



Задачи для практики

Владимир Батыгин

Разработчик

Отдел сбора структурированных данных

АУ, Санкт-Петербург, 22 сентября 2011

Сбор структурированных данных

Сбор данных из открытых веб источников

- Улучшает качество поиска
- Улучшает сервисы Яндекса: Яндекс.Карты, Яндекс.Авто и другие...

ОНТОЛОГИИ

Запросы к поиску:

- В каком городе родился Пушкин?
- Какая команда в 2008 году победила в кубке УЕФА?

ОНТОЛОГИИ

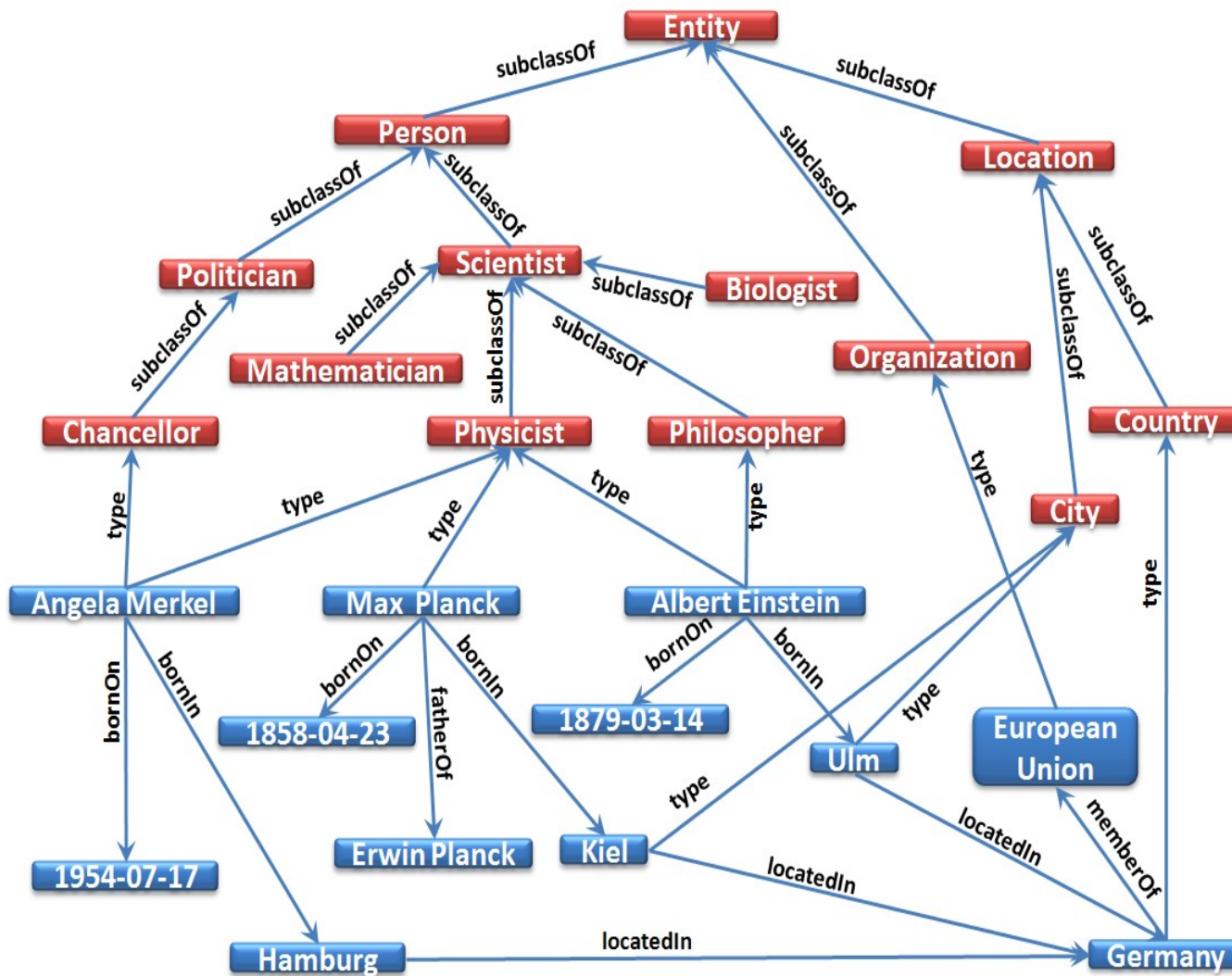
Запросы к поиску:

- В каком городе родился Пушкин?
- Какая команда в 2008 году победила в кубке УЕФА

Ответы:

- (Пушкин, Москва, родился в)
- (Зенит, 2008, победил в кубке УЕФА)

ОНТОЛОГИИ



Задача 1

Составление онтологии на основе википедии

- Выбрать тематику
- Выбрать набор отношений
- Извлечь данные из википедии в формате RDF (сущность, сущность, отношение)

Задача 2

Дополнение википедии данными из ОТКРЫТЫХ ИСТОЧНИКОВ

- Выбрать тематику
- Выбрать данные, которые нужно дополнить
- Написать робота, который дополняет информацию в википедии, используя данные из открытых веб источников



Батыгин Владимир

Разработчик

111033, Россия, Санкт-Петербург,
Свердловская наб., д. 44.

vbatygin@yandex-team.ru