A Prototype for Semantic based Diagnosis of Road Traffic Congestions

Marco Luca Sbodio, Freddy Lecue, Anika Schumann

IBM Research - Ireland
Damastown Industrial Estate, Dublin, Ireland
{(firstname.lastname)@ie.ibm.com}

Abstract. Retrieving the causes of road traffic congestions in quasi real-time is an important task that will enable city managers to get better insight into traffic issues and thus take appropriate corrective actions in a timely way. Our work, accepted at ISWC 2012, tackles this problem by integrating and reasoning over a variety of heterogeneous data sources including data streams. In this paper we present an initial prototype of our work for the city of Dublin, Ireland.

1 Introduction

Consider the case of city planning in anticipation of large events (for a example Republic of Ireland World Cup qualifier match in Croke Park, Dublin), or in reaction to unplanned events (for example a mob assembling in the Dublin Docklands area). By integrating and correlating partial observations from multiple data sources, we could infer that bad weather, coupled with a large number of people assembling in one area of the city on a normal working day, coupled with a lack of public parking, led to traffic chaos that was widely reported in the media, driving strong negative sentiment towards the handling of such events. Whilst such an analysis is a useful tool for understanding "*what went wrong*" and "*what were the causes*" after the event, our work is the first one that can **compute causes** of such unexpected situations **in quasi real-time**; other works focus on detecting, visualizing and analyzing traffic congestions [1]. We achieve this by exploiting semantic representations of historical data (such as traffic congestions data) and feeding them into an AI diagnosis approach. The work is described in a paper accepted at the In-Use track of ISWC 2012 [2].

In this paper we present an initial prototype [1] of our approach that we have developed for the city of Dublin and describe the data sets that we have semantically encoded.

2 Diagnosing Road Traffic Congestions

We start by briefly describing our diagnosis approach [2] shown in Figure 1. First, its heterogeneous input data (see next section for their details) are integrated using semantic web technologies. This then allows AI diagnosis techniques to compute off-line a

---

[1] see video at http://www.youtube.com/watch?v=xT5dPpnayZI

diagnoser representing *historical observations* over a *time window* and their *explanations* (for example Canal street was congested in 2012, May 1st at 6:00pm because of a concert event in Aviva stadium and road works in Bath avenue). Finally, quasi real-time diagnosis consists in combining semantic matching and AI diagnosis techniques for (i) retrieving "*similar*" causes (e.g., roads with heavy traffic of same duration) with "*similar*" conditions (e.g., nearby sport events) which have appeared in the past and (ii) interpreting them in the real-time context.
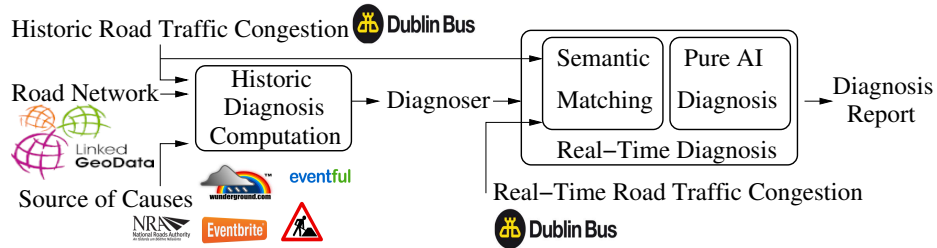

Fig. 1. Overview of the Semantics-Augmented Diagnosis Approach.

3   Data Sets

Table 1 lists the heterogeneous data sets considered in our scenario. Most of these data sources are public, and many are provided by Dublin City Council through *dublinked.ie*[2] web site, and hosted at IBM.

The **Dublin Bus Data**, encoded according to the SIRI standard (note *a* in Table 1), is transformed into a real-time stream (information about 1000 buses updated every 20 seconds) that is persisted into CSV files (one file per day). Each SIRI record (line in a CSV file) contains information about the current position (latitude and longitude) of a vehicle, its line number, its direction along the bus line, if the bus is in congestion, and if the bus is at a stop point. Information about bus lines and stops is given separately through other CSV files (such information is static, or at least it changes very rarely). We have developed a simple $\mathcal{EL}^{++}$ ontology to represent SIRI data. The actual SIRI records (lines from the CSV files) are modeled as instances of the class *VehicleAtomicUpdate*, which has a property for each field of a SIRI record. Based on a history of 217 SIRI data files (approximately 26 GB), referring to 217 days in 2011 (approximately 122 MB a day), 44.7% of the SIRI records generates 8 triples/record and 47.2% generate 11 triples/record (1 triple to define the type of the RDF resource, and 10 triples to specify the SIRI properties); the other records generate either 9 or 10 triples/record. The varying number of triples per record is due to some missing fields. The instances of VehicleAtomicUpdate with missing properties are nevertheless useful to estimate the number of buses in a bounding box in a certain time window (latitude, longitude, and timestamp are always available in SIRI records).

City Events were captured through *Eventful* and *EventBrite* web sites (notes *e, f* in Table 1). An average of 187 events a day (i.e., same days as those captured for SIRI data) have been described using some LOD vocabularies e.g., DBpedia, Talis. In addition we

---

Data Source	Provider	Format	Size
Dublin Buses Data Stream: vehicle data (GPS location, line number, delay, ...)	Dublin City Council (private)	SIRI[a](XML)	4-5 GB/day
Wunderground for Dublin: real-time weather information	Wunderground[b](public)	CSV	0.05-1.5 GB/day
Road & Weather Conditions	NRA[c](public)	CSV	0.1 GB/day
Road Works & Maintenance	Dublinked[d](public)	CSV	0.01 GB/day
Events in Dublin	Eventbrite[e] and Eventful[f](public)	XML	0.001-0.05 GB/fay
DBPedia	DBPedia[g](publifc)	RDF	3.5×10^6 concepts
Dublin roads: list of road types, junctions and GPS coordinates	Linkedgeodata[h](public)	RDF	0.1 GB

[a] SIRI (Service Interface for Real Time Information) is a standard for exchanging real-time information about public transport services and vehicles - http://siri.org.uk
[b] http://www.wunderground.com/weather/api
[c] NRA - National Roads Authority http://www.nratraffic.ie/weather
[d] http://www.dublinked.ie/datastore/datastore.php
[e] https://www.eventbrite.com/api
[f] http://api.eventful.com
[g] http://dbpedia.org
[h] http://linkedgeodata.org

enriched the events description with $\mathcal{EL}^{++}$ GCIs to capture their categories, which are used for computing not only fined grained matching between historical and new events, but also for computing the diagnosis report. Each event has been described on average through 26 RDF triples.

Similarly an average of 51 **Road Works and Maintenance** [3] records a day have also been enriched through 16 RDF triples each. An $\mathcal{EL}^{++}$ enrichment of this raw data ensures that historical and new records can be matched for diagnosis and reporting purposes. We also injected $14,316$ $\mathcal{EL}^{++}$ GCIs (6 RDF triples each) to describe 4772 **Roads and their Interconnections** [4].

The **Core Static Ontology**, which is used for representing SIRI, events, road works, road weather and Dublin weather data, is composed of 67 concepts with 24 role descriptions (25 concepts subsume the 42 remaining ones with a maximal depth of 4). Finally, a **History of** 217 **days of the Traffic Congestion Information** was computed based on buses data streams (encoded by more than 1×10^9 RDF triples). Information about past events, road works, weather information and road conditions was stored as 1.1×10^6 RDF triples.

---

[3] CSV sample in http://www.dublinked.ie/datastore/metadata064.php
[4] CSV sample in http://www.dublinked.ie/datastore/metadata125.php
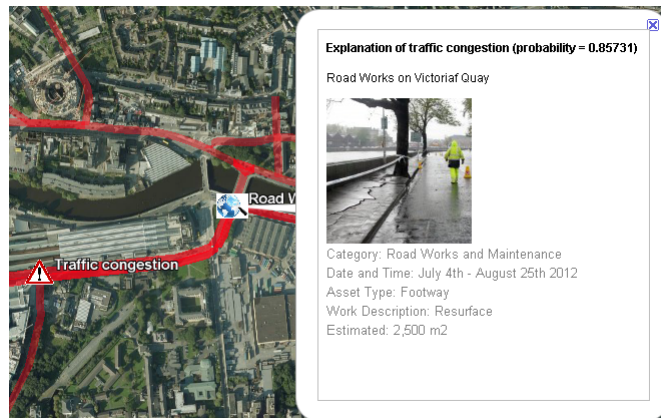
4   Prototype

Our initial prototype analyses data of Dublin buses, and displays congested roads on an interactive map (see figure 2(a)). Based on our quasi real-time diagnosis component, the system displays also explanations of selected road traffic congestions (see figure 2(b)). The explanation contains both data about its accuracy, and information extracted from the data source of the identified cause (road maintenance in the case of figure 2(b)).
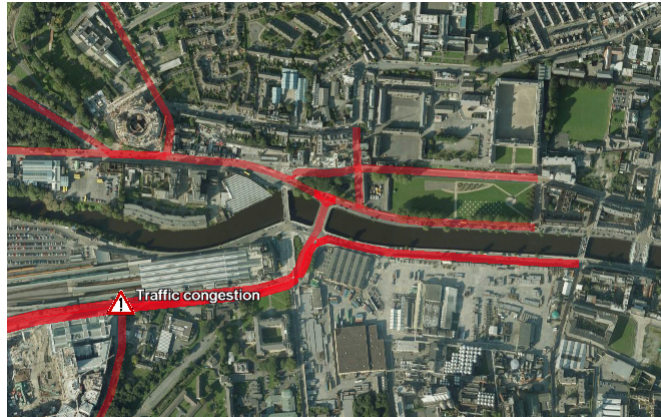
(a) Detection of traffic congestion

(b) Diagnosis of traffic congestion

Fig. 2. Screenshots of Automated Traffic Diagnosis Prototype

References

1. Biem, A., Bouillet, E., Feng, H., Ranganathan, A., Riabov, A., Verscheure, O., Koutsopoulos, H.N., Moran, C.: Ibm infosphere streams for scalable, real-time, intelligent transportation services. In: SIGMOD. (2010) 1093–1104
2. Lécué, F., Schumann, A., Sbodio, M.: Applying semanticweb technologies for diagnosing road traffic congestions. In: to appear in ISWC. (2012)