

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ Н. Э. БАУМАНА
Факультет информатики и систем управления
Кафедра теоретической информатики и компьютерных технологий

Лабораторная работа №14
по курсу «Автоматическая обработка текстов»
«Извлечение именованных сущностей
в проекте iPavlov»

Выполнил:
студент группы ИУ9-11М
Беляев А. В.

Проверила:
Лукашевич Н. В.

Москва 2018

1 Цель работы

С помощью проекта iPavlov (demo.ipavlov.ai) извлечь именованные сущности из текста и проанализировать полученный результат.

2 Ход работы

Разбор текста про Илона Маска системой iPavlov представлен на Рис. ?? . Люди выделены зеленым цветом, организации – красным, локации – Желтым.

Илон Маск (Elon Musk) — Тони Старк нашего времени . Первая компания Zip2 , основанная совместно с братом братом Кимбалом Маском , разрабатывала интернет-версии городских гидов для газетной индустрии . Во время работы Zip2 компании удалось заключить контракты с такими изданиями , как The New York Times , Chicago Tribune и другими . На тот после запуска Илон убедил совет директоров отказаться от слияния с CitySearch и на протяжении существования компании хотел занять должность CEO . Но ни один из членов совета не поддержал его в этом . В 1999 году Маск стал сооснователем X.com — компании , которая специализировалась на платёжных услугах в сети . Спустя год после основания компания объединилась с Confinity и превратилась во всемирно известную платёжную систему платежей PayPal — переименование в PayPal состоялось в 2001 году . После продажи PayPal Илон Маск не стал тратить полученные средства на отдых — в этом же году была основана Space Explorataion Technologies Corporation , сокращённо SpaceX . Основной идеей компании стала разработка космических кораблей многоразового использования для замены существующих на тот момент шаттлов NASA . На текущий момент SpaceX стала одной из самых перспективных компаний в своём сегменте . Tesla Motors стала вторым серьёзным проектом Илона Маска , нацеленным на глобальные изменения в мире . Основанная в 2003 году компания Марином Эберхардом и Джейм Би Штробелем она финансировалась из личных средств сооснователей до привлечения Маска . В феврале 2004 года Tesla Motors подняла инвестиционный раунд « А » , а Маск занял место председателя совета директоров и CEO компании . Совместно с братьями Линдоном Райвом и Питером Райвом Илон Маск основал энергетическую компанию со штаб-квартирой в Сан-Матео . Преследуемый идеями изменения транспортного сообщения между Лос-Анджелесом и Сан-Франциско Илон Маск представил проект вакуумного поезда Hyperloop .

Рис. 1: Именованные сущности в тексте

3 Выводы

В ходе разбора текста системой iPavlov, большая часть именованных сущностей была распознана верно. Можно сделать следующие выводы:

- Имена людей были выбраны верно в 100% случаев. Предположение о том, что имена распознаются как именованные сущности на русском языке, а организации – на английском – не оправдались («Elon Musk» в начале текста).
- Верно разобраны имена как известных людей (Маск), так и менее известных (Штробель), так и вымышленных персонажей (Старк).
- Названия организаций разобраны по большей части верно.
- Должность исполнительного директора (CEO, Chief Executive Officer) была разобрана как название организации, по всей видимости, из-за заглавных букв. При этом аббревиатура широко известная и неправильная ее интерпретация – значительный минус системы.
- «Сан-Матео» в предложении «... штаб-квартира в Сан-Матео» был разобран не как локация, а как организация, несмотря на наличие слова «штаб-квартира» рядом.
- В некоторых случаях именованные сущности не разбирались вовсе, хотя следующее их упоминание было разобрано корректно («... PayPal – переименование в PayPal ...» – лишь второе вхождение разобрано. SpaceX был разобран верно во всех вхождениях). Это также значительный минус, т.к. система просто «проигнорировала» уже имеющиеся слова в некоторых случаях.
- Даты не были распознаны (хотя их распознавание и не заявлено), но для полного анализа например списка компаний, основанных Маском в хронологическом порядке, эта информация пригодилась бы.

Несмотря на перечисленные проблемы, точность работы системы достаточно высока.