Demonstration Paper: Using a Knowledge Graph to Combat Human Trafficking*

Pedro Szekely[1], Craig A. Knoblock[1], Jason Slepicka[1], Andrew Philpot[1], Amandeep Singh[1], Chengye Yin[1], Dipsy Kapoor[1], Prem Natarajan[1], Daniel Marcu[1], Kevin Knight[1], David Stallard[1] Subessware S. Karunamoorthy[1], Rajagopal Bojanapalli[1], Steven Minton[2], Brian Amanatullah[2], Todd Hughes[3], Mike Tamayo[3], David Flynt[3], Rachel Artiss[3], Shih-Fu Chang[4], Tao Chen[4], Gerald Hiebel[5], and Lidia Ferreira[6]

[1] University of Southern California, Information Sciences Institute, Marina del Rey, CA, USA
{pszekely,knoblock,slepicka,philpot,dipsy,pnataraj,marcu,knight,stallard}@isi.edu
{singhama,chengyey,selvamee,bojanapa}@usc.edu
[2] InferLink Corporation, El Segundo, CA, USA
{sminton,bamanatullah}@inferlink.com
[3] Next Century Corporation Columbia, MD, USA
{todd.hughes,mike.tamayo,david.flynt,rachel.artiss}@nextcentury.com
[4] Columbia University, New York, USA
{sfchang,taochen}@ee.columbia.edu
[5] Universitat Innsbruck, Austria
gerald.hiebel@uibk.ac.at
[6] Federal University of Minas Gerais, Brazil
lidiaferreira@dcc.ufmg.br

Abstract. There is a huge amount of data spread across the web and stored in databases that we can use to build knowledge graphs. However, exploiting this data to build knowledge graphs is difficult due to the heterogeneity of the sources, scale of the amount of data, and noise in the data. In this work, we developed a system for building knowledge graphs by exploiting semantic technologies to reconcile the data continuously crawled from diverse sources, to scale to billions of triples extracted from the crawled content, and to support interactive queries on the data. We applied our approach, implemented in the DIG system, to the problem of combating human trafficking and deployed it to six law enforcement agencies and several non-governmental organizations to assist them with finding traffickers and helping victims. The demonstration will show the resulting application that is currently in use by these law enforcement agencies.

* This is a demonstration paper for the In-Use Paper titled Building and Using a Knowledge Graph to Combat Human Trafficking by Szekely et al. presented at ISWC 2015

Keywords: Linked Data, Knowledge Graphs, Entity Linkage, Data Integration, Information Extraction

1 Introduction

Human trafficking is a form of modern slavery where people profit from the control and exploitation of others, forcing them to engage in commercial sex or to provide services against their will. The statistics of the problem are shocking. In 2014 the International Labor Organization on The Economics of Forced Labour[7] reported that $99 billion came from commercial sexual exploitation. Polaris[8] reports that in the United States 100,000 children are estimated to be in the sex trade each year, and that the total number of victims is likely much larger when estimates of both adults and minors as well as sex trafficking and labor trafficking are aggregated. Estimates indicate that traffickers control an average of six victims and derive $150,000 from each victim per year. The sex trafficking industry is estimated to spend about $30 million on online advertising each year. These advertisements appear in hundreds of web sites that advertise escort services, massage parlors, etc. The total number of such advertisements is unknown, but our database of escort ads crawled from the most popular sites contains over 50 million ads.

The objective of our work is to create generic technology to enable rapid construction of knowledge graphs for specific domains together with query, visualization and analysis capabilities that enable end-users to solve complex problems. The challenge is to exploit all available sources, including web pages, document collections, databases, delimited text files, structured data such as XML or JSON, images, and videos. In this work we developed the technologies and their application to build a large knowledge graph for the human trafficking domain. This demonstration shows how the knowledge graph built using semantic technologies can be applied to combat human trafficking.

2 A Knowledge Graph for Human Trafficking

Figure 1 shows a screenshot of the DIG query interface. The interface paradigm is similar to that of popular web sites such as Amazon (`amazon.com`). Users can search using keywords, and can filter results by clicking on check-boxes to constrain the results. For example, the figure shows that the user clicked on "latina", so the results are filtered to contain only those with the selected ethnicity. The user interface issues queries to the ElasticSearch index, responding to queries over the 1.4 billion node graph in under 2 seconds.

The application of DIG to combat human trafficking is in use today. The system has currently been deployed to six law enforcement agencies and several NGOs (non-governmental organization) that are all using the tools in various

[7] http://bit.ly/1oa2cR3
[8] http://www.polarisproject.org/index.php

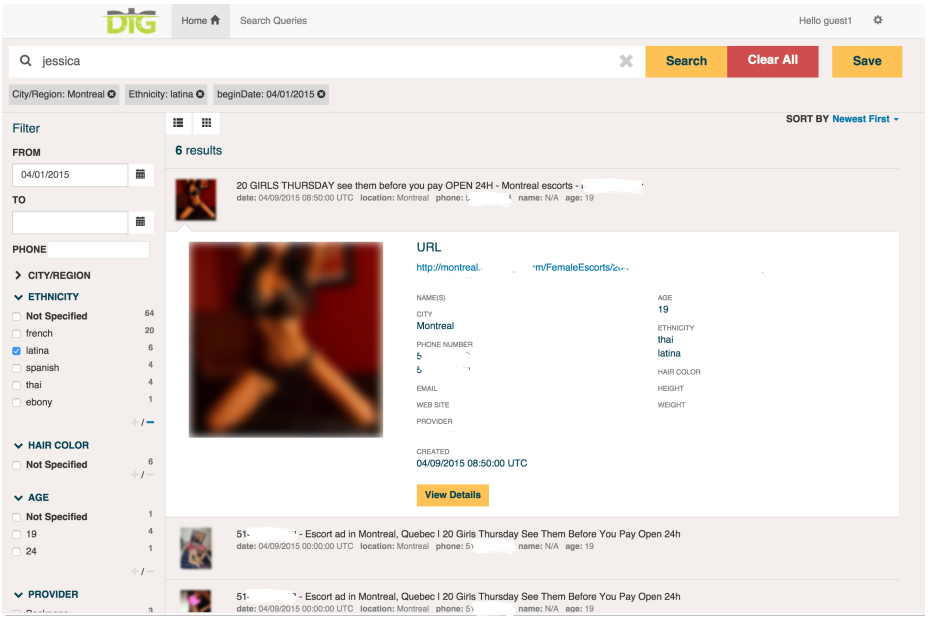

Fig. 1. Screenshot of DIG query interface showing results a query on the keyword "jessica", filtered by city/region, ethnicity and date to focus on a small number of ads

ways to fight human trafficking, such as by locating victims or researching organizations that engaging in human trafficking. The program manager for the project has also received requests from more than 100 other government agencies that are interested in using the tools produced under the DARPA Memex program, including DIG. We have had reports that the DIG tool has already been successfully used to identify several victims of human trafficking, but due to privacy concerns and the sensitivity of the topic, we have been asked not to reveal the law enforcement agencies involved or the details of any cases.

All of the data used in the deployed application comes from publicly available web sites that contain advertisements for services. The knowledge graph statistics on 30 April 2015 are the following:

- Number of ads: 52 million, new ads per day: 162,000, updated every hour
- Number of objects (RDF subjects): 1.4 billion
- Number of Feature objects: 222 million
- Number of phone numbers: 1.5 million

3 Demonstration

The full paper presented in the In Use track describes the end-to-end methods for building knowledge graphs from online sources. This demonstration will give people a sense for how such knowledge graphs can be used for a real-world

application and show the general query interface that we use for exploring a knowledge graph.

There are a number of use cases for the human trafficking knowledge graph and we will demonstrate how the system can be applied to several of these use cases. First, we will consider the case of a underage runaway that has been lured into the sex trade and show how the DIG knowledge graph can be used to locate such victims so they can be rescued. Second, we will consider the case of a organization involved in human trafficking and show how DIG can be used to research and find the individuals being controlled by such an organization. In the case of sex trafficking, such individuals are often moved from city to city, and DIG allows a law enforcement agency to find and track these individuals. The DIG knowledge graph allows law enforcement agencies to conduct the needed research to understand the full extent of the operations of a organization and to assemble a case to prosecute the individuals involved in those organizations.

4 Discussion

In this demonstration we will show how the knowledge graph built using the DIG system can be used for combating human trafficking. DIG can pull data from a combination of web pages and databases, extract, clean and integrate the data across sources, find similarities among entities, build a graph of all of the data, and then query the data to solve specific analytical problems.

DIG is not limited to this specific application and has already been applied to other problems including understanding the research trends in the field of material science, combating arms trafficking, and identifying patent trolls (also known as non-practicing entities). The underlying tools and technology are widely applicable and can be applied to many other applications, such as creating a knowledge graph of companies to build a accurate competitive landscape of companies based on their products and services or a knowledge graph of cultural heritage data that could be used by art historians. If there is interest, we can also demonstrate some of these other applications of the technology.

Acknowledgements This research is supported in part by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under contract number FA8750-14-C-0240, and in part by the National Science Foundation under Grant No. 1117913. Cloud computing resources were provided in part by Microsoft under a Microsoft Azure for Research Award. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, NSF, or the U.S. Government.