprowadzania rozwiązań mających realizować jego założenia. W dalszej części badania, pomimo niskiego stopnia adopcji nowych technologii w muzeach określona została inkluza poszczególnych typów zwiedzania. Uzyskany stopień wykorzystania audioprzewodników w polskich muzeach to 21,1%. Wnioskować więc można, że pomimo użytkowania urządzeń ułatwiających doświadczanie ekspozycji w nadchodzących latach możemy spodziewać się większego stopnia innowacji i zapoznania z nowymi technologiami w placówkach polskiego dziedzictwa kulturalnego.

1.3 Rozwiązania użyte w Bramie Poznania

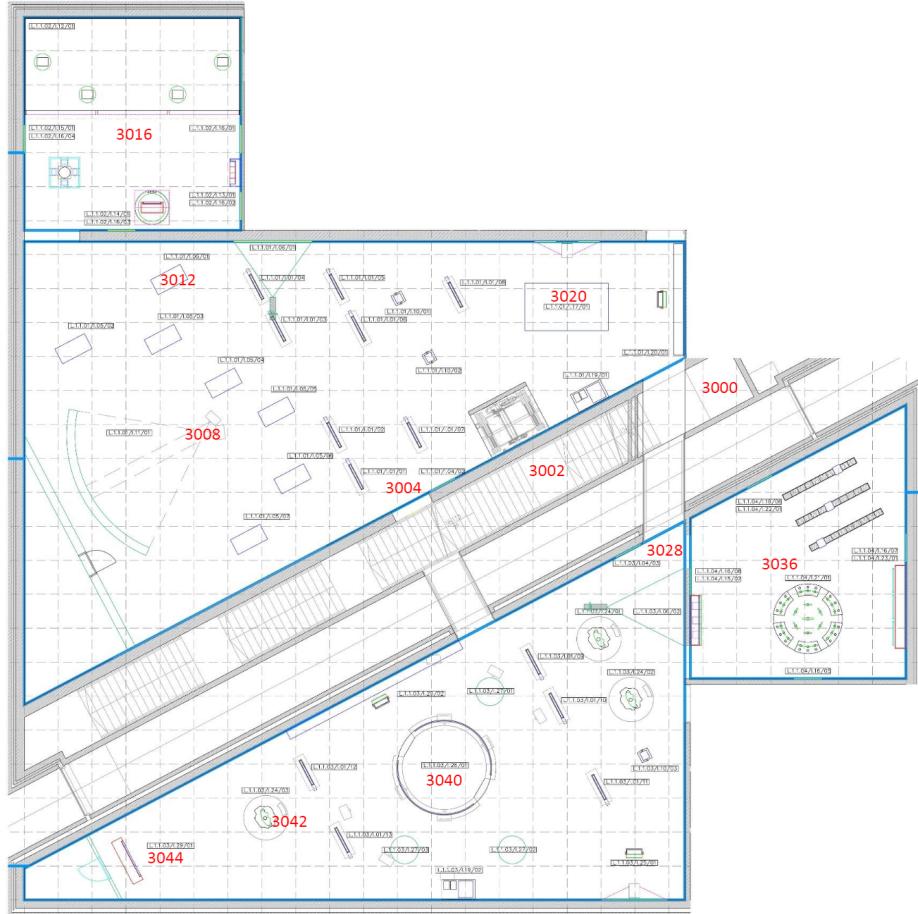
Muzeum Bramy Poznania wcześniej nazywane *Interaktywnym Centrum Historii Ostrowa Tumskiego* położone jest na Ostrowie Tumskim, poznańskiej wyspie na Warcie, oraz opowiada historię początków osadnictwa w Polsce skupiając się na historii terenu, na którym się znajduje. Zastosowanie w placówce audioprzewodników i nowoczesnych technologii było jednym z początkowych założeń realizowanych od momentu jej powstania. Na stronie Bramy Poznania charakterystyka muzeum określona jest następująco:

Brama Poznania opowiada o początkach państwa polskiego i wyspie katedralnej. Nie znajdziecie tu muzealnych eksponatów. Przedstawiamy naszą historię w nowoczesny sposób, z wykorzystaniem multimedialnych i światła. (...) Bramą Poznania jest pierwszym w Polsce centrum interpretacji dziedzictwa, stworzonym na wzór podobnych instytucji działających za granicą. Łączy ono ekspozycję, która otwiera odbiorcom drogę do poznania ważnego miejsca historycznego, jakim jest poznańska wyspa katedralna, bezpośrednio z tym miejscem.

Podczas odwiedzin w muzeum Bramy Poznania, każdy zwiedzający wyposażany jest w audioprzewodnik oraz parę słuchawek. Już od pierwszych postawionych na schodach kroków urządzenie łączy się z emitującymi sygnał o określonej częstotliwości nadajnikami, określając pozycję, w której znajduje się aktualnie jego użytkownik. Całe muzeum podzielone jest na strefy, opisane numerycznie, z których najważniejsze zostały zobrazowane na rysunkach 1.2 oraz 1.3.

Prowadzony przez audioprzewodnik zwiedzający może uczestniczyć w ogólnie zakreślonej ścieżce zwiedzania, lub poszukiwać wiedzy na własną rękę. Znajdujące się w pomieszczeniach ekranы oraz interaktywne makiety zapewniają płynne przejścia między prowadzonym zwiedzaniem oraz ciekawymi atrakcjami. Każdy audioprzewodnik wyposażony jest w diodę podczer-

woną, umożliwiającą aktywację dodatkowych opisów dostępnych na terenie wszystkich pomieszczeń oraz innych atrakcji prowadząc do uzyskania nowych, ciekawych informacji i zwiększających zaangażowanie zwiedzających w ich podróż przez historię poznańskiej wysepki.

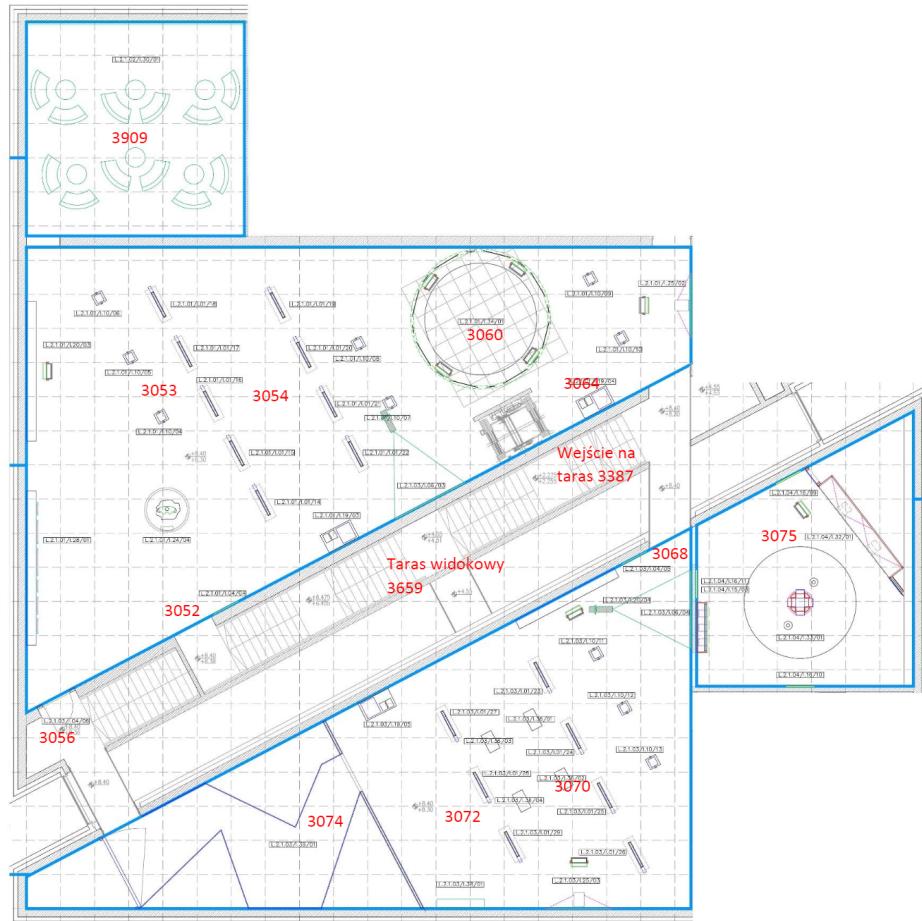


Rysunek 1.2. Rzut pionowy piętra pierwszego z zaznaczonymi strefami

Urządzenia, w sposób anonimowy, zapisują kolejno wykonywane czynności w postaci logów. Poszczególne strefy odwiedzone w trakcie sesji zwiedzania zapisywane są w pamięci podręcznej urządzenia. Gdy strefa odwiedzana przez zwiedzającego widnieje już w pamięci, materiały z nowej, pokrywającej powierzchnię poprzedniej strefy są odtwarzane. Muzeum wykorzystuje do tego wirtualne strefy nazywane przezwaniem. W ten sposób, ponowne odwiedzenie pomieszczenia powoduje odtworzenie nowych nagrań i skierowanie użytkownika na kolejne atrakcje, kontynuujące prowadzoną narrację.

Funkcjonalność urządzeń nie jest ograniczona do indywidualnego zwiedzania. W prosty sposób, przewodnik grupy może sterować biegiem zwiedzania uruchamiając lub wstrzymując nagrania w kolejnych strefach. W ten sposób, łączona jest bezpośrednia interakcja zwiedzają-

cego z pracownikiem muzeum, indywidualne doświadczenie aktywowań elementów wystawy oraz narracja płynąca z głośników słuchawek. W ten sposób, nawet w grupie, każdy z uczestników może doświadczyć ekspozycji na swój własny, unikalny sposób.



Rysunek 1.3. Rzut pionowy piętra drugiego z zaznaczonymi strefami

Jedną z zalet rozwiązania wykorzystywanego w muzeum jest ograniczenie ilości dźwięków dobiegających z głośników na sali wystawy. Umożliwia to odtwarzanie materiałów filmowych w tym samym pomieszczeniu, w którym inni zwiedzający w ciszy skupią się na obserwacji lub czytaniu dodatkowych informacji rozmieszczonych w różnych punktach sali. Ponadto, użycie słuchawek umożliwia zastosowanie efektów dźwiękowych oraz muzyki połączonej z narracją co ma bezpośredni wpływ na immersję zwiedzającego w przedstawianym świecie.

W następnych rozdziałach przedstawiona zostanie struktura danych pozyskanych z audio-przewodników poznańskiego muzeum, metodologia ich oczyszczenia i przygotowania oraz opis bibliotek wykorzystanych do przeprowadzenia analizy.

1.4 Podsumowanie

Rozwój Internetu rzeczy prowadzi instytucje oraz prywatnych użytkowników do integracji technologii w szerokim zakresie codziennych czynności oraz życia czego efektem jest wzrost znaczenia idei Inteligentnych Miast i turystyki. Muzea decydujące się na wykorzystanie w swoich ekspozycjach elementów takich jak audioprzewodniki cieszą się renomą i rosnącą popularnością a wygenerowane w takich instytucjach dane rzadko poddawane są analizom i pozostają niewykorzystane. Takie zbiory mogą umożliwiać wgląd w decyzyjność osób zwiedzających i pomagać w zrozumieniu efektywności ekspozycji. Są one jednak trudne do uzyskania, ekstrakcji oraz oczyszczenia. Dane uzyskane z muzeum *Bramy Poznania* są więc niecodziennym przypadkiem, gdzie tak surowe i nieprzetworzone zbiory udostępniane są na cele akademickie. W następnym rozdziale opisane zostaną metody analizy danych strukturalnych generowanych przez takie urządzenia za pomocą Teorii Grafów oraz Analizy Sieci Społecznościowych.

Rozdział 2

Teoria grafów w badaniu ścieżek

2.1 Analiza

Analiza (z łac. *analysis*) polega na rozkładzie na czynniki określonych zjawisk lub przedmiotów w celu wyodręgnienia ich cech, właściwości oraz składników. Wyróżnia się wiele rodzajów analiz, dzielonych ze względu na różnice w wykorzystywanych podczas nich metodach oraz zakresach ich zastosowania. Analizy materialne są to reakcje chemiczne, umożliwiające badanie substancji i ich składu. Analizy niematerialne obejmują zagadnienia z dziedziny matematyki, informatyki, literatury, ekonomii, kryminologii, systemów i innych niematerialnych bytów lub konceptów.

2.2 Teoria grafów

Teoria grafów po raz pierwszy opisana została przez Leonarda Eulera w referacie naukowym dotyczącym Zagadnienia Siedmiu Mostów na rzece Pregole, opublikowanym w 1736 roku. W swojej pracy poszukując sposobu przejścia przez wszystkie znajdujące się w Królewcu mosty tak, aby każdy z nich przekroczyć tylko raz Euler sformułował warunki, które muszą spełniać grafy spójne (grafy, w których dla każdej pary wierzchołków istnieje zbiór krawędzi je łączących) aby można było przechodzić je w sposób podobny do mostów królewieckich. Grafy takie nazywane są grafami eulerowskimi. Kontynuacja rozważań rozpoczętych przez szwajcarskiego matematyka doprowadziła do rozwoju dziedziny matematyki zwanej topologią (L. 1813), oraz dalszych badań nad sformalizowanymi podklasami grafów, takich jak drzewa (Cayley 1881).

2.2.1 Definicja Grafu

Graf G składa się z niepustego zbioru wierzchołków $V(G)$ oraz zbioru krawędzi $E(G)$ łączących wierzchołki. Można więc zapisać graf jako:

$$G = (V(G), E(G)) \quad (2.1)$$

Aby zobrazować kierunkowość relacji między wierzchołkami stosuje się grafy skierowane, nazywane również digrafami, w których każda krawędź oznaczona jest dodatkowo grotkiem zwróconym w stronę wierzchołka docelowego. Graf skierowany D składa się z niepustego zbioru wierzchołków $V(D)$ oraz zbioru łuków $A(D)$ będącym skończonym uporządkowanym zbiorem par wierzchołków zawartych w $V(D)$. Można więc zapisać graf skierowany jako:

$$D = (V(D), A(D)) \quad (2.2)$$

W celu zobrazowania dodatkowych cech krawędzi między poszczególnymi wierzchołkami, stosuje się grafy ważone, w których każda krawędź posiada przypisaną wartość liczbową – wagę (Wilson 1998).

W grafie ważonym krawędziowo zbiór tworzący graf jest rozszerzony o funkcję $\delta : E \cup A \rightarrow K$ taką, że dla każdej krawędzi $e \in E$, $\delta(e)$ jest wagą danej krawędzi.

2.2.2 Trawersowanie grafów

W celu określenia metod przechodzenia między wierzchołkami grafów wyróżnione zostały metody nazywane Trasami, ścieżkami oraz drogami. (Ross i Wright 2005)

- Trasa – Skończony ciąg par krawędzi postaci

$$e_s, e_e \quad (2.3)$$

gdzie e_s oznacza krawędź początkową a e_e krawędź końcową, w którym e_s i e_e są sąsiednie lub identyczne. Można ją również oznać jako ciąg wierzchołków:

$$v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_m \quad (2.4)$$

gdzie v_0 nazywamy wierzchołkiem początkowym a v_m wierzchołkiem końcowym.

- Droga – Wyznaczona przez krawędzie trasa pomiędzy unikalnymi wierzchołkami, w której podróż między wierzchołkami wykonywana jest po łączących je krawędziach.
- Ścieżka – Trasa wyznaczana przez wierzchołki, w której wszystkie krawędzie są od siebie różne.

2.3 Analiza sieci społecznościowych

Analiza sieci społecznościowych to wywodząca się z zarządzania dziedzina nauki skupiająca się na definiowaniu zależności oraz procesów poprzez zastosowanie metod macierzowych, teorii grafów oraz analizy strukturalnej na obiektach zwanych sieciami, które przedstawiają relacje zachodzące pomiędzy poszczególnymi ich elementami (Batorski i Zdziarski 2009).

Analiza sieciowa skupia się na określaniu właściwości powiązań między punktami (węzłami) w sieci. Umożliwia poszukiwanie powiązań między zachowaniami ludzkimi a np. organizacjami, tematami lub podejmowanymi decyzjami. Podejście sieciowe umożliwia modelowanie, analizę oraz wizualizację skomplikowanych procesów, których struktura uzależniona jest od poprzedzających je akcji.

Analogicznie do przedstawionych założeń teorii grafów Analiza Sieci Społecznościowych definiuje i opisuje sieci za pomocą zbioru aktorów oraz zachodzących między nimi relacji.

Analiza istnienia relacji wykorzystuje strukturę standardowego, nieskierowanego grafu, w którym obiektem badania jest istnienie relacji (krawędzi) między aktorami (węzłami), przyczyną jej istnienia oraz jej siły.

W celu przedstawienia bardziej zaawansowanych sieci wykorzystywane są zasady budowy grafów skierowanych. Nadanie kierunku relacji umożliwia określenie relacji jednostronnych oraz odwzajemnionych.

Analiza sieci społecznościowych umożliwia również określenie między innymi:

- Stopnia odwzajemnienia relacji;
- Najkrótszej drogi w sieci;
- Średnicy sieci mierzonej jako największy dystans między parą węzłów;
- Centralności – miary wpływu "uprzywilejowanych" węzłów na resztę sieci;
- Grupowości – ilości, struktury oraz warunków przez które poszczególne komponenty sieci ulegają skupieniu.

2.4 Sieci temporalne

Sieci temporalne definiowane są jako sieci, których krawędzie oznaczone są przez następujące chronologicznie interaktywne wydarzenia pomiędzy węzłami. Można więc zapisać je w sposób:

$$G = (V(G), E(G), A(G)) \quad (2.5)$$

gdzie analogicznie do definicji przedstawionej w rozdziale 2.2.1 $V(G)$ oznacza zbiór węzłów, $E(G)$ zbiór krawędzi a zbiór $A(G)$ oznacza zbiór wydarzeń. Każda krawędź pomiędzy węzłami określona jest przez wydarzenia uporządkowane chronologicznie:

$$a_0 \rightarrow a_1 \rightarrow a_2 \rightarrow \dots \in A(G) \quad (2.6)$$

gdzie a_i oznacza wydarzenie w momencie czasu t_i .(Zuo i in. 2018)

Sieci temporalne wykorzystywane są do modelowania sytuacji, w których zdarzenie bezpośrednio poprzedzające inne zdarzenie oraz czynniki na nie wpływające posiada bezpośredni wpływ na czas zajścia oraz parametry zdarzenia po nim następującego.

Przykładami wykorzystania Sieci Temporalnych oraz metod modelowania przez nie oferowanych są między innymi (Holme i Saramäki 2012):

- Sieci komunikacji między jednostkami;
- Sieci bliskości fizycznej wykorzystywane podczas badań rozprzestrzeniania patogenów;
- Badania dotyczące biologii komórkowej;
- Sieci neuronowe reprezentujące działanie autonomicznego oprogramowania oraz odwzorowujące funkcjonowanie ludzkiego mózgu;
- Grupowości – ilości, struktury oraz warunków przez które poszczególne komponenty sieci ulegają skupieniu.

Jedną z metod stosowanych przy analizie sieci temporalnych jest Process Hawkesa, w którym punktowo analizowane wydarzenia przedstawione w czasie dyskretnym przed momentem t mogą wpływać na wystąpienie aktualnego wydarzenia. W metodzie tej stosowana jest warunkowa funkcja intensywności (2.7) opisująca szybkość nadejścia zdarzeń sekwencyjnych, która może zostać określona jako ilość zdarzeń zachodzących w krótkim okresie czasu $[t, t + \Delta t]$ bio-

rąc pod uwagę wszystkie zdarzenia historyczne $H(t)$.

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{E[N(t + \Delta t|H_t)]}{\Delta t} \quad (2.7)$$

2.5 Implementacja w języku Python

2.5.1 NetworkX Graph

Biblioteka NetworkX Graph to dostępna na licencji BSD biblioteka zawierająca narzędzia ułatwiające budowę, analizę oraz manipulacje dużych sieci. Pierwsza ogólnodostępna wersja ukazała się w roku 2005. Na oficjalnej stronie projektu funkcjonalność NetworkX Graph definiowana jest następująco (w tłumaczeniu autora):

NetworkX jest dedykowanym Pythonowi pakietem służącym do kreacji, manipulacji oraz badania struktury, dynamiki oraz funkcji złożonych sieci

Oferowana funkcjonalność oparta jest na strukturze słowników, zawartych w słownikach co czyni ją szybką, efektywną oraz wysoce skalową. Biblioteka umożliwia pracę na grafach o ilości węzłów oraz krawędzi ograniczonej przez możliwości urządzenia.

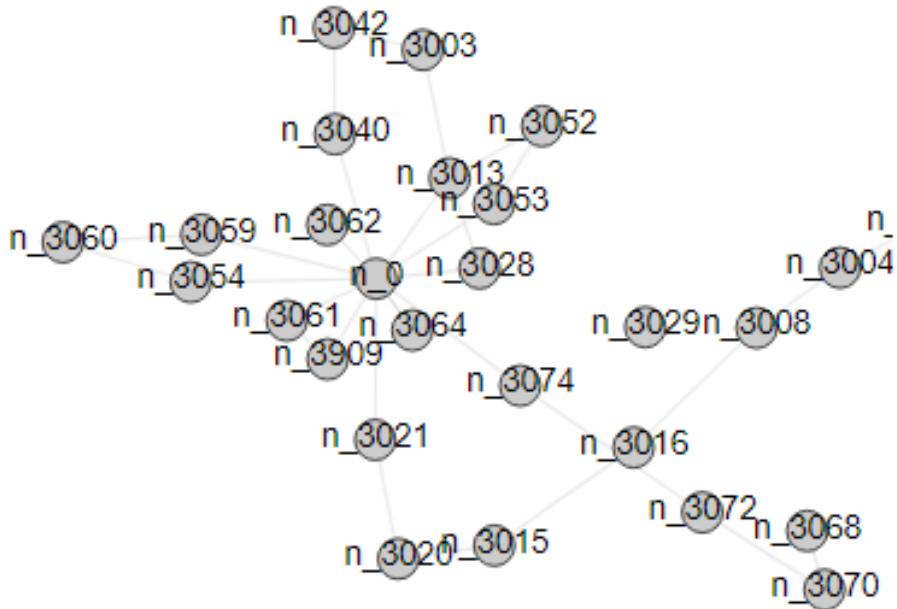
NetworkX umożliwia tworzenie prostych grafów, oraz budowanie złożonych, skierowanych grafów w których węzły mogą być dowolnym indeksowalnym obiektem, a krawędzie mogą przechowywać różne parametry. Zbiór węzłów oraz krawędzi przechowywany jest oddzielnie, co umożliwia wykonywanie operacji na ograniczonych zbiorach elementów.

Wizualizacja sieci i grafów tworzonych za pomocą biblioteki NetworkX dokonywana jest za pomocą metod zawartych w bibliotece MatPlotLib, wysoce znanym oraz poważanym projektem umożliwiającym tworzenie dokładnych wykresów oraz wizualizacji danych. Wizualizacja umożliwia eksploracyjną analizę danych przedstawionych już w ustrukturyzowanej postaci.

Na potrzeby pracy zaadaptowane zostały założenia skierowanego grafu skierowanego z wykorzystaniem klasy DiGraph zaczerpniętej z wyżej opisanej biblioteki.

2.5.2 Pathpy2

Pathpy2 jest to przeznaczona dla Pythona biblioteka, podobna do NetworkX, umożliwiająca analizę danych uszeregowanych czasowo w postaci sieci wyższego rzędu. Stworzona została na Uniwersytecie w Zuryczu oraz Wuppertal w 2018 roku. Udostępniana jest na licencji



Rysunek 2.1. Siatka w momencie $t = 1$, w którym niewidoczne są akcje

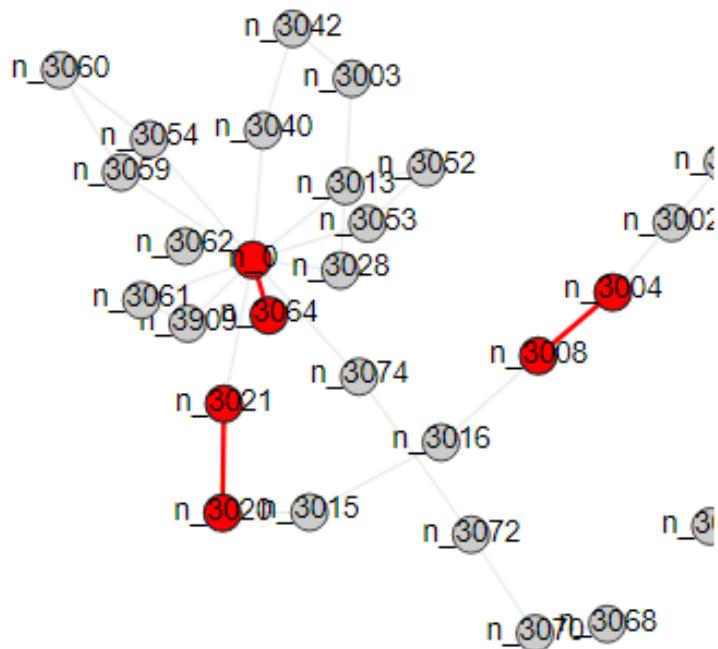
Źródło: Opracowanie własne.

GNU. (<https://www.pathpy.net>) Bardzo dobrze sprawdza się w przypadkach modelowania sieci temporalnych, oraz sieci w których powstają niezależne, uwarunkowane czasowo ścieżki, które wymagają modelowania relacji przyczynowo skutkowych (<https://pypi.org/project/pathpy2/>).

W odróżnieniu od NetworkX, pathpy umożliwia pracowanie na sieciach ulegająccych zmianie w czasie. Dzięki temu istnieje możliwość szerszej analizy zbiorów, wyszczególnienia ścieżek o określonej długości na bazie różnic odstępu czasu między zaistnieniem dwóch kolejnych zdarzeń (odpowiedników krawędzi łączących węzły).

Pathpy2 jest biblioteką oferującą interaktywne wizualizacje, o szerokich możliwościach ich dostosowania. Ograniczona jest jednak możliwość modyfikacji atrybutów węzłów oraz krawędzi a co za tym idzie, utrudnione jest budowanie statycznych wizualizacji reprezentujących zdarzenia w obiektach lub przestrzeniach.

Obie wymienione biblioteki potrafią ze sobą współpracować. Stworzony za pomocą NetworkX graf możemy załadować jako źródło danych do sieci PyPath, oraz na podstawie wag lub stempli czasowych dogłębić ją przeanalizować lub zwizualizować.



Rysunek 2.2. Siatka z widocznymi zajściami akcji

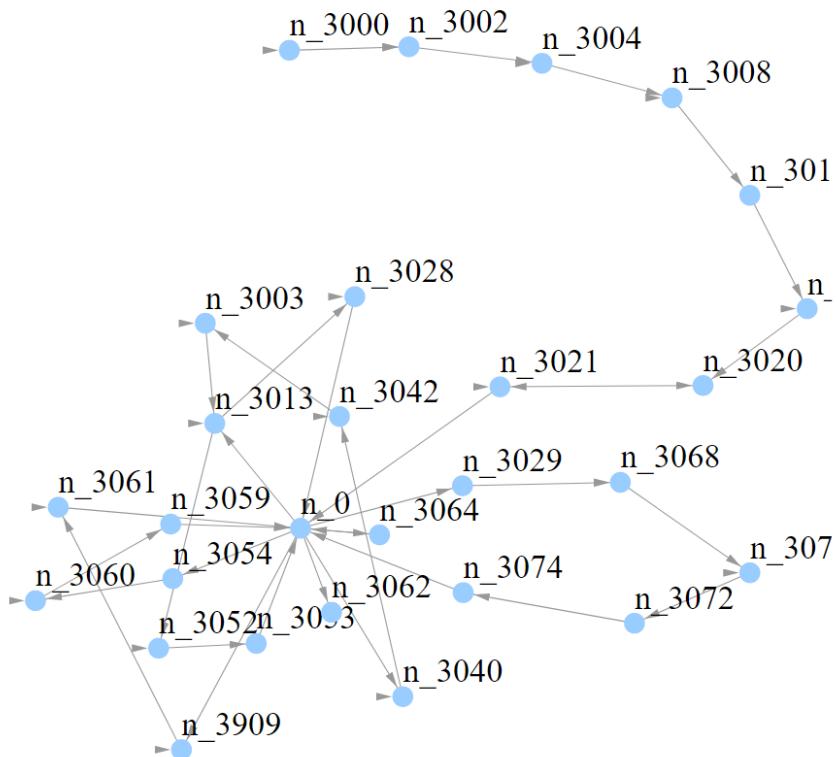
Źródło: Opracowanie własne.

Przykładowo, Pypath umożliwia budowę wizualizacji ulegających zmianie w czasie, na których reprezentowane są momenty uruchomienia węzłów oraz krawędzi oraz przedstawienie wygenerowanej po symulacji aktywacji kolejnych węzłów i krawędzi sieci statycznej będącej swego rodzaju siatką zbudowaną na podstawie danych.

2.5.3 Pandas

Pandas jest biblioteką umożliwiającą analizę oraz dokonywanie zmian na danych udostępnioną na licencji BSD. Powstała ona w roku 2008 w AQR Capital Management w wyniku zapotrzebowania na oprogramowanie umożliwiające szybką analizę finansową. Jego twórca, Wes McKinney w roku 2009 wynegocjował z AQR wprowadzenie otwartego licencjonowania projektu. W roku 2015 Pandas stał się projektem finansowanym przez NumFOCUS.

Podstawową i najbardziej znaną funkcjonalnością biblioteki jest budowa Ramek Danych (z ang. *DataFrame*) umożliwiających szybkie dokonywanie zmian oraz obliczeń na zindeksowanych danych w nich zawartych.



Rysunek 2.3. Sieć statyczna stworzona na bazie dynamicznej sieci temporalnej

Źródło: Opracowanie własne .

Wprowadzenie tabelarycznego przedstawienia danych oraz możliwości intuitywnego wykonywania na nich działań, takich jak wyświetlanie warunkowe, klasyfikacja danych itp skutkowało rosnącą popularnością tej biblioteki wśród użytkowników Pythona. Wraz z czym format Ramek Danych stawał się coraz bardziej rozpoznawalny. Doprowadziło to do budowy zależności między klasami wywodzącymi się z Pandas a innymi bibliotekami takimi jak Matplotlib umożliwiającymi wizualizację danych zawartych w Ramkach Danych. Zakres funkcjonalności oferowanej przez bibliotekę obejmuje wczytywanie oraz zapis danych, ich analizę, restrukturyzacje, grupowanie oraz wiele innych. Wszystkie operacje wykonywać można zarówno na danych uszeregowanych czasowo jak i statycznych. Dodatkową zaletą jest szybkość wykonywanych operacji, które zostały zoptymalizowane pod względem wykorzystania Ramek Danych w środowisku Python. Oferują one ogromny zakres ilości możliwie przechowywanych danych w połączeniu z szybkim ich przetwarzaniem oraz wyświetlaniem.

| plec | wiek | KLASA_MIEJSCEWOSCI | WOJEWODZTWO | eduk4_2011 | lata_nauki_2011 | fc4 | fc11 | fc37 | fd4 | |
|------|------|--------------------|-------------|------------|-----------------|-----|------|------|-----|----|
| 0 | 1 | 48 | 1 | 2 | 3 | 14 | 1 | 4 | 11 | 40 |
| 1 | 1 | 55 | 1 | 2 | 2 | 11 | 7 | 2 | 12 | 35 |
| 2 | 1 | 22 | 1 | 2 | 2 | 9 | 4 | 1 | 12 | 40 |
| 3 | 1 | 36 | 1 | 2 | 3 | 12 | 1 | 1 | 12 | 40 |
| 4 | 1 | 34 | 1 | 2 | 4 | 17 | 3 | 1 | 11 | 20 |

Rysunek 2.4. Przykład Ramki Danych

Źródło: Opracowanie własne.

2.6 Podsumowanie

Istnieje wiele metod analizy danych dotyczących działań sekwencyjnych. Niektóre odnajdują zastosowanie w zarządzaniu, chemii, biologii oraz wielu innych dziedzinach nauki (Holme i Saramäki 2012). Opisane metody oraz narzędzia posłużą analizie posiadanych danych. Teoria grafów oraz podstawowe założenia Analizy Sieci Społecznych w połączeniu z biblioteką NetworkX zostanie wykorzystana podczas tworzenia wizualizacji sieci przejść pomiędzy kolejnymi elementami wystawy muzelanej. Pandas zostanie wykorzystany w procesach manipulacji danych. W następnym rozdziale, nieprzygotowane jeszcze dane uzyskane od muzeum *Bramy Poznania* zostaną poddane oczyszczeniu oraz restrukturyzacji. Celem tych działań jest umożliwienie wyodrębnienia poszczególnych sesji zwiedzania oraz danych ogólnych je opisujących.

Rozdział 3

Analiza ścieżek odwiedzających Bramę Poznania

3.1 Pozyskanie i analiza danych z audioprzewodników

Wszystkie dane wykorzystane w tej pracy dostarczone zostały przez pracowników muzeum Bramy Poznania w postaci logów zapisanych w formacie XML. Zbiór zawiera zapisy działań z 50 audioprzewodników, zebranych na przedziale od 10 kwietnia 2016 roku do 23 marca 2018 roku.

```
<?xml version="1.0"?>
- <STATYSTYKA>
  - <PRZEDZIAŁ_CZASU>
    <OD>2012-01-01 21:12:46</OD>
    <DO>2019-11-26 21:12:47</DO>
  </PRZEDZIAŁ_CZASU>
  - <WYBRANE_AUDIOPRZEWODNIKI>
    <NAZWA>1000;AUDIOGUIDE 1000 -</NAZWA>
  </WYBRANE_AUDIOPRZEWODNIKI>
  - <SZCZEGÓLowe>
    - <ZAAWANSOWANE>
      <ILOŚĆ_WPISÓW>195557</ILOŚĆ_WPISÓW>
      <DATA_CZAS_AKCJA>2016-04-10*08:40:24*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*08:40:24*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*08:48:32*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*08:50:00*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*11:41:04*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:02:04*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:25:20*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:26:08*Wybór języka; PL</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:26:08*Rozpoczęcie Nagrania; 65020, 255</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:26:20*Zakończenie nagrania; 65020, 255</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:26:48*Wybór ścieżki L1</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:26:52*Rozpoczęcie Nagrania; 65020, 191</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:16*Rozpoczęcie Nagrania; 3000, 3</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:16*Przycisk - przewin w przód</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:16*Rozpoczęcie Nagrania; 3000, 4</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:16*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:24*Włączenie urządzenia</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:24*Rozpoczęcie Nagrania; 65020, 255</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:27:36*Zakończenie nagrania; 65020, 255</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:31:16*Wybór ścieżki L0</DATA_CZAS_AKCJA>
      <DATA_CZAS_AKCJA>2016-04-10*12:31:16*Rozpoczęcie Nagrania; 65020, 12</DATA_CZAS_AKCJA>
```

Rysunek 3.1. Zrzut ekranu z formatem danych dla jednego z audioprzewodników

Źródło: Opracowanie własne.

W celu ułatwienia pracy nad danymi zostały one przekonwertowane do postaci plików z wartościami oddzielonymi przecinkami (Comma Separated Values; CSV) oraz zapisane w formacie: *plik audioprzewodnika, data, godzina, akcja, strefa, nagranie*. Na tym etapie liczebność zbioru wynosiła **8 461 509 wpisów**.

Dane zostały przeanalizowane podczas wstępnych oględzin podczas których zidentyfikowane zostały problemy z ich wiarygodnością oraz strukturą. Pierwszą trudnością było określenie przedziału czasowego z jakiego zebrane dane powinny być analizowane. Pomimo iż zapisy obejmują zakres dwóch lat, zawierają one również dane dla dni podczas których muzeum nie było dostępne dla zwiedzających.

Przykładem takich zapisów są kontrolne przejścia pracowników muzeum podczas których kontrolowane jest działanie wystawy prowadzonej i jej elementów. Zaobserwowane również zostało cykliczne uruchamianie się i wyłączanie urządzeń bez podjęcia żadnej akcji (3.2).

```
1000.xml,2016-04-10,08:40:24,Włączenie urządzenia,NA,NA
1000.xml,2016-04-10,08:40:24,Wyłączenie urządzenia,NA,NA
1000.xml,2016-04-10,08:48:32,Włączenie urządzenia,NA,NA
1000.xml,2016-04-10,08:50:00,Wyłączenie urządzenia,NA,NA
1000.xml,2016-04-10,11:41:04,Włączenie urządzenia,NA,NA
1000.xml,2016-04-10,12:02:04,Wyłączenie urządzenia,NA,NA
1000.xml,2016-04-10,12:25:20,Włączenie urządzenia,NA,NA
```

Rysunek 3.2. Cykliczne restartowanie się urządzeń

Źródło: Opracowanie własne.

Zachowanie to zostało potwierdzone z pracownikami muzeum jako skutek długiego ładowania urządzenia, podczas którego naładowanie w pełni urządzenia powoduje jego uruchomienie. Między niektórymi elementami zbioru nie dochodziło do zachowania struktury, co mogłoby prowadzić do utrudnienia ich analizy. Przykładowo na rysunku 3.3 widoczne jest wielokrotne rozpoczęcie nagrań o indeksie 2 w strefach 3008, 3016 oraz 3015. Zarejestrowane zostało zaakończenie jedynie nagrania 2 w strefie 3015.

Ostatnim problemem było wyodrębnienie sesji zwiedzania, dla poszczególnych użytkowników. Urządzenia zapisywały swoje działanie w sposób ciągły, oraz często nie były wyłączone między kolejnymi zwiedzającymi. Takie zachowanie audioprzewodników wymaga poszukiwania rozwiązania umożliwiającego przybliżenie i wyodrębnienie takich sesji za pomocą warunków logicznych.

```
1000.xml,2016-04-10,12:33:00,Zakończenie nagrania, 3004, 13  
1000.xml,2016-04-10,12:33:00,Rozpoczęcie Nagrania, 3008, 2  
1000.xml,2016-04-10,12:33:00,Rozpoczęcie Nagrania, 3008, 2  
1000.xml,2016-04-10,12:34:28,Rozpoczęcie Nagrania, 3016, 2  
1000.xml,2016-04-10,12:35:40,Rozpoczęcie Nagrania, 3015, 2  
1000.xml,2016-04-10,12:35:52,Zakończenie nagrania, 3015, 2
```

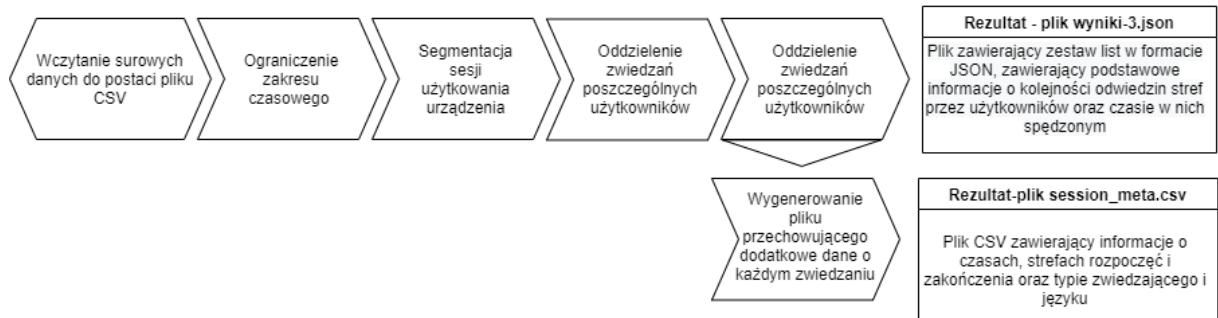
Rysunek 3.3. Przykład niedokładności zapisu danych

Źródło: Opracowanie własne.

Po przeglądzie, dane zostały załadowane do środowiska PyCharm oraz Jupyter w celu ich dalszego oczyszczania i analizy.

3.2 Proces czyszczenia danych

Zobrazowany na Rysunku 3.4 został schemat procesu wykorzystanego w celu uzyskania dodatkowych informacji z danych oraz ich oczyszczania ma na celu ułatwienie dalszej pracy oraz zaproponowanie rozwiązań problemów opisanych w sekcji 3.1.



Rysunek 3.4. Diagram poszczególnych etapów przygotowania danych

Źródło: Opracowanie własne.

3.2.1 Ograniczenie okresu czasowego

Aby zwiększyć wiarygodność danych, konieczne było odrzucenie wszystkich aktywności zachodzących w nocy oraz dni, w których muzeum jest zamknięte dla zwiedzających.

W tym celu, dane zostały wczytane do postaci Ramki Danych za pomocą biblioteki pandas a następnie ograniczone do wartości zawartych w odpowiednich ramach czasowych poprzez wykorzystanie metod zawartych w bibliotece `DateTime`.

Ponieważ muzeum jest zamknięte dla zwiedzających w poniedziałki, wartości te zostały odrzucone w całości. W rezultacie, liczelnosc wierszy danych została zmniejszona do 5 755 383.

Na rysunku 3.7 zobrazowane zostały uzyskane różnice między zbiorami. Jako wartość osi X obrane zostały poszczególne godziny dnia (hour). Oś Y przedstawia sumaryczną liczbę wykonanych akcji w danym dniu o pewnej godzinie. Porównując oba rysunki można zauważyć, że pomimo niedostępności muzeum w poniedziałki, na rysunku 3.5 wciąż obecne były wartości dla tego dnia. Założono, że są to wartości fałszywe, wygenerowane przez pracowników muzeum oraz opisany wcześniej mechanizm samouruchamiania urządzeń.

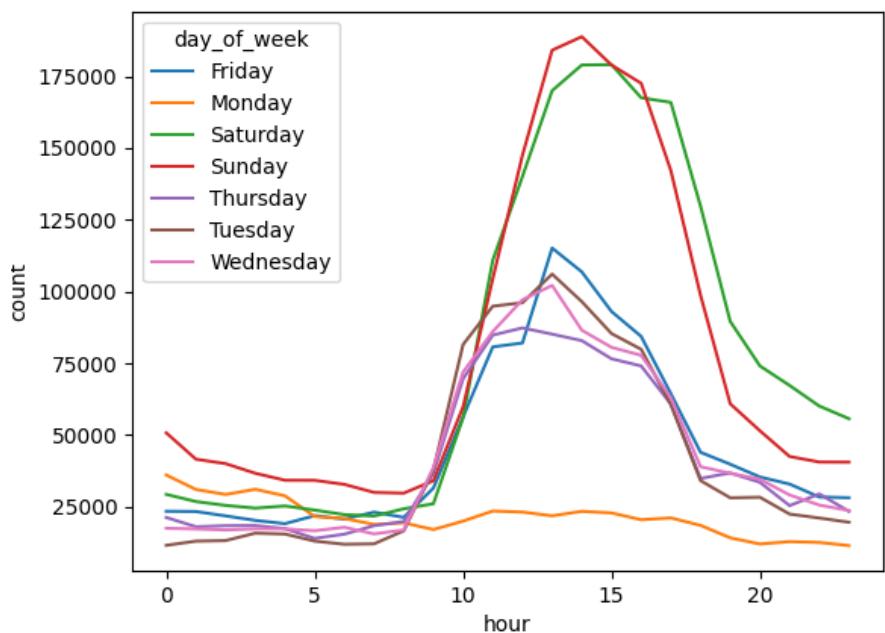
Na przedstawionym wykresie 3.6 wyraźnie zauważalna jest dominacja Soboty i Niedzieli w popularności zwiedzania, oraz dynamiczny wzrost aktywności w piątki między godziną 12 a 15. Wartości dla dni tygodni ograniczone zostały do godzin funkcjonowania muzeum. Aby uwzględnić w danych wydarzenia takie jak Noc Muzeum, wartości dla Soboty oraz Niedzieli nie zostały ograniczone względem godzin. Aby umożliwić prace nad danymi zarówno w perspektywie całościowej, jak również przy podziale na poszczególne sesje lub ich grupy, konieczne jest wydzielenie z danych reprezentujących wszystkie podjęte akcje w danym dniu danych o unikatowych zwiedzaniach.

3.2.2 Segmentacja sesji działania urządzeń

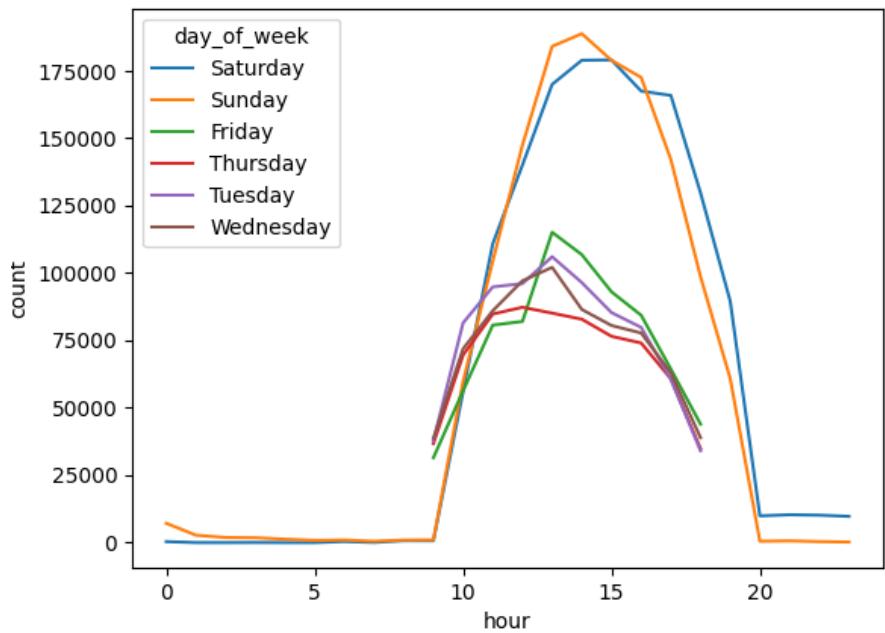
Na tym etapie dane przedstawiają wciąż wszystkie podejmowane akcje takie jak przyciszańie, pogłaszczenie urządzeń itp. Wśród nich znajdują się dane dotyczące wybranej przez użytkownika ścieżki językowej, które posłużą nam jako baza do określenia narodowości oraz wieku zwiedzającego. Dane przetwarzane są według schematu opisanego na Rysunku 3.8.

Struktura danych po operacjach przedstawionych na Rysunku 3.8 wzbogacone zostają o kolumnę zawierającą numer sesji działania, kolumnę język, grupa (przyjmującą wartości binarne) oraz wiek. Jego struktura prezentuje się następująco: *numer urządzenia, godzina akcji, akcja, strefa, nagranie, język, wiek, grupa*.

Wartość nagranie przyjmuje dla stref w których odtwarzane jest automatyczne nagranie dźwiękowe wartość równą numerowi tego nagrania. Dla stref uruchamianych przez użytkownika za pomocą urządzenia, przyjmuje ona wartość 'RF TAG'. Ze względu na wysoką zależność



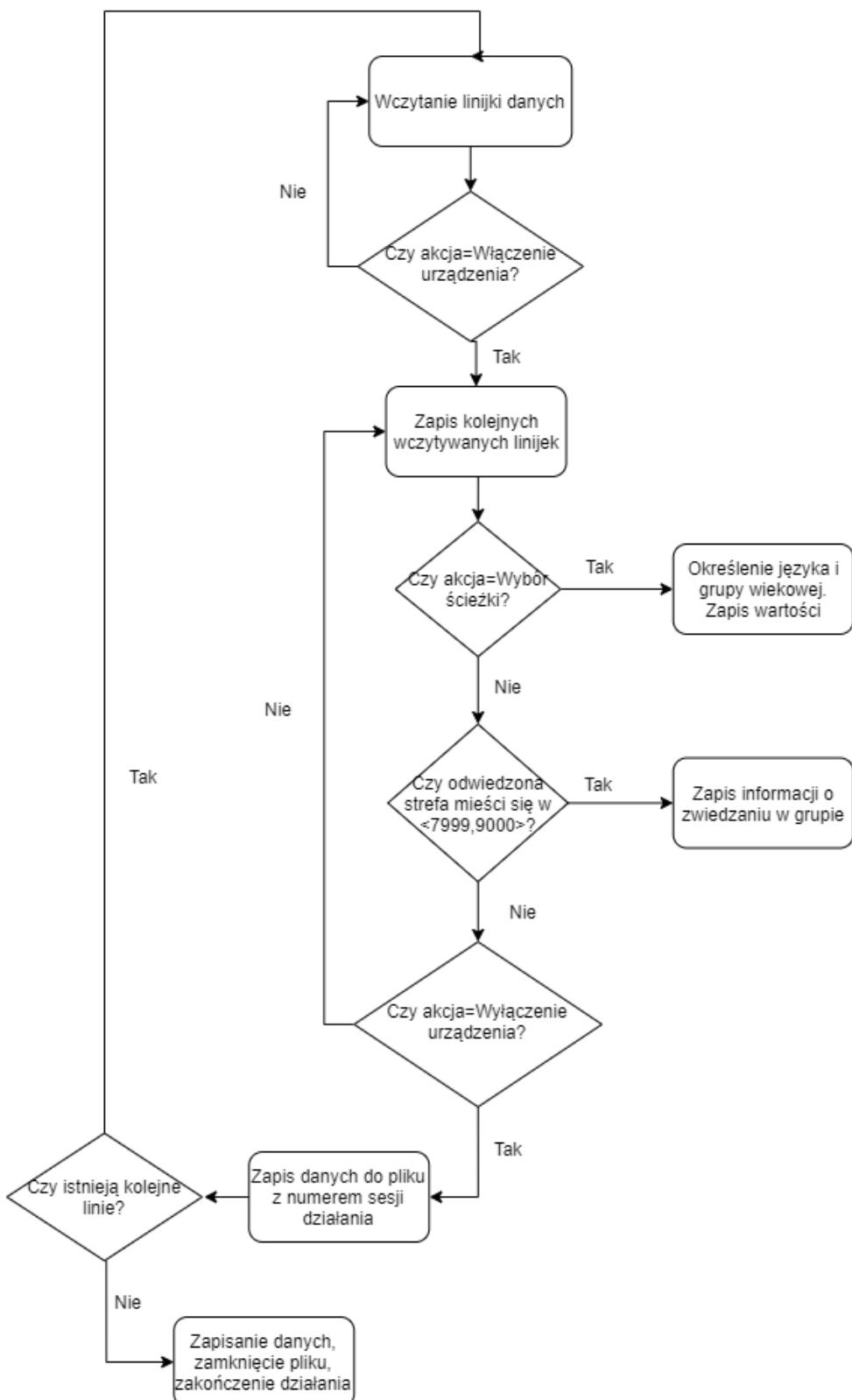
Rysunek 3.5. przed oczyszczeniem



Rysunek 3.6. po oczyszczeniu

Rysunek 3.7. Rezultaty ograniczenia okresu czasowego dla danych

Źródło: Opracowanie własne.



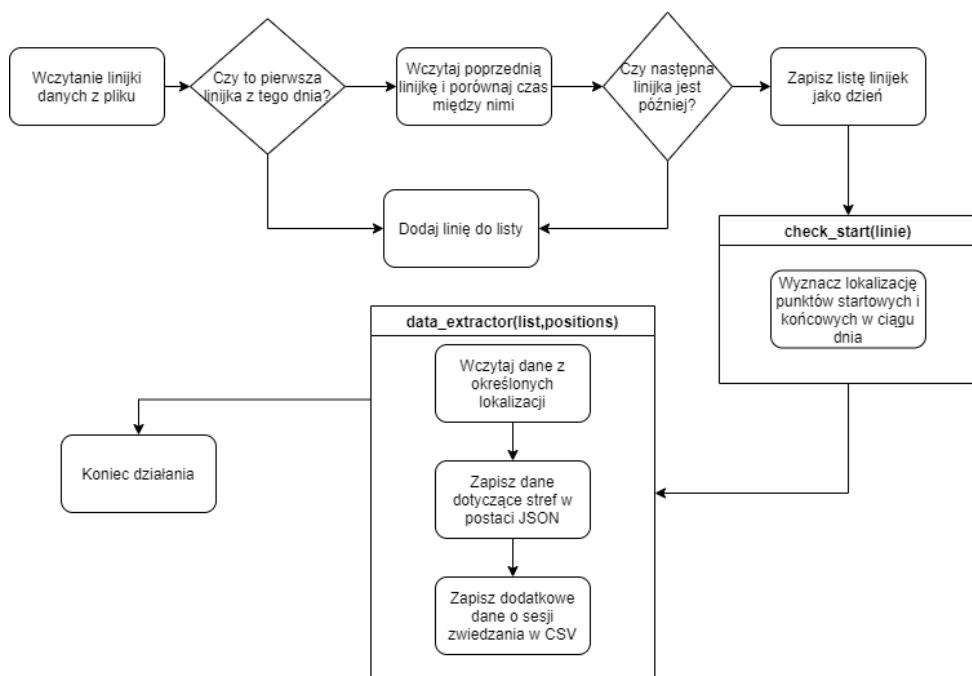
Rysunek 3.8. Segmentacja sesji działania urządzenia

Źródło: Opracowanie własne.

wiarygodności sesji użytkowania urządzenia od zachowania zwiedzającego, konieczna jest dalsza separacja sesji opisana w następnym podrozdziale. Wygenerowany na tym etapie plik zawierający dodatkowe dane o sesjach działania urządzenia zawiera 30182 sesje aktywności. Jego struktura prezentuje się następująco: *numer audioprzewodnika, data, godzina rozpoczęcia, godzina zakończenia, język, wiek, grupa, liczba akcji*.

3.2.3 Segmentacja sesji użytkowników

Wstępnie wydzielony plik jest następnie segmentowany ponownie, zgodnie ze schematem 3.9.



Rysunek 3.9. Segmentacja sesji użytkowników

Źródło: Opracowanie własne.

Dane na podstawie kolejności godzin dzielone są na fragmenty obejmujące poszczególne dni, z których następnie ekstraktowane są punkty rozpoczęcia oraz zakończenia sesji zwiedzania. Aby określić momenty rozpoczęcia oraz zakończenia, każda sesja trawersowana jest do momentu uzyskania wystarczającej ilości danych umożliwiających stwierdzenie czy sesja jest kompletna, rozpoczęła się w początku trasy, była efektem restartu urządzenia oraz gdzie została zakończona. W ten sposób, uzyskane wartości powinny uwzględniać możliwość ponownego uruchomienia urządzenia i kontynuowania zwiedzania, oraz wcześniejszego jego wyłączenia.

Założenia logiczne i schemat poszukiwania momentów początku i końca sesji przedstawione są na Rysunku 3.10.

Z pomocą określonych w ten sposób wartości rozpoczęcia oraz zakończenia sesji, zbiór dzielony jest na części obejmujące informacje zawarte w każdej z sesji, a następnie zapisuje je do dwóch plików.

Zapis danych o kolejności przechodzonych stref dokonywany jest według schematu pokazanego na Rysunku 3.11 do pliku zawierającego listę danych w formacie JSON o następującej strukturze o nazwie *wynik-3.json*:

```
sesja = {  
    'numer_urzadzenia': nr_urzadzenia,  
    'numer_sesji':numer_sesji_uzycowania,  
    'sesje_dzialania':sesje_dzialania,  
    'liczba_akcji': len(temp_actions),  
    'sciezki': lista_stref  
}
```

Program 3.1. Struktura danych w formacie JSON

gdzie lista stref przyjmuje format listy wielu 4 elementowych list formatu: [*numer strefy, numer nagrania, typ nagrania ('nagranie' lub 'RF TAG'), czas w sekundach od poprzedniej akcji*]. Zapis danych opisujących poszczególne sesje zwiedzania dokonywany jest do pliku CSV, o formacie zobrazowanym na rysunku 3.12.

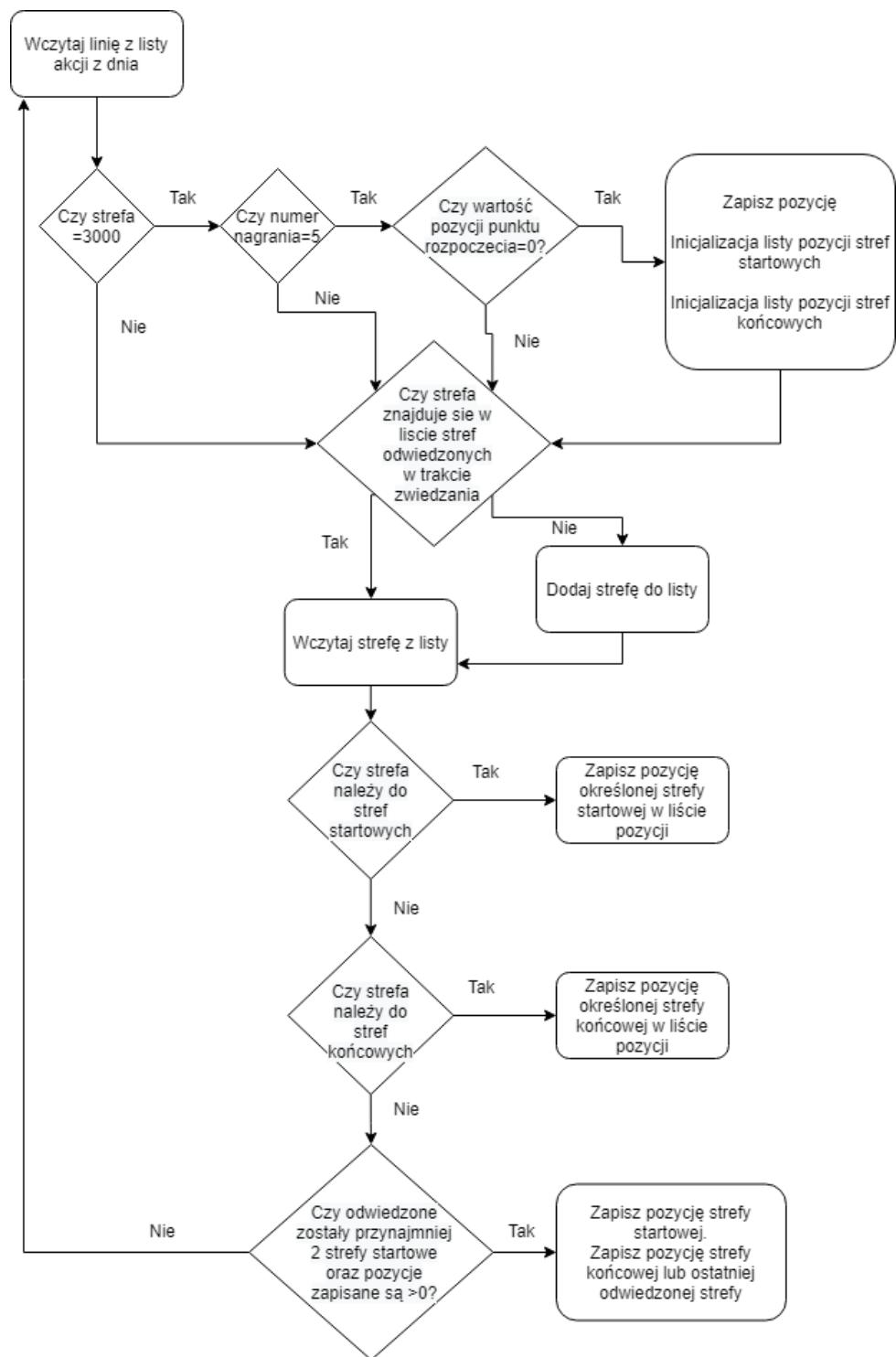
W powyżej opisany sposób określonych zostało 17592 unikatowych sesji zwiedzania, czyli średnio prawie 352 sesje na każde urządzenie w badanym okresie.

3.3 Statystyki zwiedzających (średni czas trwania, opis zwiedzających osób)

3.3.1 Czas zwiedzania

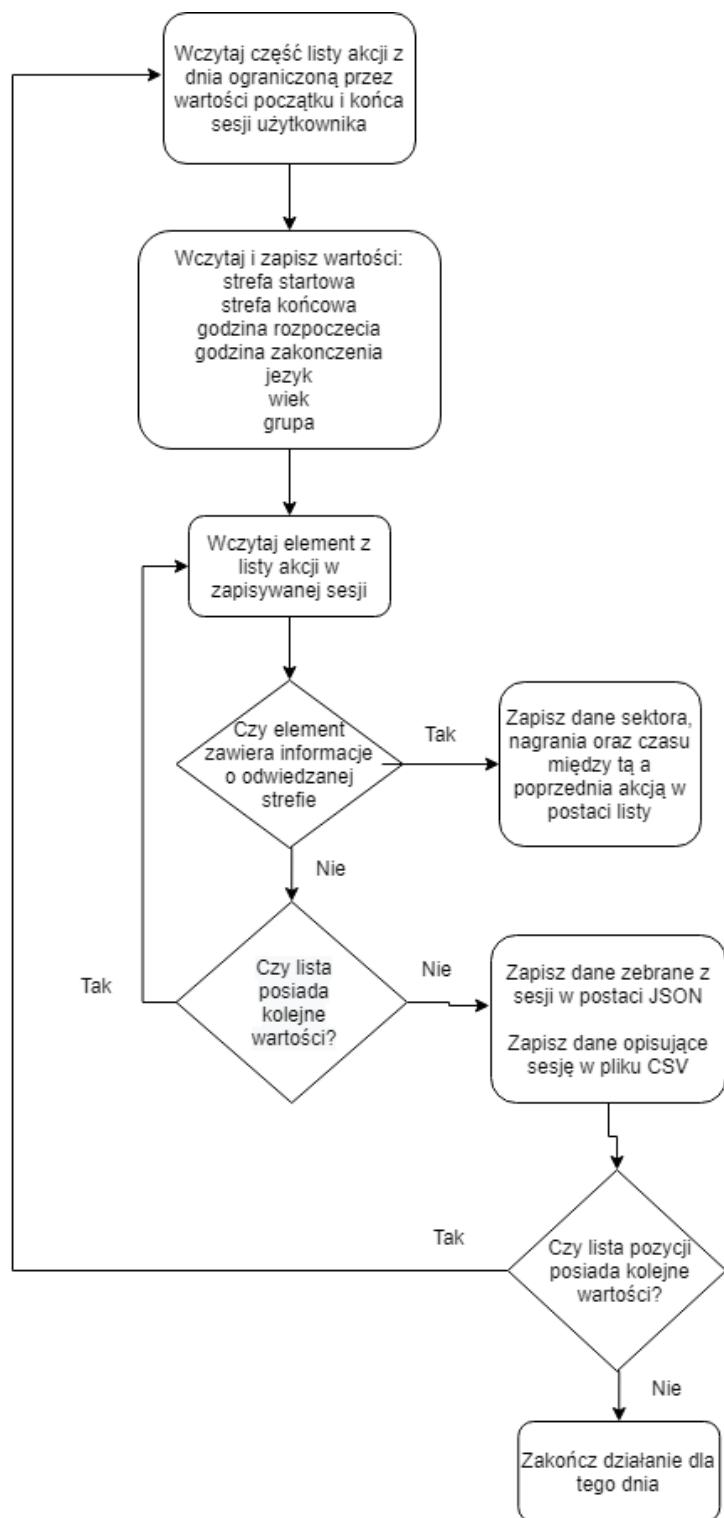
Wykorzystując dane zawarte w pliku zobrazowanym na Rysunku 3.12 poprzez obliczenie różnicy między godziną rozpoczęcia oraz zakończenia i ograniczeniu danych do sesji trwających więcej niż 3 minuty i mniej niż 5 godzin otrzymano histogram zobrazowany na rysunku 3.13. Ograniczenie zostało wprowadzone w celu wykluczenia wartości krótkich przejść, testów urządzeń przez pracowników w trakcie godzin otwarcia muzeum:

```
meta['godz_zak'] = pd.to_datetime(meta['godz_zak'],format= '%H:%M:%S')  
meta['godz_rozp'] = pd.to_datetime(meta['godz_rozp'],format= '%H:%M:%S')  
meta['czas_trwania']=meta['godz_zak']-meta['godz_rozp']
```



Rysunek 3.10. Poszukiwanie momentów kluczowych sesji - funkcja `check_start(linia)`

Źródło: Opracowanie własne.



Rysunek 3.11. Ekstrakcja i zapis danych finalnych

Źródło: Opracowanie własne.

| | numer_urzadzenia | numer_sesji | godz_rozp | godz_zak | strefa_rozpoczenia | strefa_zakonczenia | grupa_jazykowa | grupa_wiekowa | grupa | liczba_akcji |
|-------|------------------|-------------|-----------|----------|--------------------|--------------------|----------------|---------------|-------|--------------|
| 0 | 1000 | 1 | 12:31:36 | 12:59:00 | 3000.0 | 3062.0 | polski | dorośli | 0 | 99 |
| 1 | 1000 | 2 | 12:59:56 | 13:32:44 | 3000.0 | 3.0 | polski | dorośli | 0 | 11 |
| 2 | 1000 | 3 | 13:32:44 | 13:34:48 | 3000.0 | 3002.0 | polski | dorośli | 0 | 7 |
| 3 | 1000 | 4 | 13:34:52 | 14:58:48 | 3000.0 | 3074.0 | polski | dorośli | 0 | 228 |
| 4 | 1000 | 5 | 13:45:40 | 14:11:24 | 3000.0 | 3000.0 | polski | dzieci | 0 | 129 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 17587 | 1049 | 17588 | 16:32:32 | 17:35:56 | 3000.0 | 3000.0 | polski | dzieci | 0 | 303 |
| 17588 | 1049 | 17589 | 15:27:36 | 15:29:28 | 3000.0 | 3002.0 | polski | dorośli | 0 | 7 |
| 17589 | 1049 | 17590 | 15:29:32 | 16:07:48 | 3000.0 | 3062.0 | polski | dorośli | 0 | 126 |
| 17590 | 1049 | 17591 | 17:16:12 | 18:20:08 | 3000.0 | 3000.0 | polski | dorośli | 0 | 242 |
| 17591 | 1049 | 17592 | 19:07:56 | 19:09:36 | 3000.0 | 3002.0 | polski | dorośli | 0 | 12 |

17592 rows × 10 columns

Rysunek 3.12. Struktura danych dodatkowych

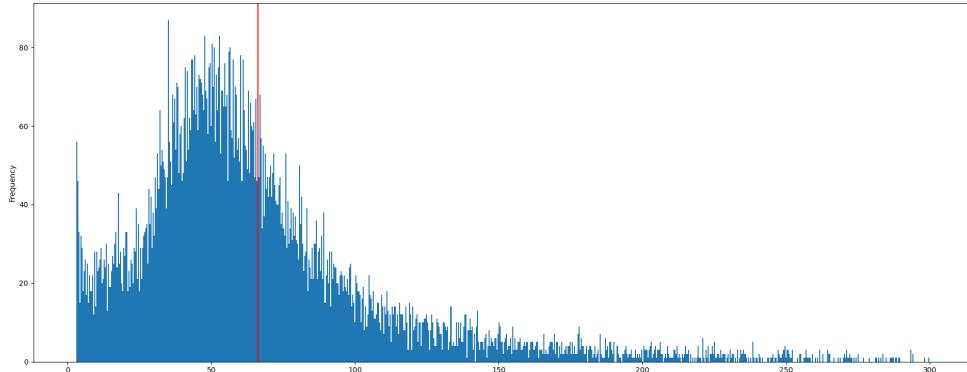
Źródło: Opracowanie własne.

```

meta['godz_zak']=meta['godz_zak'].dt.time
meta['godz_rozp']=meta['godz_rozp'].dt.time
from datetime import timedelta
d=timedelta(minutes=3)
l=timedelta(hours=5)
meta = meta.drop(meta[meta['czas_trwania']<d].index)
meta = meta.drop(meta[meta['czas_trwania']>l].index)

```

Program 3.2. Przygotowanie danych dotyczących długości czasu zwiedzania.



Rysunek 3.13. Struktura czasu zwiedzania w minutach

Źródło: Opracowanie własne.

Na osi X przedstawione zostały wartości długości zwiedzania wyrażone w minutach. Na rysunku 3.13, pionową czerwoną linią zaznaczona została średnia wartość długości czasu zwiedzania, wynosząca 66.28 minut. Na wykresie wyraźnie widoczna jest asymetria lewostronna,

co informuje o przeważającej ilości zwiedzających, których czas zwiedzania jest krótszy od średniej. Przeważająca większość wartości znajduje się w przedziale od 25 minut do 75 minut czasu zwiedzania.

3.3.2 Charakterystyka grup zwiedzających muzeum

Na podstawie przygotowanej w rozdziale 3.3.1 Ramki Danych dokonano zgrupowania wartości w celu przedstawienia struktury grup językowo wiekowych wśród osób zwiedzających muzeum Bramy Poznania. W tym celu, wartości zostały wpierw pogrupowane za pomocą funkcji `DataFrame.groupby()`:

```
meta_grupy=meta.groupby(['grupa_językowa', 'grupa_wiekowa', 'grupa'])
meta_grupy.count()
```

Program 3.3. Wyodrębnienie liczebności poszczególnych typów osób zwiedzających

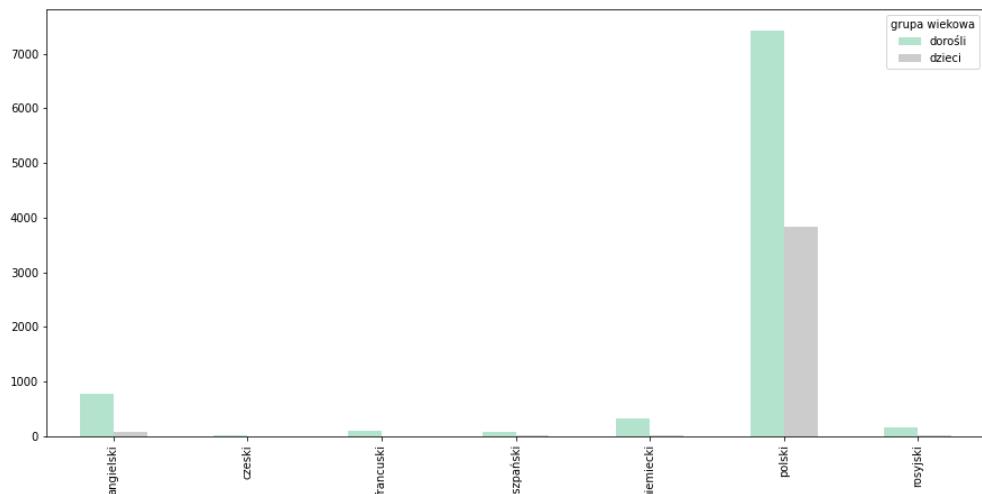
Otrzymany w ten sposób obiekt przedstawiony na Rysunku 3.14 (wizualizowany na 3.15) zawierał liczebności dla każdej kombinacji wybieranych języków oraz wieku.

| grupa_językowa | grupa_wiekowa | grupa | numer_urzadzenia | numer_sesji | godz_rozp | godz_zak | strefa_rozpoczecia | strefa_zakonczenia | liczba_akcji | czas_trwania |
|----------------|---------------|-------|------------------|-------------|-----------|----------|--------------------|--------------------|--------------|--------------|
| angielski | dorośli | 0 | 771 | 771 | 771 | 771 | 771 | 709 | 771 | 771 |
| | | 1 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 |
| | dzieci | 0 | 78 | 78 | 78 | 78 | 78 | 74 | 78 | 78 |
| czeski | dorośli | 0 | 25 | 25 | 25 | 25 | 25 | 23 | 25 | 25 |
| | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| francuski | dorośli | 0 | 107 | 107 | 107 | 107 | 107 | 97 | 107 | 107 |
| | | 1 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| hiszpański | dorośli | 0 | 92 | 92 | 92 | 92 | 92 | 80 | 92 | 92 |
| | | 1 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| niemiecki | dorośli | 0 | 326 | 326 | 326 | 326 | 326 | 290 | 326 | 326 |
| | | 1 | 21 | 21 | 21 | 21 | 21 | 20 | 21 | 21 |
| polski | dorośli | 0 | 7403 | 7403 | 7403 | 7403 | 7403 | 6795 | 7403 | 7403 |
| | | 1 | 27 | 27 | 27 | 27 | 27 | 26 | 27 | 27 |
| | dzieci | 0 | 3834 | 3834 | 3834 | 3834 | 3834 | 3658 | 3834 | 3834 |
| | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| rosyjski | dorośli | 0 | 173 | 173 | 173 | 173 | 173 | 162 | 173 | 173 |
| | | 1 | 19 | 19 | 19 | 19 | 19 | 18 | 19 | 19 |

Rysunek 3.14. Ramka danych zawierająca informacje o typach zwiedzających

Źródło: Opracowanie własne.

Najczęściej występującą wartością było zwiedzanie w języku polskim dla dorosłych, takich sesji odbyło się ponad 7000. Drugą najczęściej wybieraną kombinacją było zwiedzanie dla dzieci w języku polskim, drugim najbardziej popularnym językiem był język angielski. Pozostałe języki posiadają stosunkowo niskie liczebności.



Rysunek 3.15. Wykres typów zwiedzających

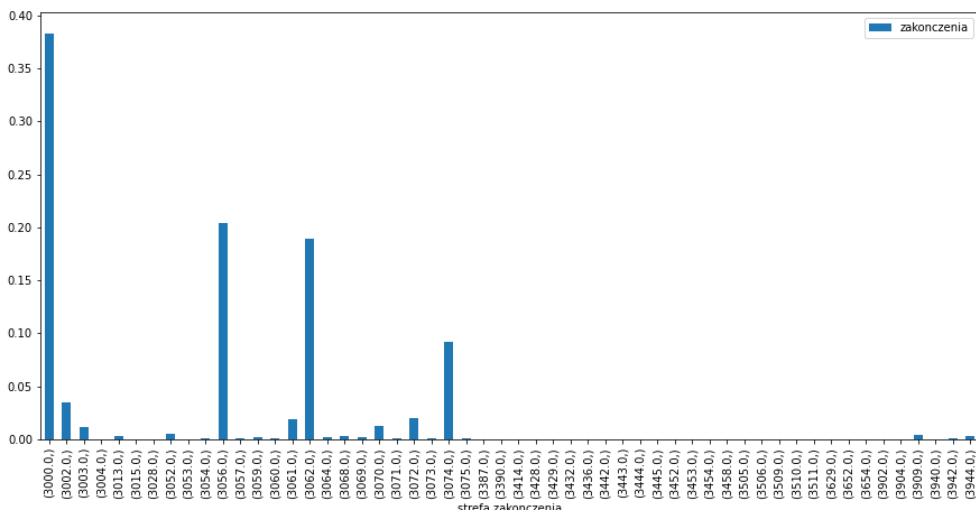
Źródło: Opracowanie własne.

3.3.3 Stopień wykorzystywania audioprzewodników

Aby przybliżyć stopień w jakim zwiedzający wykorzystuje audioprzewodnik w trakcie trasy, stworzony został wykres stopnia wykorzystania na podstawie danych ze zbioru danych opisujących zwiedzania przedstawiony na rysunku 3.16.

Wysoki wynik strefy 3000, będącej strefą początkową jest skutkiem posiadania przez nią jednego nagrania kończącego zwiedzanie, odtwarzającego się na schodach prowadzących do wyjścia. Zakończone w ten sposób zwiedzania, można uznać więc za pełne.

Pozostałe trzy główne strefy na których kończyło się zwiedzanie to strefa 3056, strefa 3062 oraz strefa 3074, które znajdują się odpowiednio u góry schodów prowadzących do wyjścia, przy wejściu do przedostatniego pomieszczenia na drugim piętrze oraz w korytarzu kończącym cykl zwiedzania, po którym można zejść do wyjścia, lub udać się na taras. Można więc założyć, że około trzy czwarte zwiedzających spędza wizytację ekspozycji używając czynnie swoich urządzeń. Taki rozkład świadczy o dość wysokim stopniu wykorzystania audioprzewodników. Użytkownicy cenią sobie narrację, którą zapewniają oraz nie zaprzestają korzystania z nich zbyt wcześnie.



Rysunek 3.16. Stopień wykorzystania audioprzewodników, poprzez określenia stref zakończeń zwiedzań

Źródło: Opracowanie własne.

3.4 Przygotowanie struktury grafu atrakcji muzealnych

3.4.1 Zbiór przejść - krawędzie grafu skierowanego ważonego

Za pomocą funkcji z biblioteki json, dane o poszczególnych sesjach zwiedzania wczytane zostały do formatu Ramki Danych postaci zaprezentowanej na Rysunku 3.17. Następnie, dane dla każdego zwiedzania przekształcone zostały w postać zawierającą strefę wyjściową oraz wejściową (gdzie strefa wyjściowa, to strefa opuszczana przez zwiedzającego, a wejściowa, to strefa po niej następująca), wraz z czasem spędzonym w strefie wyjściowej. Na tym etapie odrzucone zostały strefy systemowe, strefy zwiedzania grupowego, oraz strefy, które nie znajdowały się w muzeum. W ten sposób zebrane pary stref zostały zliczone, czego wynikiem była tabela liczebności "przejść" pomiędzy nimi (Rysunek 3.18). Ponieważ w audioprzewodnikach funkcjonuje system przewań powodujący zmianę wartości strefy pod warunkiem odwiedzenia innej strefy wartości zostały zmienione zgodnie z informacjami dostarczonymi od pracowników. Przykładowo, po odwiedzeniu jednej ze stref 3042, 3044, 3420 lub 3383 strefa o numerze 3040 odczytywana jest jako strefa 3041. Podczas tworzenia grafu przejść zamierzonym sposobem identyfikacji węzłów jest nazwanie ich numerem strefy, w celu zachowania spójności wszystkie strefy na których wywołane zostało przezwanie zostają zmienione na ich pierwotną wartość.

| | numer_urzadzenia | numer_sesji | sesje_dzialania | liczba_akcji | sciezki |
|-------|------------------|-------------|-----------------|--------------|---|
| 0 | 1000 | 1 | [3] | 99 | [[3000, 5, nagranie, 0], [3000, 5, nagranie, 2... |
| 1 | 1000 | 2 | [3] | 11 | [[3000, 5, nagranie, 0], [3000, 6, nagranie, 2... |
| 2 | 1000 | 3 | [3] | 7 | [[3000, 5, nagranie, 0], [3000, 6, nagranie, 2... |
| 3 | 1000 | 4 | [3] | 228 | [[3000, 5, nagranie, 0], [3000, 5, nagranie, 2... |
| 4 | 1000 | 5 | [7] | 129 | [[3000, 5, nagranie, 0], [3000, 5, nagranie, 2... |
| ... | ... | ... | ... | ... | ... |
| 17587 | 1049 | 17588 | [63048] | 303 | [[3000, 5, nagranie, 0], [3000, 6, nagranie, 2... |
| 17588 | 1049 | 17589 | [63056] | 7 | [[3000, 5, nagranie, 0], [3000, 6, nagranie, 2... |
| 17589 | 1049 | 17590 | [63056] | 126 | [[3000, 5, nagranie, 0], [3002, 2, nagranie, 2... |
| 17590 | 1049 | 17591 | [63066] | 242 | [[3000, 5, nagranie, 0], [3000, 5, nagranie, 2... |
| 17591 | 1049 | 17592 | [63067] | 12 | [[3000, 5, nagranie, 0], [3000, 5, nagranie, 2... |

Rysunek 3.17. Dane dotyczące stref odwiedzonych podczas sesji

Źródło: Opracowanie własne.

| | 0 | 1 | size |
|------|------|------|-------|
| 0 | 3000 | 3000 | 49090 |
| 1 | 3000 | 3001 | 2177 |
| 2 | 3000 | 3002 | 8174 |
| 3 | 3000 | 3003 | 231 |
| 4 | 3000 | 3004 | 1195 |
| ... | ... | ... | ... |
| 6644 | 3950 | 3621 | 6 |
| 6645 | 3950 | 3650 | 2 |
| 6646 | 3950 | 3655 | 5 |
| 6647 | 3950 | 3900 | 26 |
| 6648 | 3950 | 3950 | 10842 |

6649 rows × 3 columns

Rysunek 3.18. Tabela liczebności "przejść"

Źródło: Opracowanie własne.

Poprzez zachowanie kolejności oraz kierunku zapisanego w parach stref, wizualny efekt przedstawiony na grafie pozostanie niezmienny, a wartości liczebności krawędzi będą odzwierciedlać wszystkie rodzaje przejść. Liczebności każdego przejścia posłużą do określenia wagi krawędzi odzwierciedlającej częstość obierania danej ścieżki. Widoczna jednak na Rysunku 3.18 dysproporcja względem najwyższych wartości oraz wartości najniższych stworzyła konieczność normalizowania wartości. W tym celu zastosowana została funkcja `log()` znajdująca się w bibliotece `numpy`, dokonująca normalizacji na sposób opisany w rozdziale 2. Tak przygotowane wartości zapisane zostały w postaci trzyelementowych krotek formatu:

(strefa wyjścia, strefa wejścia, waga krawędzi)

1

Program 3.4. Struktura krotki określającej krawędź grafu.

Przygotowane w ten sposób wartości posłużą do zobrazowania krawędzi skierowanego grafu ważonego przedstawiającego trasę obierane przez użytkowników audioprzewodników.

3.4.2 Zbiór stref – węzły grafu

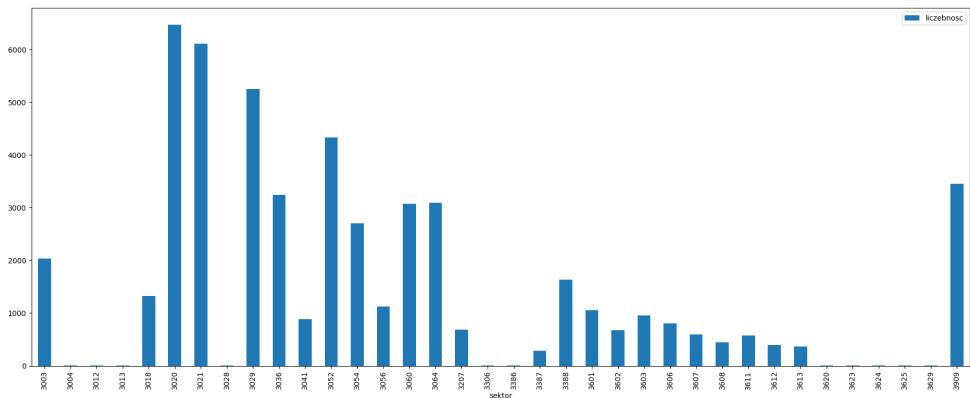
Kolejnym etapem przygotowania danych było przygotowanie zbioru węzłów z lub do których wykonywane były przejścia. Otrzymany zbiór wszystkich unikalnych stref posłuży do identyfikacji węzłów:

```
uniqueValues = (przejscia_skierowane[0].append(przejscia_skierowane[1])).1
unique()2
3 Unikalne strefy wszystkich przejść4
5 [3000 3001 3002 3004 3005 3006 3007 3008 3009 3010 3011 3012 3014 3016
6 3019 3020 3028 3036 3040 3042 3044 3052 3053 3054 3056 3060 3063 3064
7 3068 3070 3072 3074 3075 3201 3202 3203 3205 3207 3208 3209 3210 3211
8 3212 3301 3302 3303 3304 3305 3306 3307 3308 3309 3310 3311 3312 3313
9 3314 3315 3316 3334 3335 3622 3381 3621 3384 3385 3386 3387 3389 3390
10 3391 3401 3402 3403 3404 3405 3406 3407 3408 3409 3410 3411 3412 3413
11 3414 3415 3416 3417 3418 3419 3420 3421 3422 3423 3424 3425 3426 3427
12 3428 3429 3430 3431 3432 3433 3434 3435 3436 3437 3438 3439 3440 3441
13 3442 3443 3444 3445 3446 3447 3448 3449 3450 3451 3452 3453 3454 3455
14 3456 3457 3458 3471 3474 3501 3502 3503 3504 3505 3506 3507 3508 3509
15 3510 3511 3600 3601 3602 3603 3605 3606 3607 3608 3610 3611 3612 3613
16 3620 3623 3624 3625 3626 3629 3650 3651 3652 3653 3654 3655 3656 3657
3658 3659 3900 3902 3904 3909 3940 3942 3944 3950]
```

Program 3.5. Wyznaczenie unikalnych stref znajdujących się w danych

Widoczna jest niezgodność między ilością stref wyznaczonych tym sposobem, oraz stref, które zostały oznaczone na mapie muzeum. Zostały utworzone dwa dodatkowe zbiory danych. Zbiór stref znajdujących się na mapie, gdzie każdej z nich nadana została również jej pozycja na wykresie grafu oraz zbiór stref, w których dokonywana była aktywacja czujnika RF wraz z ich

liczebnościami. Struktura zbioru stref w których zaszła aktywacja dodatkowej funkcjonalności ekspozycji przedstawiona została na rysunku 3.19.



Rysunek 3.19. Wykres częstości wykorzystania stref zawierających dodatkowe informacje

3.5 Wizualizacja ścieżek obieranych przez zwiedzających

Do wizualizacji ścieżek zwiedzania wykorzystana została biblioteka NXGraph, umożliwiająca rysowanie skierowanych grafów ważonych z dużych zbiorów danych. Wszystkie unikalne wartości stref odwiedzanych zostały załadowane do wcześniej zainicjalizowanego obiektu grafu:

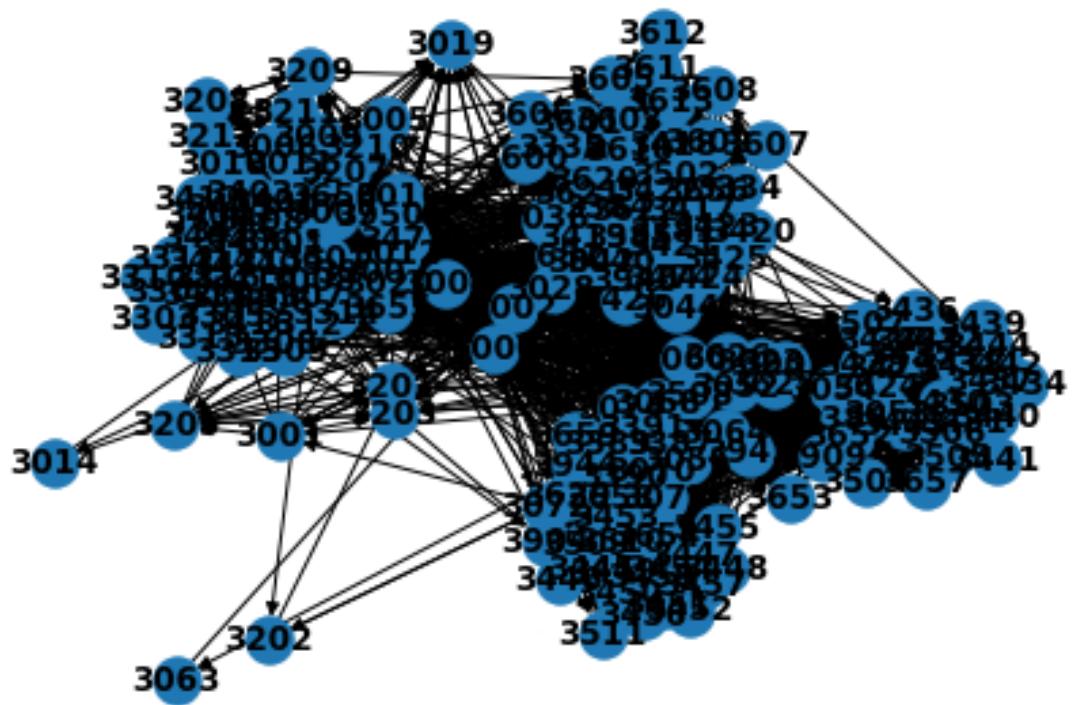
```
G = nx.DiGraph()
G.add_nodes_from(uniqueValues)
print(list(G.nodes))
G.add_weighted_edges_from(list(przejscia_skierowane.to_records(index=False)))
nx.draw(G, with_labels=True, font_weight='bold')
```

Program 3.6. Inicjalizacja i budowa pierwszej formy grafu.

Tak przygotowany graf, przedstawiony na Rysunku 3.20 jest nieczytelny. Widoczne są jednak zgrupowania stref oraz istnienie skierowanych krawędzi między nimi. Aby podnieść widoczność informacji zawartych na grafie, zestawienie map obu pięter zostało załadowane jako tło, a następnie za pomocą wcześniej określonych wartości strefy, których pozycja była znana, zostały umieszczone na grafie jako statyczne. Pozostałe strefy układają się na grafie zgodnie z wartościami obliczonymi przez funkcję:

```
pos = nx.spring_layout(G, pos=pozycje, fixed = fixed_nodes)
```

Program 3.7. Funkcja ustalająca pozycje węzłów



Rysunek 3.20. Pierwsza forma grafu

Źródło: Opracowanie własne.

Która określa pozycję węzłów na podstawie ilości oraz wadze jego powiązań z innymi węzłami. Dla każdej krawędzi z jej krotki wczytana została wartość jej wagi, wpływająca na szerokość linii łączącej dwa węzły. Kod wykorzystany do stworzenia drugiej wersji grafu:

```

G = nx.DiGraph()
G.add_nodes_from(uniqueValues)
print(list(G.nodes))
G.add_weighted_edges_from(list(przejscia_skierowane.to_records(index=False)))
    )
pos = nx.spring_layout(G, pos=pozycje, fixed = fixed_nodes)
edge_widths = [w for (*edge, w) in G.edges.data('weight')]
print(edge_widths)
fig, ax = plt.subplots()
fig.set_size_inches(26, 21)
axes = plt.gca()
y_min, y_max = axes.get_ylim()
print(y_min, y_max)
ax.imshow(img)
nx.draw(G, pos, width=edge_widths, connectionstyle='arc3, rad=.15',
       with_labels=True, font_weight='bold',
       node_size=40,
       font_size=12, edge_color='Green')
plt.show()

```

Program 3.8. Inicjalizacja pierwszego grafu naniesionego na mapę

Uzyskany w ten sposób graf (Rysunek 3.21 przedstawał wszystkie strefy, oraz ważone powiązania między nimi, jednak wciąż nie był wystarczająco czytelny. Obrany schemat kolorystyczny utrudniał rozróżnienie wartości poszczególnych krawędzi a duża ilość stref uniemożliwiała ich identyfikację.

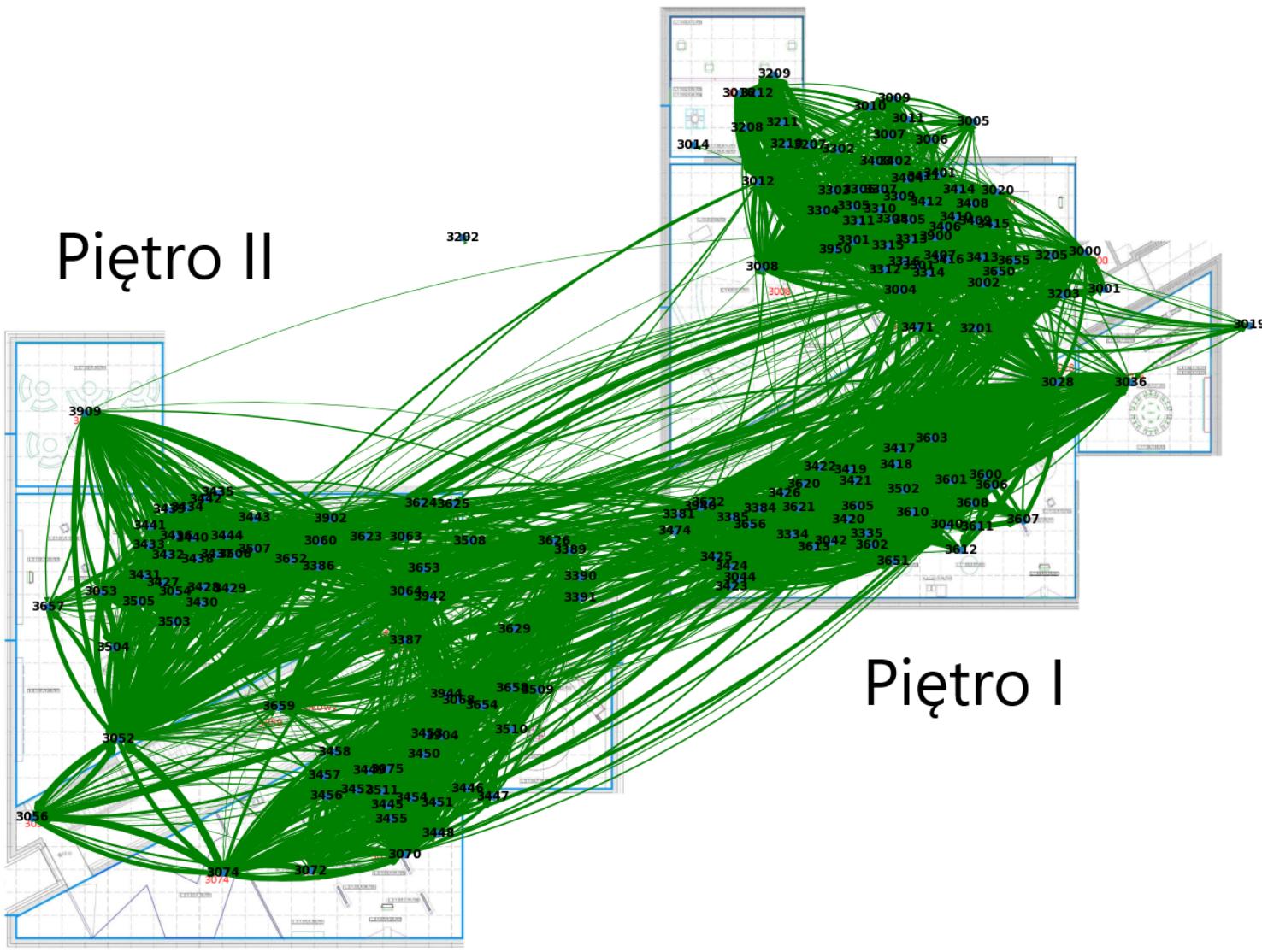
W celu ułatwienia rozróżnienia stref, zbiór służący do rysowania grafu pozostał niezmieniony, aby zapewnić rysowanie wszystkich krawędzi między strefami. Po utworzeniu grafu, usunięte zostały strefy niebędące częścią grupy stref w której zwiedzający mogą skorzystać z czujników RF oraz grupy stref naniesionych na mapę. Wartości kolorów zostały dodatkowo uzależnione od wartości znormalizowanej za pomocą metody min-max wagi krawędzi, dzięki czemu najważniejsze połączenia zostały wyszczególnione. Kod wykorzystany do stworzenia finalnej wersji grafu (Rysunek 3.22):

```

G = nx.DiGraph()
G.add_nodes_from(uniqueValues)
print(list(G.nodes))
G.add_weighted_edges_from(list(przejscia_skierowane.to_records(index=False)))
    )
pos = nx.spring_layout(G, pos=pozycje, fixed = fixed_nodes)
G.remove_nodes_from(surplus)
edge_widths = [w for (*edge, w) in G.edges.data('weight')]
edges, weights = zip(*nx.get_edge_attributes(G, 'weight').items())
print(edge_widths)
weights_norm = [float(i)/max(weights) for i in weights]
fig, ax = plt.subplots()
fig.set_size_inches(26, 21)
axes = plt.gca()

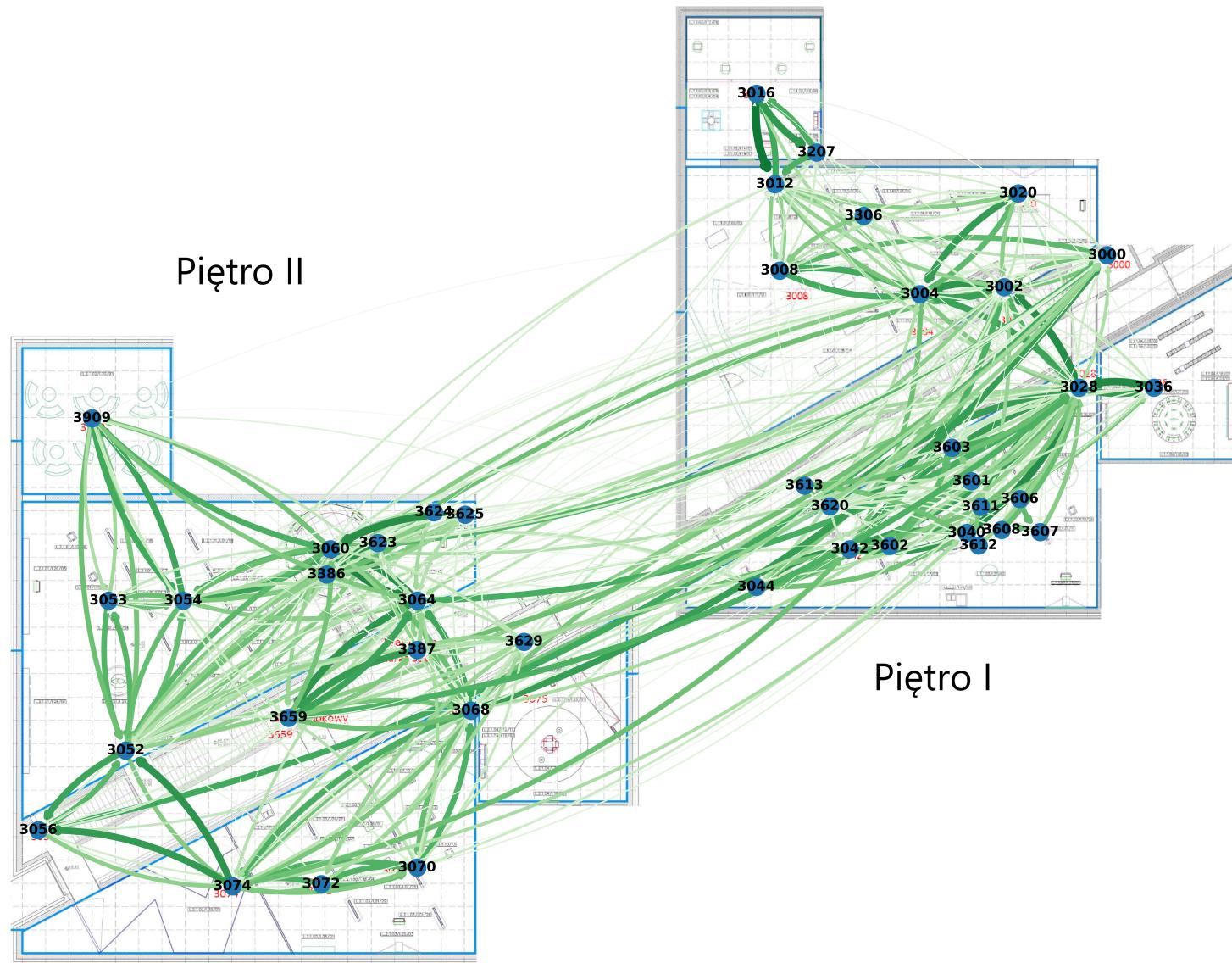
```

Piętro II



Rysunek 3.21. Graf naniesiony na mapę z wagami krawędzi

Źródło: Opracowanie własne.



Rysunek 3.22. Graf naniesiony na mapę z wagami krawędzi ograniczony do stref o znanym położeniu oraz aktywowanych

Źródło: Opracowanie własne.

```

y_min, y_max = axes.get_ylim()                                14
print(y_min, y_max)                                         15
ax.imshow(img)                                              16
nx.draw(G, pos, edgelist=edges, width=edge_widths, connectionstyle='arc3, 17
    rad=.15,
,with_labels=True, font_weight='bold', node_size=360,           18
font_size=15,edge_color=weights_norm,edge_cmap=plt.cm.Greens)
plt.show()                                                 19
fig.savefig('trocheczytelnygrafzrfid.png',dpi=300)          20
                                                               21

```

Program 3.9. Inicjalizacja naniesionego na mapę grafu o zwiększonej czytelności

W ten sposób rysowany graf jest czytelniejszy. Widoczne na grafie jest zróżnicowanie obieranych przez zwiedzających tras, oraz ogólne przedstawienie "głównej ścieżki".

3.6 Wnioski

Muzeum Bramy Poznania od początku swojego istnienia zakładało ścisłą współpracę z nowymi technologiami oraz zaprojektowane zostało aby dostarczać zwiedzającym zarówno rozrywki jak i wiedzy. Wysoki stopień retencji wykorzystania audioprzewodników, popularność muzeum oraz częste korzystanie przez zwiedzających z dodatkowych elementów wystawy utwierdza jego pozycję jako miejsca umożliwiającego poszukiwanie wiedzy. Graf prezentujący ścieżki zwiedzania wskazuje na wysoką indywidualność wśród zwiedzających, które wykorzystują główną trasę jako punkt odniesienia podczas odwiedzin muzeum. Wysoki stopień przyjęcia audioprzewodników i ich skuteczne działanie wskazuje na wysoką pozycję Bramy Poznania jako miejsca otwartego i oferującego wiedzę na miarę drugiego tysiąclecia.

Podsumowanie

Rozwój technologii Internetu rzeczy umożliwia wykorzystanie nowatorskich rozwiązań w muzeach na całym świecie. Placówki takie jak *Brama Poznania* są miejscami, gdzie technologia współpracuje z ekspozycją, umożliwiając zmianę tradycyjnego modelu zwiedzania w doświadczenie umożliwiające interaktywne poszukiwanie wiedzy przez zwiedzających.

Uzyskane z audioprzewodników logi są zbiorem nietypowym oraz unikalnym. Pochodzą bezpośrednio z badanej placówki, więc są wiarygodne w swoim przedstawieniu rzeczywistości muzealnej. Poprzez trudność dostępu, takie dane rzadko podlegają analizie i prezentacji.

Otrzymane dane sprawiały dużo trudności pod względem doprowadzenia ich do stanu umożliwiającego analizę. Zastosowane podejście było wystarczające do ekstrakcji chociaż części informacji zawartych w tym zbiorze.

Przeprowadzona analiza umożliwia wgląd w zachowanie zwiedzających, wskazując na wysoką indywidualność wśród zwiedzających oraz duży stopień wykorzystania audioprzewodników. Dzięki takim badaniom, istnieje możliwość zaproponowania ulepszeń oraz nowych rozwiązań jeszcze lepiej dopasowanych do charakterystyki profilu osoby zwiedzającej.

W przyszłości pracy nad danymi koniecznością będzie ulepszenie procesu selekcji danych a w szczególności sesji poszczególnych zwiedzających. Podniesienie wiarygodności danych w ten sposób uzyskanych jest kluczowe dla dalszej ich analizy.

Badanie przeprowadzone w tej pracy jest wstępnym spojrzeniem na informacje możliwe do uzyskania ze zbioru otrzymanego z *Bramy Poznania*. Do rozwoju przeprowadzonego badania warto byłoby zastosować analizę grupowości, skupienia oraz odstępstw od wyznaczonej najczęstszej trasy zwiedzania. Większa integracja z opisanymi metodami badań Sieci Społecznościowych oraz Sieci Temporalnych umożliwiłaby również wgląd w przyczynowość obieranych decyzji, oraz wyszczególnienie zbiorów obranych ścieżek i ich kategoryzację.

Przeprowadzone oczyszczenie danych, ekstrakcja ścieżek oraz ich analiza są więc pierwszym etapem badania, które daje nam wgląd w często nieznaną część cyfrowej rzeczywistości.

Bibliografia

- Ahlgren, B., Hidell, M. & Ngai, E. H. (2016). Internet of Things for Smart Cities: Interoperability and Open Data. *IEEE Internet Computing*, 20(06), 52–56. <https://doi.org/10.1109/MIC.2016.124>
- Ashton, K. (2009). That 'Internet of Things' Thing. *RFID Journal*, 22, 97–114.
- Barnaghi, P., Compton, M., Corcho, O., Castro, R. G., Graybeal, J., Herzog, A., Janowicz, K., Neuhaus, H., Nikolov, A. & Page, K. (2011). *Semantic Sensor Network XG Final Report*. World Wide Web Consortium. <http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/>
- Batorski, D. & Zdziarski, M. (2009). Analiza sieciowa i jej zastosowania w badaniach organizacji i zarządzania. *Problemy zarządzania*, 7, 157–184.
- Caragliu, A., Bo, C. D. & Nijkamp, P. (2011). Smart Cities in Europe. *Journal of Urban Technology*, 18(2), 65–82.
- Cayley, P. (1881). On the Analytical Forms Called Trees. *American Journal of Mathematics*, 4(1), 266–268. <http://www.jstor.org/stable/2369158>
- Chui, K., Liu, W., Lytras, M. & Zhao, M. (2019). Big data and IoT solution for patient behaviour monitoring. *Behaviour Information Technology*, 38, 1–10. <https://doi.org/10.1080/0144929X.2019.1584245>
- Decathlon. (2019). *Decathlon rewolucjonizuje zastosowanie RFID w Polsce!* Pobrane marzec 26, 2021 z <http://media.decathlon.pl/84427-decathlon-revolucionizuje-zastosowanie-rfid-w-polsce>
- Duhigg, C. (2019). *Is Amazon Unstoppable?* Pobrane marzec 26, 2021 z <https://www.newyorker.com/magazine/2019/10/21/is-amazon-unstoppable>
- Evans, D. (2011). How the Next Evolution of the Internet Is Changing Everything.

- Holme, P. & Saramäki, J. (2012). Temporal networks [Temporal Networks]. *Physics Reports*, 519(3), 97–125. <https://doi.org/https://doi.org/10.1016/j.physrep.2012.03.001>
- Ivudria, G. (2021). *Global Internet Of Things (IoT) Revenues To Jump By 53% And Hit \$677B By 2025*. Pobrane marzec 26, 2021 z <https://www.busiweek.com/global-internet-of-things-iot-revenues-to-jump-by-53-and-hit-677b-by-2025/#:~:text=The%5C%20government%5C%20ranked%5C%20second%5C%20with,to%5C%20%5C%2490.2bn%5C%20by%5C%202025>.
- Kiatkawsin, K., Sutherland, I. & Lee, S. K. (2020). Determinants of Smart Tourist Environmentally Responsible Behavior Using an Extended Norm-Activation Model. *Sustainability*, 12, 4934. <https://doi.org/10.3390/su12124934>
- L., C. A. (1813). Recherche sur les polyèdres – premier mémoire. *Mémoire sur la polyèdrométrie*, 1812–1813.
- Landt, J. (2005). The history of RFID. *IEEE POTENTIALS*, 24, 8–11.
- Lee, S. (2017). A review of audio guides in the era of smart tourism. *Information Systems Frontiers*, 19. <https://doi.org/10.1007/s10796-016-9666-6>
- Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P. & Sheth, A. P. (2018). Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks*, 4(3), 161–175. <https://www.sciencedirect.com/science/article/pii/S235286481730247X>
- Marketing Land. (2020). *Roughly 1 in 4 U.S. adults now owns a smart speaker, according to new report: It estimates there are now 157 million of these devices in American homes*. Pobrane marzec 26, 2021 z <https://marketingland.com/roughly-1-in-4-u-s-adults-now-owns-a-smart-speaker-according-to-new-report-273994>
- Naramski, M. (2020). The Application of ICT and Smart Technologies in Polish Museums—Towards Smart Tourism. *Sustainability*, 12(21). <https://www.mdpi.com/2071-1050/12/21/9287>
- Oteafy, S. M. A. & Hassanein, H. S. (2018). IoT in the Fog: A Roadmap for Data-Centric IoT Development. *IEEE Communications Magazine*, 56(3), 157–163. <https://doi.org/10.1109/MCOM.2018.1700299>
- Ross, K. A. & Wright, C. R. (2005). *Matematyka dyskretna*.

Smart Card Alliance. (2007). *Contactless Payments: Frequently Asked Questions: A Smart Card Alliance Contactless Payments Council Publication*. Pobrane marzec 26, 2021 z https://www.securetechalliance.org/resources/pdf/Contactless_Payments_FAQ.pdf

Tsai, C., Lai, C., Chiang, M. & Yang, L. T. (2014). Data Mining for Internet of Things: A Survey. *IEEE Communications Surveys Tutorials*, 16(1), 77–97. <https://doi.org/10.1109/SURV.2013.103013.00206>

Vanolo, A. (2014). Smartmentality: The Smart City as Disciplinary Strategy. *Urban Studies*, 51(5), 883–898. <https://doi.org/10.1177/0042098013494427>

Wilson, R. J. (1998).

Zuo, Y., Liu, G., Lin, H., Guo, J., Hu, X. & Wu, J. (2018). Embedding Temporal Network via Neighborhood Formation, 2857–2866. <https://doi.org/10.1145/3219819.3220054>

Spis rysunków

| | | |
|------|--|----|
| 1.1 | Retencja użytkowania audioprzewodników wykazana przez Lee (2017) | 11 |
| 1.2 | Rzut pionowy piętra pierwszego z zaznaczonymi strefami | 13 |
| 1.3 | Rzut pionowy piętra drugiego z zaznaczonymi strefami | 14 |
| 2.1 | Siatka w momencie $t = 1$, w którym niewidoczne są akcje | 21 |
| 2.2 | Siatka z widocznymi zajściami akcji | 22 |
| 2.3 | Sieć statyczna stworzona na bazie dynamicznej sieci temporalnej | 23 |
| 2.4 | Przykład Ramki Danych | 24 |
| 3.1 | Zrzut ekranu z formatem danych dla jednego z audioprzewodników | 25 |
| 3.2 | Cykliczne restartowanie się urządzeń | 26 |
| 3.3 | Przykład niedokładności zapisu danych | 27 |
| 3.4 | Diagram poszczególnych etapów przygotowania danych | 27 |
| 3.5 | przed oczyszczeniem | 29 |
| 3.6 | po oczyszczeniu | 29 |
| 3.7 | Rezultaty ograniczenia okresu czasowego dla danych | 29 |
| 3.8 | Segmentacja sesji działania urządzenia | 30 |
| 3.9 | Segmentacja sesji użytkowników | 31 |
| 3.10 | Poszukiwanie momentów kluczowych sesji - funkcja <code>check_start(linie)</code> | 33 |
| 3.11 | Ekstrakcja i zapis danych finalnych | 34 |
| 3.12 | Struktura danych dodatkowych | 35 |
| 3.13 | Struktura czasu zwiedzania w minutach | 35 |
| 3.14 | Ramka danych zawierająca informacje o typach zwiedzających | 36 |
| 3.15 | Wykres typów zwiedzających | 37 |
| 3.16 | Stopień wykorzystania audioprzewodników, poprzez określenia stref zakończeń zwiedzań | 38 |

| | |
|--|----|
| 3.17 Dane dotyczące stref odwiedzonych podczas sesji | 39 |
| 3.18 Tabela liczebności "przejść" | 39 |
| 3.19 Wykres częstości wykorzystania stref zawierających dodatkowe informacje | 41 |
| 3.20 Pierwsza forma grafu | 42 |
| 3.21 Graf naniesiony na mapę z wagami krawędzi | 44 |
| 3.22 Graf naniesiony na mapę z wagami krawędzi ograniczony do stref o znanym położeniu oraz aktywowanych | 45 |

Spis programów

| | | |
|-----|--|----|
| 3.1 | Struktura danych w formacie JSON | 32 |
| 3.2 | Przygotowanie danych dotyczących długości czasu trwania zwiedzania. | 32 |
| 3.3 | Wyodrębnienie liczebności poszczególnych typów osób zwiedzających | 36 |
| 3.4 | Struktura krotki określającej krawędź grafu. | 40 |
| 3.5 | Wyznaczenie unikalnych stref znajdujących się w danych | 40 |
| 3.6 | Inicjalizacja i budowa pierwszej formy grafu. | 41 |
| 3.7 | Funkcja ustalająca pozycje węzłów | 41 |
| 3.8 | Inicjalizacja pierwszego grafu naniesionego na mapę | 43 |
| 3.9 | Inicjalizacja naniesionego na mapę grafu o zwiększonej czytelności | 43 |