

Entities are Defined by Facts



Entities are Defined by Facts

- Repositories of facts containing this information can be found at many places, like data.worldbank.org, Wikipedia infoboxes etc.
- Countries are popular and finite, finding complete knowledge bases is possible.
- What about less popular entities?
 - What is the population of Arbit Apartments, Powai?
 - What is the GDP of Sugarcane Industry of India?
 - Percent of Internet users in Mumbai?

Entities are Defined by Facts

- Good news: the Web is huge, probably, there is some page which contains the information we are looking for.

Key Intuition

The way in which you express a fact about an entity depends on the fact, and not the entity.

- We may expect the sentence structure to be similar.
 - Population of India reached 1.3 billion, making it the second largest country in the world
 - Population of Arbit Apartments, Powai reached 1300

Problem Statement

- Given a repository of numerical facts about entities, can we train extractors that run over the web and pull similar facts about other entities?