AKADEMIA GÓRNICZO – HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

---

kierunek studiów: Informatyka

Analiza datasetów

Przedmiot: Teoria kompilacji i kompilatory

Paweł Łosek

Krzysztof Czuryło

Spis datasetów:

1. Dane przejazdów Ubera
2. Dane przejazdów taksówek
3. Awarie autobusów
4. Wypadki samochodowe
5. Pomiary aktywności ludzi
6. Pomiary dziennej aktywności pobrane z sensorów rozmieszczonych w domu
7. Pomiary zużycia energii elektrycznej w gospodarstwie domowym
8. Dane postaci w grze MMO World of Warcraft
9. Zbiór tweetów pro-ISIS zebrany po zamachach w Paryżu
10. Wezwania pomocy (911) w USA
11. Pomiary składu powietrza
12. Wydarzenia drogowe w Kraju Basków

Opisy datasetów

- Dane przejazdów Ubera

Nazwa zbioru:

Uber TLC FOIL Response

Link do zbioru:

https://github.com/fivethirtyeight/uber-tlc-foil-response

Opis słowny:

Zbiór danych zawiera informacje o przewozach zamówionych za pomocą aplikacji taksówkowej Uber. Dataset dotyczy samochodów zrzeszonych w 8 bazach znajdujących się w Nowym Jorku. Repozytorium zawiera przejazdy Ubera rejestrowane pomiędzy kwietniem a wrześniem 2014 r. oraz pomiędzy styczniem a czerwcem 2015 roku. Ponadto, umieszczono tam także informacje o przejazdach z 10 innych firm wynajmujących samochody oraz zbiorcze dane z 329 takich firm. Te pozostałe rejestrowane były w dniach 3 sierpnia oraz 15 i 22 września 2015 roku.

Wszystkie dane zostały otrzymane od Komisji ds. Taksówek i limuzyn Nowego Jorku.

Opis formatu:

Interesujący nas dataset (przejazdy z aplikacji Uber) jest w formacie csv. Plik zawiera nagłówek określający kolejno:

- "Date/Time" - data oraz czas startu przejazdu samochodem Uber,
- “Lat” - szerokośc geograficzna miejsca, gdzie zamówiony był samochód,
- “Lon” - długość geograficzna miejsca, gdzie zamówiony był samochód,
- “Base” - baza, do której przypisane było auto realizujące przejazd

Dwa przykładowe wiersze datasetu wyglądają następująco (zgodnie z kolejnością opisaną powyżej, nie są to kolejne wiersze):

"4/1/2014 0:55:00" , 40.7524, -73.996 , "B02512"
"4/5/2014 21:27:00", 40.7156, -74.0074, "B02598"

Uwagi i komentarze:

Dataset zawiera dosć mało informacji, ponadto wartości typu String ujęte są w cudzysłowiu, co wymaga uwzględnienia w implementacji mechanizmu parsującego plik, albo uzwględnienia przy tworzeniu wyszukiwanych wyrażeń (oczekiwane wartości również wpisywać w cudzysłowiu).

- Dane przejazdów taksówek

Nazwa zbioru:

2015 Yellow Taxi Trip Data

Link do zbioru:

https://data.cityofnewyork.us/view/ba8s-jw6u

Opis słowny:

Dataset zawiera rejestr przejazdów oficjalnymi, “żółtymi”, nowojorskimi taksówkami. Dane zbierane były pomiędzy styczniem a czerwcem 2015 roku i zawierają oprócz daty i precyzyjnego miejsca początku kursu również analogiczne dane dotyczące końca kursu. Dodatkowo załączone sa informacje na temat samego przejazdu jak ilość osób, wartość kursu czy metoda płatności. Dane zostały otrzymane od Komisji ds. taksówek i limuzyn Nowego Jorku. Zbiór zawiera ponad miliard rekordów.

Opis formatu:

Dataset jest w formacie csv. Plik zawiera nagłówek opisujący poszczególne wartości występujące w zbiorze:

- passenger_count - ilość pasażerów,
- trip_distance - długość kursu,
- pickup_longitude - długość geograficzna miejsca, gdzie zamówiona była taksówka,
- pickup_latitude - szerokość geograficzna miejsca, gdzie zamówiona była taksówka,
- store_and_fwd_flag - flaga dot. metadanych - czy przejazd został od razu przesłany na serwer, czy został zapisany w pamięci urządzenia
- dropoff_longitude - długość geograficzna miejsca końca kursu,
- dropoff_latitude - szerokość geograficzna miejsca końca kursu,
- fare_amount - opłata za kurs,
- extra - dodatkowe opłaty,
- mta_tax - podatek,
- tip_amount - napiwek dla taksówkarza
- tolls_amount - suma opłat związanych z przejazdem,
- total_amount - całkowita opłata, którą obciążono pasażerów,
- vendor_id - identyfikator technologii która dostarczyła rekord,
- pickup_datetime - stempel czasowy startu kursu,
- dropoff_datetime - stempel czasowy końca kursu,
- rate_code - strefa taryfowa na koniec przejazdu,
- payment_type - metoda płatności.

Dwa przykładowe wiersze datasetu(ze względu na dużą ilość kolumn poniżej zrzut ekranu):

1	1.25	-74.00766754	40.74103165	N	-74.00669861	40.75136948	6	0.5	0.5	1.1	0	8.4	2	2015 Jun 23 23:27:25	2015 Jun 23 23:32:03	1	1
5	3.9	-73.98355865	40.74322128	N	-73.98396301	40.70235825	15	0.5	0.5	2	0	18.3	2	2015 May 22 00:07:11	2015 May 22 00:23:29	1	1

Uwagi i komentarze:

Zbiór danych jest bardzo duży (ponad 11 GB po dekompresji), dlatego do jego przechowywania na potrzeby przetwarzania należałoby użyć bazy danych.

- Awarie autobusów

Nazwa zbioru:

Bus breakdowns and delays

Link do zbioru:

https://data.cityofnewyork.us/Transportation/Bus-Breakdown-and-Delays/ez4e-fazm

Opis słowny:

Zbiór danych zawiera informacje o awariach i opóźnieniach autobusów przewożących uczniów do szkół w Nowym Jorku. Dane przesyłane były od kierowców autobusów w czasie rzeczywistym w przypadku wystąpienia ww. zdarzeń. System zbierający dane wykorzystywany jest do informowania rodziców uczniów oraz szkół korzystających z usług firm logujących zdarzenia w systemie. Cały system jest dostępny publicznie i zawiera dane aktualizowane w czasie rzeczywistym.

Opis formatu:

Zbiór danych zapisany jest w formacie csv. Plik zawiera nagłówek definiujący pola:

- School_Year - rok szkolny zdarzenia,
- Busbreakdown_ID - identyfikator rekordu,
- Run_Type - rodzaj przejazdu,
- Bus_No - numer boczny autobusu,
- Route_Number - numer trasy,
- Reason - powód opóźnienia wprowadzony przez pracownika,
- Schools_Serviced - szkoły obsługiwane w ramach przejazdu,
- Occurred_On - stempel czasowy zdarzenia,
- Created_On - stempel czasowy wprowadzenia zdarzenia do systemu,
- Boro - dzielnica,
- Bus_Company_Name - firma, do której należy autobus którego dotyczy zdarzenie,
- How_Long_Delayed - długość opóźnienia,
- Number_Of_Students_On_The_Bus - ilość pasażerów,
- Has_Contractor_Notified_Schools - czy firma świadcząca usługę powiadomiła szkoły,
- Has_Contractor_Notified_Parents - czy firma świadcząca usługę powiadomiła rodziców,
- Have_You_Alerted_OPT - czy system został powiadomiony o zdarzeniu,
- Informed_On - stempel czasowy momentu powiadomienia (dot. 3 powyższych pól),
- Incident_Number - identyfikator zgłoszenia w przypadku zgłoszenia przez klienta,
- Last_Updated_On - stempel czasowy ostatniej modyfikacji rekordu,
- Breakdown_or_Running_Late - określa, czy zdarzenie to awaria czy opóźnienie,
- School_Age_or_PreK - określa, jakiej szkoły dotyczy trasa obsługiwana przez autobus

Dwa przykładowe wiersze datasetu(ze względu na dużą ilość kolumn poniżej zrzut ekranu):

2015-2016	1227241	Pre-K/EI	9815	3	Heavy Traffic	C329	11/04/2015 08:46:00 AM	11/04/2015 08:49:00 AM	Bronx	G.V.C. LTD.	15MIN	16	Yes	Yes	Yes	11/04/2015 08:49:00 AM	11/04/2015 08:49:05 AM	Running Late	Pre-K
2015-2016	1268267	Pre-K/EI	1015	UCP-3	Heavy Traffic	C522	05/13/2016 02:20:00 PM	05/13/2016 02:32:00 PM	Bronx	PHILLIPS BUS SERVICE	40MINS	0	Yes	Yes	Yes	05/13/2016 02:32:00 PM	05/13/2016 02:32:28 PM	Running Late	Pre-K

Uwagi i komentarze:

API umożliwia pobranie zbioru w wielu formatach(csv, json, xml itp.), ale optymalny ze względu na odczytywanie i procesowanie w ramach stworzonej aplikacji jest format csv.

- Wypadki samochodowe

Nazwa zbioru:

NYPD Motor Vehicle Collisions

Link do zbioru:

https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

Opis słowny:

Zbiór danych zawiera szczegóły kolizji i wypadków drogowych udostępniony przez ratusz Nowego Jorku. Dane dotyczą okresu od stycznia 2012 roku do czerwca 2016.

Opis formatu:

Zbiór danych zapisany jest w formacie csv, zawiera nagłówek definiujący następujące pola:

- DATE - stempel czasowy - data,
- TIME - stempel czasowy - godzina,
- BOROUGH - dzielnica,
- ZIP CODE - kod pocztowy,
- LATITUDE - szerokość geograficzna,
- LONGITUDE - długość geograficzna,
- LOCATION - lokalizacja (dwie poprzednie wartości jako jedno pole),
- ON STREET NAME - ulica na której doszło do zdarzenia,
- CROSS STREET NAME - najbliższe skrzyżowanie,
- OFF STREET NAME - nazwa ulicy ,w okolicy której doszło do zdarzenia,
- NUMBER OF PERSONS INJURED - ilość osób rannych,
- NUMBER OF PERSONS KILLED - ilość osób zabitych,
- NUMBER OF PEDESTRIANS INJURED - ilość pieszych rannych,
- NUMBER OF PEDESTRIANS KILLED - ilość pieszych zabitych,
- NUMBER OF CYCLIST INJURED - ilość rowerzystów rannych,
- NUMBER OF CYCLIST KILLED - ilość rowerzystów zabitych,
- NUMBER OF MOTORIST INJURED - ilość motocyklistów rannych,
- NUMBER OF MOTORIST KILLED - ilość motocyklistów zabitych,
- CONTRIBUTING FACTOR VEHICLE 1 .. 5 - czynnik, który spowodował że pojazd (1..5) brał udział w zdarzeniu,

- UNIQUE KEY - identyfikator zdarzenia,
- VEHICLE TYPE CODE 1 .. 5 - typ pojazdu (1-5) biorącego udział w zdarzeniu,

Ze względu na bardzo duża ilość kolumn i nieczytelność pojedynczego wiersza nawet na zrzucie ekranu przykład nie zostaje umieszczony w raporcie. Po wejściu w odnośnik do zbioru danych jest możliwość podglądu danych bez ich pobierania.

Uwagi i komentarze:

API umożliwia pobranie zbioru w wielu formatach(csv, json, xml itp.), ale optymalny ze względu na odczytywanie i procesowanie w ramach stworzonej aplikacji jest format csv. Część danych dubluje się (longitude i latitude tworzą kolumnę location), występuje duża ilość pól pustych (szczególnie w kolumnach występujących pięciokrotnie).

- Pomiary aktywności ludzi

Nazwa zbioru:

Localization Data for Person Activity

Link do zbioru:

https://archive.ics.uci.edu/ml/datasets/Localization+Data+for+Person+Activity

Opis słowny:

Zbiór zawiera informacje pobierane z sensorów umieszczonych na ciałach 5 osób w trakcie wykonywania określonych czynności. Każda z osób korzystała z czterech takich sensorów (na kostkach, w pasie i na klatce piersiowej). Pomiary dokonywane były w trakcie sekwencji ruchów, które zostały powtórzone pięciokrotnie dla każdej z osób.

Opis formatu:

Dane są w formacie csv, plik nie posiada nagłówka. Dane pogrupowane są według kolumn:

- Numer sekwencji
- Identyfikator sensora
- Unikalny identyfikator czasu
- Stempel czasowy
- Współrzędna x sensora
- Współrzędna y sensora
- Współrzędna z sensora
- Rodzaj czynności

Przykładowe dwa wiersze zbioru:

A01	010-000-024-033	633790226053442677	27.05.2009 14:03:25:343	3.95849609375	1.70356285572052	0.5110414028167725	walking
D03	020-000-033-111	633790159925606016	27.05.2009 12:13:12:560	4.377655029296875	1.2984321117401123	0.3073897957801819	lying

- Pomiary dziennej aktywności pobrane z sensorów rozmieszczonych w domu

Nazwa zbioru:

Activities of Daily Living (ADLs) Recognition Using Binary Sensors

Link do zbioru:

https://archive.ics.uci.edu/ml/datasets/Activities+of+Daily+Living+(ADLs)+Recognition+Using+Binary+Sensors

Opis słowny:

Repozytorium zawiera dwa zestawy datasetów (pomiary dla dwóch osób). Informacje dotyczą czynności wykonywanych przez te osoby w trakcie przebywania w domu, rejestrowane za pomocą 12 sensorów rozmieszczonych w różnych pomieszczeniach. Datasety opisują 35 pełnych dni z listopada i grudnia 2011 roku. Zestawy zawierają zbiór opisujący czynności oraz zbiór opisujący odczyty sensorów. Dane pozyskiwane były w ramach bezprzewodowej sieci i opisywane manualnie.

Opis formatu:

Zbiory danych zawierają nagłówki i zapisane są w formacie tsv. Zbiory podzielone są na zbiory dla osoby A oraz B, pliki z czynnościami mają następującą strukurę:

- Start time - stempel czasowy początku czynności
- End time - stempel czasowy końca czynności
- Activity - czynność

Przykładowe rekordy:

2011-11-28 02:27:59	2011-11-28 10:18:11	Sleeping
2011-11-30 18:57:39	2011-11-30 19:37:10	Spare_Time/TV

Pliki z odczytami sensorów mają z kolei następujące pola:

- Start time - stempel czasowy początku odczytu
- End time - stempel czasowy końca odczytu
- Location - lokalizacja sensora
- Type - typ sensora
- Place - pomieszczenie, gdzie umieszczony był sensor

Przykładowe rekordy:

2012-11-11 21:14:21	2012-11-12 00:21:49	Seat	Pressure	Living
2012-11-12 09:41:14	2012-11-12 09:41:17	Door	PIR	Living

- Pomiary zużycia energii elektrycznej w gospodarstwie domowym

Nazwa zbioru:

Electric power consumption

Link do zbioru:

https://www.kaggle.com/adriferher/electric-power-consumption-data-set

Opis słowny:

Zbiór danych zawiera pomiary zużycia energii elektrycznej w pojedynczym gospodarstwie domowym w watogodzinach na przestrzeni 47 miesięcy (od grudnia 2006 do listopada 2010), przy czym pomiary dokonywane są co jedną minutę.

Opis formatu:

Dane zgromadzone są w formacie SSV (semicolon separated values). Zbiór zawiera nagłówek definujący poniższe pola:

- Date - data pomiaru
- Time - godzina pomiaru
- Global_active_power - uśredniona minutowa globalna moc czynna
- Global_reactive_power - uśredniona minutowa globalna moc bierna
- Voltage - uśrednione minutowe napięcie
- Global_intensity - uśrednione minutowe globalne natężenie prądu
- Sub_metering_1 - pomiar zużycia w kuchni
- Sub_metering_2 - pomiar zużycia w pralni
- Sub_metering_3 - pomiar zużycia energii przez klimatyzację i bojler

Przykładowe dwa rekordy:

16/12/2006;17:24:00;4.216;0.418;234.840;18.400;0.000;1.000;17.000
16/12/2006;18:35:00;6.072;0.000;232.480;26.400;0.000;27.000;17.000

Uwagi i komentarze:

Nie wszystkie rekordy datasetu są kompletne - zdarzają się stemple czasowe nie zawierające wartości odczytów.

- Dane postaci w grze MMO World of Warcraft

Nazwa zbioru:

World of Warcraft Avatar History

Link do zbioru:

https://www.kaggle.com/mylesoneill/warcraft-avatar-history/

Opis słowny:

Zbiór zawiera szczegółowe informacje dotyczące postaci w grze komputerowej World of Warcraft i przedstawia ich rozwój w czasie rozgrywki. Opisany zbiór dotyczy informacji zebranych na jednym z serwerów gry dla jednej z frakcji w roku 2008 Informacje pobierane były w interwałach 10 minutowych dla postaci mających status on-line.

Opis formatu:

Dane zgromadzone są w formacie csv, posiadają nagłówek definiujący pola:

- char - unikalny id postaci
- level - poziom doświadczenia postaci
- race - rasa postaci
- charclass - klasa postaci
- zone - lokalizacja postaci w momencie pomiaru
- guild - gildia, do której należy postać
- timestamp - stempel czasowy pomiaru

Dwa przykładowe rekordy datasetu:

22307,70,Orc,Warrior,Orgrimmar,174,01/01/08 00:02:19
27711,70,Undead,Priest,Stranglethorn Vale,103,01/01/08 00:04:31

Uwagi i komentarze:

Zbiór dostępnym pod podanym adresem jest zbiorem ograniczonym. Szerszy dataset można pobrać z linku w opisie na stronie podanej powyżej.

- Zbiór tweetów pro-ISIS zebrany po zamachach w Paryżu

Nazwa zbioru:

How ISIS Uses Twitter

Link do zbioru:

https://www.kaggle.com/kzaman/how-isis-uses-twitter

Opis słowny:

Dataset zawiera 17.000 wpisów umieszczonych przez ponad 100 znanych sympatyków Państwa Islamskiego z całego świata po zamachach terrorystycznych dokonanych w Paryżu w listopadzie 2015 roku.

Opis formatu:

Informacje zapisane są w formacie csv, z nagłówkiem definiującym następujące właściwości:

- Name - imie autora wpisu
- Username - nazwa użytkownika
- Description - opis wpisu
- Location - lokalizacja z której dodano wpis
- Followers - ilość osób śledzących wpis w momencie pobrania danych przez API
- Numberstatuses - ilość wpisów umieszczonych przez autora w momencie pobrania danych przez API
- Time - stempel czasowy wpisu
- Tweets - treść wpisu

Przykładowe rekordy z datasetu:

Rain Qattal	1515Ummah	21:15 For they fled from the Swords from the drawn Sword and from the bent Bow and	Punch Jammu And Kasl	214	169	2/19/2016 16:06	@IbnElHijaz lool
راعي البقر حوز الهند	Uncle_SamCoco	Here to defend the American freedom and also the freedom of coconut . Cat Lover or H	United States	1376	2852	2/19/2016 16:49	@Abduhark US special forces don't use "Takfiri" or "Muhajirin" please

Uwagi i komentarze:

Duża część wpisów jest w językach arabskich, co utrudnia analizę. Oprócz tego, część danych nie jest zapisana zgodnie z konwencją csv - występuje problem z ich odczytaniem, przez skorzystaniem z datasetu należaloby zwalidować plik i oczyścić go z takich wartości.

- Wezwania pomocy (911) w USA

Nazwa zbioru:

Emergency - 911 Calls

Link do zbioru:

https://www.kaggle.com/mchirico/montcoalert

Opis słowny:

Zbiór zawiera spis wezwań pomocy poprzez numer telefonu 911 w stanie Pensylwania w Stanach Zjednoczonych. Zebrane dane dotyczą okresu od grudnia 2015 do końca sierpnia 2016 roku. Informacje dotyczą trzech typów wypadków: pożary, wypadki drogowe oraz wezwania ze względów medycznych (EMS).

Opis formatu:

Dane zapisane są w formacie csv wraz z nagłówkiem opisującym poszczególne pola:

- Lat - szerokość geograficzna miejsca zdarzenia
- Lng - długość geograficzna miejsca zdarzenia
- Desc - opis zdarzenia
- Zip - kod pocztowy miejsca zdarzenia
- Title - tytuł nadany przy wprowadzeniu zdarzenia do zbioru
- timeStamp - stempel czasowy wydarzenia
- Twp - region zdarzenia
- Addr - adrs zdarzenia
- E - nieokreslona wartość (dla wszystkich rekordów równa 1)

Przykładowe rekordy ze zbioru:

40.0841613	-75.3083857	BROOK RD & COLWELL LN; PLYMOUTH; 2015-12-10 @ 16:32:10;	19428	Traffic: VEHICLE ACCID	2015-12-10 17:40:02	PLYMOUTH	BROOK RD & COLWELL LN
40.2249227	-75.5280446	LINFIELD TRAPPE RD; LIMERICK; 2015-12-10 @ 18:50:23-Station:STA51;	19468	Fire: VEHICLE ACCIDEN	2015-12-10 18:52:00	LIMERICK	LINFIELD TRAPPE RD

Uwagi i komentarze:

Dataset zawiera jedną kolumnę danych o nieznanym znaczeniu (“e”).

- Pomiary składu powietrza

Nazwa zbioru:

Air Quality Monitoring: Unverified Hourly Air Quality and Meteorological Data

Link do zbioru:

ftp://ftp.env.gov.bc.ca/pub/outgoing/AIR/Hourly_Raw_Air_Data/Air_Quality/

Opis słowny:

Repozytorium zawiera datasety opisujące stężenia składników powietrza oraz parametry meteorologiczne z ostatnich 30 dni zbierane przez Ministerstwo Środowiska Kanady. Informacje dotyczą obszaru Kolumbii Brytyjskiej. Dane odczytywane są przez stacje z interwałem jednogodzinnym. Informacje dostępne w repozytorium nie są zweryfikowane.

Opis formatu:

Datasety dostępne w repozytorium zapisane są w formacie csv, pliki zawierają nagłówki. Przykładowe dane zawarte w datasecie z pomiarami pyłów PM10:

- DATE_PST - stempel czasowy pomiaru
- EMS_ID - identyfikator pomiaru
- STATION_NAME - nazwa stacji pomiarowej
- PARAMETER - mierzony parametr
- AIR_PARAMETER - mierzony parametr
- INSTRUMENT - urządzenie pomiarowe
- RAW_VALUE - wynik odczytu

- UNIT - jednostka pomiaru
- STATUS - status
- AIRCODESTATUS - kod ogólnego stan powietrza
- STATUS_DESCRIPTION - opis statusu
- REPORTED_VALUE - zareportowana wartość

Przykładowe rekordy dla powyższego datasetu:

2016-08-19 18:00	0220204	Powell River Cranberry Lake_60	PM10	PM10	PM10_R&P_TEOM	30.62275	ug/m3	1	n/a	Data Ok	30.6
2016-08-24 20:00	E286369	Castlegar Zinio Park	PM10	PM10	PM10_R&P_TEOM	9.1000000	ug/m3	1	n/a	Data Ok	9.1

- Wydarzenia drogowe w Kraju Basków

Nazwa zbioru:

Traffic incidents in the Basque Country

Link do zbioru:

https://datahub.io/dataset/basque-traffic-incidents

Opis słowny:

Zbiór danych opisuje wydarzenia drogowe opisujące aktualną sytuację w Kraju Basków w Hiszpanii. Informacje dotyczą zdarzeń drogowych takich jak objazdy, wydarzenia sportowe czy roboty drogowe i odwzorowują faktyczny stan w momencie pobierania datasetu.

Opis formatu:

Zbiór danych zapisany jest w formacie XML. Plik pobrany w ramach testu zawierał następujące informacje:

- Tipo - typ zdarzenia
- Fechahora_ini - stempel czasowy
- Autonomia - obszar w Kraju Basków
- Causa - powód wydarzenia
- Provincia - prowincja
- Nombre - nazwa drogi
- Pk_initial - początek odcinka drogi
- Sentido - parametr
- Matricula - rejest
- Carretera - numer drogi
- Poblacion - gmina
- Pk_final - koniec odcinka drogi
- Nivel - oznaczenie

Przykładowy rekord datasetu:

```
<incidencia>
    <tipo>Accidente</tipo>
    <autonomia>Euskadi</autonomia>
    <provincia>GIPUZKOA</provincia>
    <matricula>SS</matricula>
    <causa>Alcance</causa>
    <poblacion>Errenteria</poblacion>
    <fechahora_ini>2016-09-05 19:06:17</fechahora_ini>
    <nivel>Blanco</nivel>
    <carretera>GI-2132</carretera>
    <pk_inicial>15</pk_inicial>
    <pk_final>15</pk_final>
    <sentido>REKALDE</sentido>
  </incidencia>
```

Uwagi i komentarze:

Dane zawierają niestandartowe znaki z alfabetu hiszpańskiego.