

Софийски университет
„Св. Климент Охридски”

Дипломна работа

Катедра „Софтуерни технологии”



Прилагане на методи от машинния превод за изграждане на диалогова система за отговаряне на често задавани въпроси

Дипломант:

Мартин Боянов

Специалност:

**Извличане на информация и откриване на
знания**

Факултетен номер:

25187

Научни ръководители:

Д-р Преслав Наков, Проф. Иван Койчев

София 2017г.

Съдържание

1	Увод	2
1.1	Актуалност на проблема и мотивация	2
1.2	Цел и задачи на дипломната работа	2
1.3	Структура на дипломната работа	3
2	Преглед на предметната област	4
2.1	Въведение	4
2.2	Диалогови системи: мотивация и история	4
2.3	Машинен превод	4
2.4	Отговаряне на често задавани въпроси	4
3	Използвани технологии и платформи	5
3.1	intro	5
3.2	tensorflow	5
3.3	python	5
3.4	kelp	5
3.5	BPE	5
3.6	Data	5
3.6.1	UDC	5
3.6.2	QL	5
4	Реализация, експерименти и анализ на резултатите	6
4.1	Реализация	6
5	Заклучение	7
5.1	Обобщение на изпълнението на началните цели	7
6	Приложения	8

1 Увод

1.1 Актуалност на проблема и мотивация

Бързо разрастващата се сфера на мигновените съобщения (instant-messaging) поражда голям интерес за създаване на интелигентни системи за водене на диалог, още известни като чатботове.

Изграждането на такова приложение е трудна задача, поради липсата на структура в текста. Трудността се засилва и от факта, че при чат приложенията правописът и граматиката не са приоритет за потребителите. При зададен въпрос от потребител, системата трябва да генерира смислен и релевантен отговор.

Подобна задача се решава при машинния превод. При дадена поредица от думи, трябва да се генерира друга поредица от думи. Целта на тази дипломна работа е да се адаптира модел от машинния превод за изграждане на чатбот.

1.2 Цел и задачи на дипломната работа

Основната цел на тази работа е изследване на възможностите за изграждане на диалогова система базирана на модела seq2seq. Този модел е в основата на съвременните системи за машинен превод, използващи изкуствени невронни мрежи. Ще се изследват и възможностите за интегриране на модела в система за автоматично отговаряне на въпроси.

Задачи, произтичащи от целта:

- Обзор на подходите за решаване на този вид задачи
- Набавяне на набор от данни - Ubuntu Dataset Corpus и Qatar-Living
- Предварителна обработка на данните
- Имплементация и интеграция на seq2seq модела

-
- Анализ на процеса на машинно самообучение
 - Тестване, настройка и оценяване на работата на системата
 - Анализ на резултатите
 - Интеграция в системата за автоматично отговаряне на въпроси
 - Разработване на интерактивно демонстративно приложение

1.3 Структура на дипломната работа

2 Преглед на предметната област

2.1 Въведение

Изграждането на отворена диалогова система е предизвикателна задача, с която са се сблъскали много екипи. През годините са опитвани различни подходи за изграждане на подобна система и в тази глава ще се запознаем с основните методи и тяхната еволюция. Благодарение на прогреса в сферата на машинното самообучение и на нарастващата изчислителна мощ бяха постигнати подобрения в задачата за автоматичен машинен превод. Ще обобщим тяхната архитектура и ще коментираме приложимостта на подобен модел за изграждане на диалогова система. Ще разгледаме и проблема за отговаряне на често задавани въпроси и връзката му с диалоговите системи

<https://research.fb.com/the-long-game-towards-understanding-dialog/>

2.2 Диалогови системи: мотивация и история

-мотивация -ELIZA -PARRY? -ALICE -games -assistants

2.3 Машинен превод

- statistical machine translation - neural machine translation - seq2seq

2.4 Отговаряне на често задавани въпроси

3 Използвани технологии и платформи

3.1 intro

3.2 tensorflow

3.3 python

3.4 kelp

3.5 BPE

3.6 Data

3.6.1 UDC

3.6.2 QL

4 Реализация, експерименти и анализ на резултатите

4.1 Реализация

5 Заключение

5.1 Обобщение на изпълнението на началните цели

6 Приложения