

# Contents

<b>1</b>	<b>Motivation</b>	<b>2</b>
1.1	Motivation0 . . . . .	2
1.2	Motivation1 . . . . .	3
1.3	Motivation2 . . . . .	3
<b>2</b>	<b>Linear Function</b>	<b>3</b>
2.1	Slope-intercept form . . . . .	3
2.2	point-slope form . . . . .	3
<b>3</b>	<b>Quadratic Function</b>	<b>3</b>
3.1	something1 form . . . . .	3
3.2	something2 form . . . . .	3

# My preparation for thesis in L<sup>A</sup>T<sub>E</sub>X

Cesar Sinchiguano

February 26, 2019

## 1 Motivation

Besides the obvious application of WE in the search engine, there are also so-called vertical web services where WE is widely used too. Such websites allow to explore and work with the information on a specific topic. Such services usually do not produce the content by themselves, and rather they automatically collect the data from various sources, processes this data in some way and present to the user in the appealing form. Such services also called *aggregators*. A broad list of aggregators topics includes news, retail other advertisements, reviews, flight tickets, videos and pictures, recipes, social networks and many other. The list is very extensive, virtually for any subject discussed and presented on the Internet there are aggregators exist which try to grasp all other sources into one.

The relationship between aggregators and original source makers today reminds the problem of chicken and egg. That's why search engines are very accurate with automatic extraction at the moment of showing the result for the search query. If Google shows you the correct answer right when you type the question, the website with this particular answer where Google took the information from will lose the user. Since this site loses the user, it also loses monetizing and resources to produce the 'correct answers' in future; hence Google will lose the sources of correct answers. That's the reason why WE among Internet industry players is a controversial topic.

### 1.1 Motivation0

In the research area, WE is typically used in the tasks related to natural language processing and text mining. The reason for this is that the web page content is the semi-structured text mixed with the HTML markdown. Many research project centered around the analysis of various news sources as news articles, tweets, social networks, etc. Some of the possible tasks are sentiment analysis of news, topic recognition, summarization.

Some projects aim to build the system which collects huge amount of data, extract entities from them, determine relationships between them and my an-

swer on human-language question (IBM Watson [?], Calais by Reuters [?], Wolfram Language [?]).

The extraction of data from the web page is closely linked with the Document Object Model (DOM) processing and implies the understanding of the web page structure from the 'browser point of view'.

## **1.2 Motivation1**

In this way, WE is becoming a quite interesting, actual and challenging task. It includes both the technical background for the data retrieving and manipulation as well as theoretical background for the reasonable model building. This thesis includes all necessary parts of this task.

## **1.3 Motivation2**

Aggregators are usually very popular because they provide the user with the big database, rich functionality, flexibility and instant updates, what helps a user to save time. Due to popularity of aggregator the 'original' content maker is becoming popular too and therefore the content maker has resources to produce new interesting content. The other side of this practice is traffic reduction on the original content maker websites. It happens for example if the aggregator is unfair and doesn't show the source of information.

# **2 Linear Function**

## **2.1 Slope-intercept form**

The slope intercept form of the linear function is given by  $y = x + 2$ .

## **2.2 point-slope form**

# **3 Quadratic Function**

## **3.1 something1 form**

## **3.2 something2 form**