What are the Important Properties of an Entity?

Comparing Users and Knowledge Graph Point of View

Ahmad Assaf†, Ghislain A. Atemezing‡, Raphaël Troncy‡ and Elena Cabrio‡

†SAP Research, SAP Labs France SAS
Mougins, France
ahmad.assaf@sap.com

‡EURECOM
Sophia Antipolis, France
firstname.lastname@eurecom.fr

Entities are generally described in knowledge bases with a lot of properties, this is the case for DBpedia. It is, however, difficult to assess which ones are more "important" than others

Motivation

1. Question Answering system such as QakisMedia: http://qakis.org/
2. Second screen application for a TV program: http://www.linkedtv.eu/demos/linkednews/
3. Data augmentation in business intelligence applications

We assume in our experiments that the properties displayed for an entity are type and context-dependent (country, query, time, etc.) which affect the results

Reverse Engineering the Knowledge Graph

- Get all DBpedia concepts that have at least one instance which is owl:sameAs with a Freebase resource
- Filter out generic concepts by excluding those who are direct subclasses of owl:Thing since they will trigger ambiguous queries

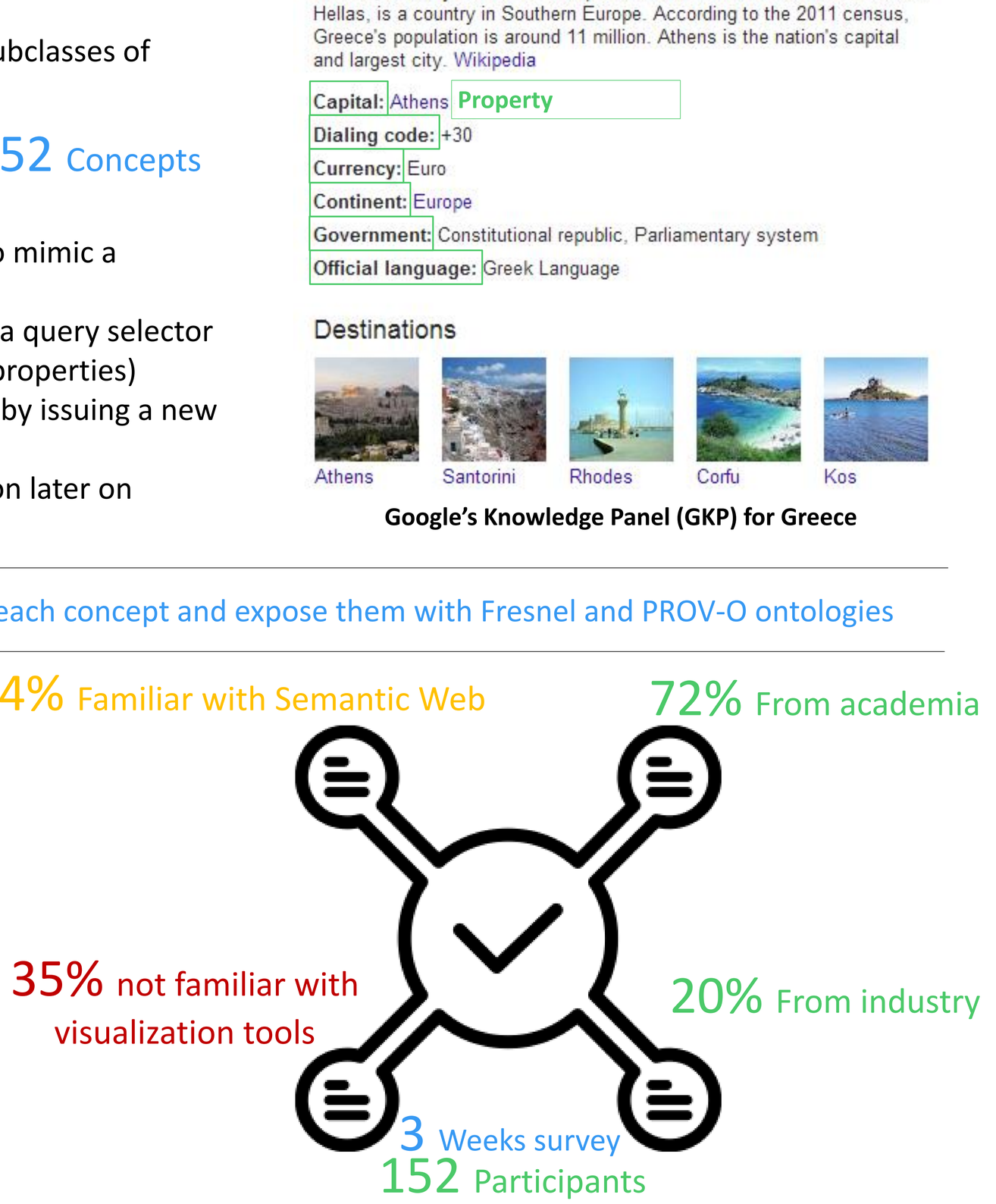
» 352 Concepts

- For each of these concepts, we retrieve 100 instances
- Google does not serve the GKP for all user agents and we had to mimic a browser behavior by setting a browser User Agent
- We use CSS selectors to extract data from a GKP. An example of a query selector is ._om (all elements with class name om -> this returns all the properties)
- If we do not find a GKP in a SERP, we disambiguate the instance by issuing a new query with the concept type attached
- If no GKP was found again, we capture that for manual inspection later on

Google's Knowledge Panel (GKP) for Greece

We aggregate the properties captured for all the instances of each concept and expose them with Fresnel and PROV-O ontologies

- We empirically define the most important properties when there is an agreement between one of the biggest knowledge base (Google) and users preferences
- We are aware that this knowledge is highly dynamic, the Google Knowledge Graph panel varies across top level countries and time
- We have provided the code that enables to perform new calculation at run time and we aim to study the temporal evolution of what are important properties on a longer period



By analyzing the results of the survey we discovered that concepts like Book and Museum are pretty stable (in agreement) compared to other concepts like Person/Agent
All results are reproducible from our code base at https://github.com/ahmadassaf/KBE