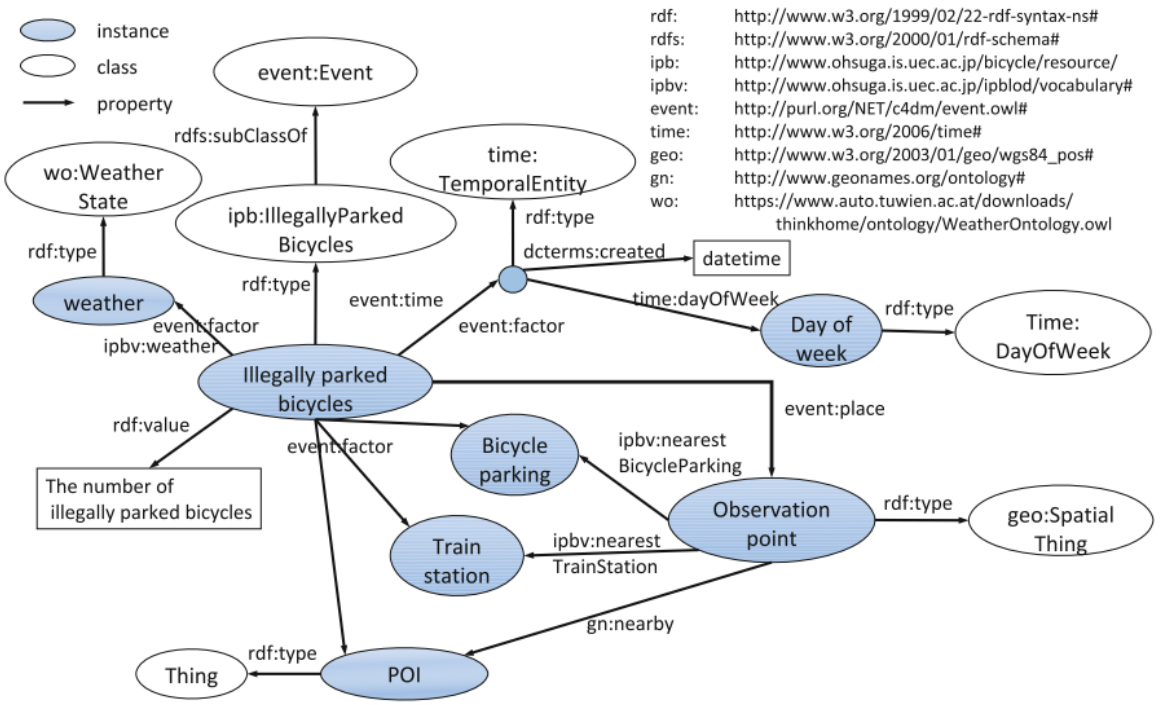
Martin Chvíla, 26.11.2016

Propojená data na webu, minireferát – abstrakt článku:

EGAMI, Shusaku, Takahiro KAWAMURA a Akihiko OHSUGA. Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo. *The Semantic Web – ISWC 2016: 15th International Semantic Web Conference Kobe, Japan, October 17–21, 2016 Proceedings, Part II*. Springer, 2016, 291-307. ISSN 0302-9743.

Článek pojednává o návrhu a implementaci řešení problému s ilegálně zaparkovanými koly v Tokiu. Během posledních 40 let se počet jízdních kol v Japonsku ztrojnásobil. Kvůli tomu se objevuje velké množství ilegálně zaparkovaných kol, které brání plynulé dopravě, působí dopravní nehody, zvyšují množství krádeží a kazí vzhled ulic. Jako pomoc při řešení problému navrhli autoři zveřejnit data o nelegálně zaparkovaných kolech jako LOD (Linked Open Data) a vytvořit vizualizace o rozložení zaparkovaných kol po městě. To může sloužit jako nástroj pro navržení optimálních míst pro vybudování parkovišť a pomoci s odstraňováním zaparkovaných kol.

Jako základní strukturu pro modelování dat o ilegálně zaparkovaných kolech byla použita ontologie Event[1]. Každé pozorování tedy prakticky pozorovaná nelegálně zaparkovaná kola je považováno jako event. V rámci ontologie má event vlastnosti place, time, agent, factor a product. Po průzkumu článků o nelegálně zaparkovaných kolech na internetu a byla zanalyzována častá klíčová slova a tím odvozeny nejdůležitější údaje o zaparkovaných kolech. Na základě této analýzy byla použitá ontologie přizpůsobena na konkrétní problém. Například vlastnost agent byla vynechána, protože je obtížné získat data o vlastnících kol a byla nahrazena statistickými údaji o obyvatelích na daném místě. Výsledné LOD schéma lze vidět na Obr. 1.



Obr. 1: LOD schéma

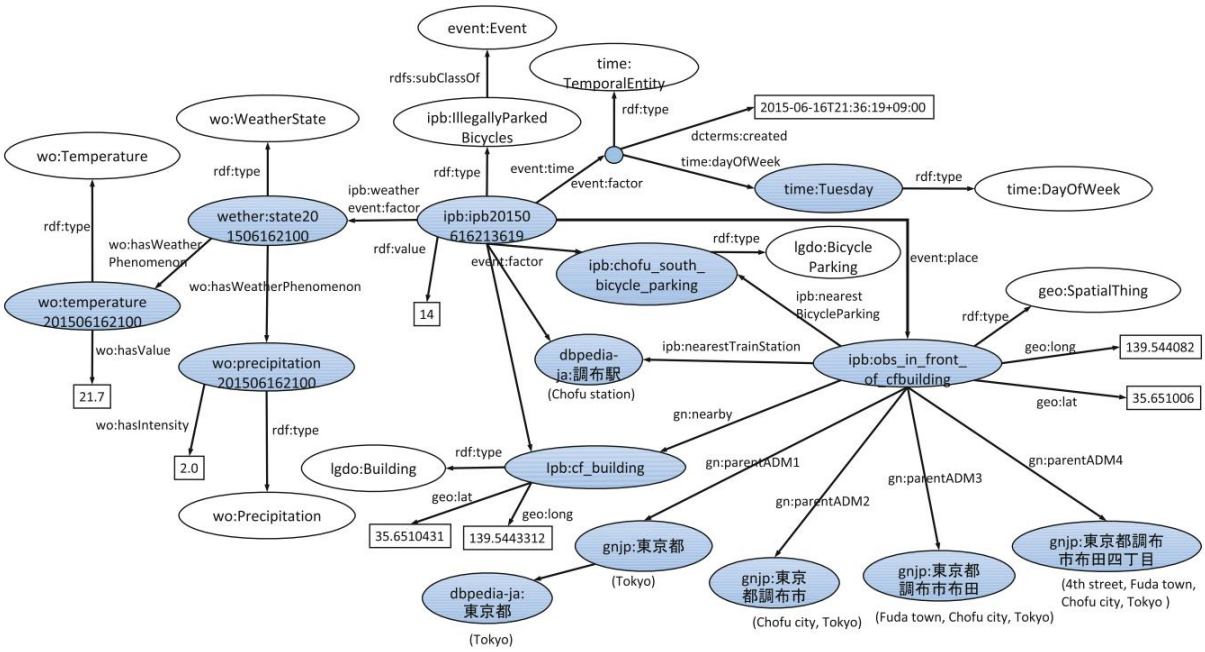
[1] http://purl.org/NET/c4dm/event.owl

Pro pomoc se sběrem dat byla vytvořena aplikace pro odeslání příspěvku na twitteru o nalezení ilegálně zaparkovaných kol obsahující lokaci, čas a počet zaparkovaných kol (Obr. 2). Ke každé lokaci, ve které byly nalezeny nelegálně zaparkovaná kola byly nasbírány blízké POI z Google Places[2] a Foursquare API[3]. Dále byly také získány informace o parkovacích místech (lokace, cena, kapacita,…) z webu tokijského úřadu[4] ve formátu csv a o obyvatelstvu v okolí z japonského statistického webu e-stat[5]. Jako poslední externí data použili informace o počasí z webu Japonského meteorologického úřadu[6]. Všechna tato data byla převedena do RDF podle navrženého LOD schématu. Příklad části grafu dat odpovídající konkrétnímu pozorování můžeme vidět na Obr. 3. Celý graf je veřejně dostupný pomocí SPARQL endpointu[7].



Obr. 2: Screenshot aplikace pro twitter

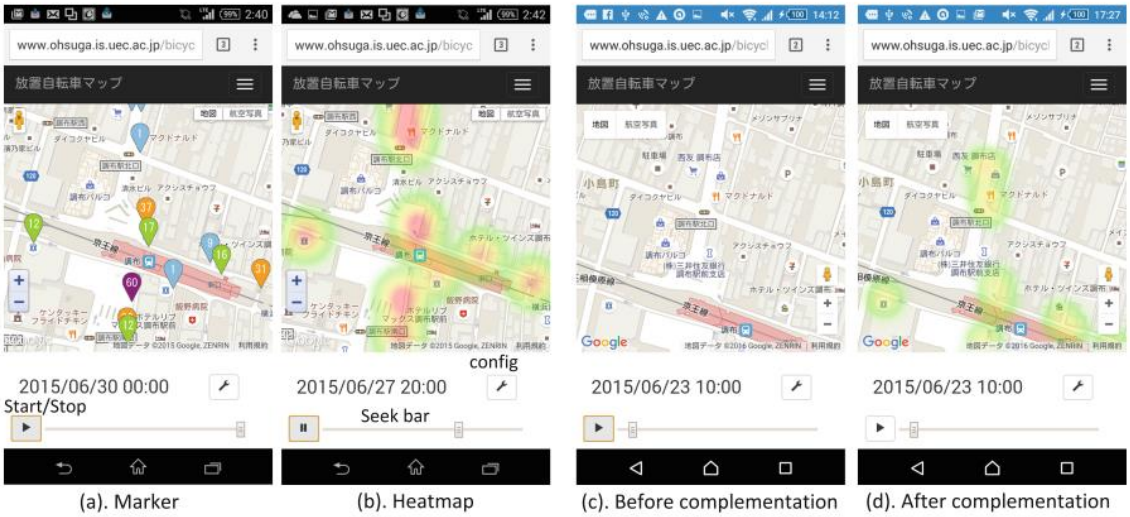
[2] https://developers.google.com/places/?hl=en
[3] https://developer.foursquare.com/
[4] http://www.soumu.metro.tokyo.jp/30english/index-en.htm
[5] http://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do
[6] http://www.jma.go.jp/jma/indexe.html
[7] http://www.ohsuga.is.uec.ac.jp/sparql



Obr. 3: Příklad grafu pro jedno pozorování

Výsledkem celé práce má být vizualizace rozložení ilegálně zaparkovaných kol po městě přes celý den. Nicméně jako zdrojová data mají autoři pouze pozorování na určitém místě v konkrétní čas. Proto se rozhodli použít predikční model, který odhadne počet ilegálně zaparkovaných kol na daných místech v jiný čas na základě faktorů dostupných z doplňujících dat jako jsou údaje o počasí, blízké POI nebo blízké parkoviště pro kola. Pro tento odhad použili model Bayesovské sítě. Jako vstupní proměnné do modelu byly použity následující příznaky: lokace, den, hodina, srážky, teplota, denní sazba nejbližšího parkoviště, měsíční sazba, počet obyvatel/1km^2 , počet dojiždějících veřejnou dopravou, blízká POI (46 kategorií POI, každý příznak je boolean pro daný typ POI). Po několika iteracích modelu dosáhli průměrné přesnosti predikce 71% při 10-násobné crossvalidaci.

Na základě těchto odhadů vytvořili autoři webovou aplikaci pro vizualizaci. V aplikaci je možné zobrazit jak původní pozorování, tak odhady rozložení ilegálně zaparkovaných kol v čase (screenshoty na Obr. 4).



Obr. 4: Screenshoty webové aplikace pro vizualizaci počtu ilegálně zaparkovaných kol

Výsledkem projektu jsou tedy celkem 3 komponenty:

- Aplikace pro sběr pozorování
- RDF data o ilegálně zaparkovaných kolech a dalších faktorech souvisejících s místem a časem pozorování
- Vizualizace rozdělení počtu ilegálně zaparkovaných kol na určitých místech v čase

Hlavním problémem studie je nedostatek zdrojových dat (pouze 897 pozorování). Autoři předpokládají, že jejich dosavadní výsledek rozšíří povědomí o problému a motivuje občany poskytovat více pozorování, které budou sloužit k rozšíření a zdokonalení výsledků modelu.