

# Usage of Regular Expressions in text processing

Dobrydnev Nikolai

November 2023

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Article Overview . . . . .	1
1.2	Regular Expression Definition . . . . .	2
<b>2</b>	<b>Pattern Matching and Searching</b>	<b>2</b>
2.1	Finding patterns in text . . . . .	2
2.2	Examples of pattern matching . . . . .	2
<b>3</b>	<b>Text Cleaning and Preprocessing</b>	<b>2</b>
3.1	Cleaning text with regex . . . . .	2
3.2	Real-world use cases . . . . .	2
<b>4</b>	<b>Practical Applications</b>	<b>2</b>
4.1	Data mining . . . . .	2
4.2	Text Analysis and Linguistic Applications . . . . .	2
4.3	Natural Language Processing . . . . .	2
<b>5</b>	<b>Conclusion</b>	<b>2</b>

## 1 Introduction

### 1.1 Article Overview

This article aims to provide information on the the utilization of *Regular Expressions* when writing, processing and retrieving data.

In this article I am going to define the term *Regular Expression*, explore the history of its development, examine algorithms it deploys and analyze its contribution to working with data.

## 1.2 Regular Expression Definition

Regular Expression (*RegEx*) is a set of characters based on strict rules that represents particular pattern to find in the text. It is widely used in finding numerous variations of strings that follow the same pattern.

One example of Regular Expression can be: `https?://[^\s]+`

While *RegEx* instantly identifies all the valid links (*http* or *https* protocols) in the text, *common search* (which implies matching sub string) requires no less than two queries for the same outcome.

Regular Expression is a beneficial tool to process any kind of text or data. Apart from its convenience, it provides an opportunity to create complex queries that cannot be achieved through the common search.

## 2 Pattern Matching and Searching

### 2.1 Finding patterns in text

### 2.2 Examples of pattern matching

## 3 Text Cleaning and Preprocessing

### 3.1 Cleaning text with regex

### 3.2 Real-world use cases

## 4 Practical Applications

### 4.1 Data mining

### 4.2 Text Analysis and Linguistic Applications

### 4.3 Natural Language Processing

## 5 Conclusion