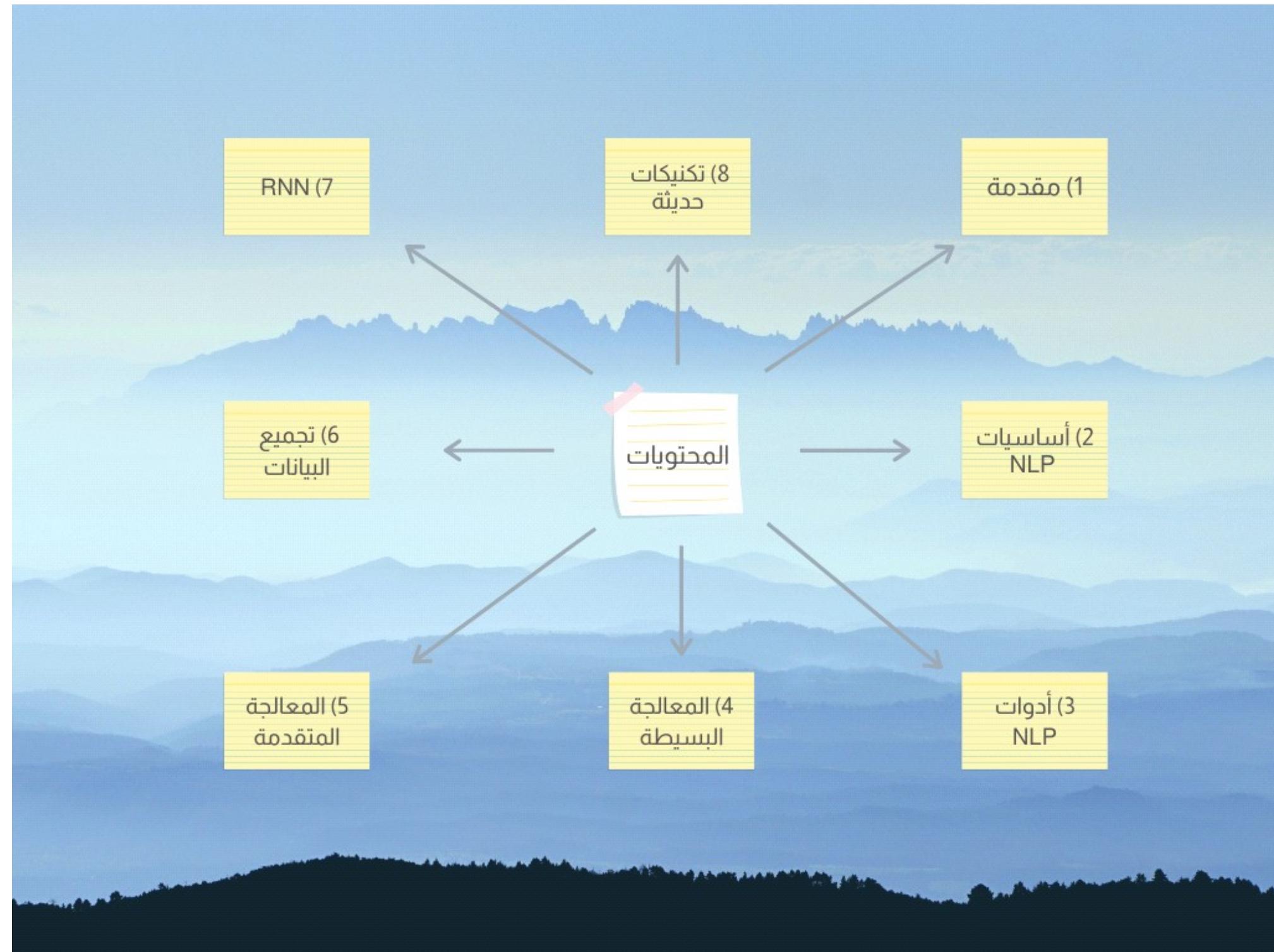


# NATURAL LANGUAGE PROCESSING

# المعالجة اللغوية الطبيعية



# المحتويات

				التطبيقات	العقبات و التحديات	NLP تاريخ ملفات pdf	ما هو NLP الملفات النصية	المحتويات المكتبات	1) مقدمة
T.Visualization	Syntactic Struc.	Matchers	Stopwords	NER	Stem & Lemm	POS	Sent. Segm.	Tokenization	2) أساسيات NLP
	Dist. Similarity	Text Similarity	TF-IDF	BOW	Word2Vec	T. Vectors	Word embed	Word Meaning	3) أدوات NLP
T. Generation	NGrams	Lexicons	GloVe	L. Modeling	NMF	LDA	T. Clustering	T. Classification	4) المعالجة البسيطة
	Summarization & Snippets		Ans. Questions		Auto Correct	Vader	Naïve Bayes	Sent. Analysis	5) المعالجة المتقدمة
Search Engine	Relative Extraction		Information Retrieval		Information Extraction		Data Scraping	Tweet Collecting	6) تجميع البيانات
					Rec NN\TNN	GRU	LSTM	Seq to Seq	7) RNN (
Chat Bot	Gensim	FastText	Bert	Transformer	Attention Model	T. Forcing	CNN	Word Cloud	8) تكنولوجيا حديثة

## القسم الثاني : أساسيات NLP

### الجزء الرابع : البحث في النصوص

نتناول الان مكتبة `re` و هي اختصار regular expression

و هي التي تقوم بعمل بحث كامل في النصوص ، لفلترة حروف معينة ، او رموز او ارقام ، و ذلك للبحث عن كلمات معينة

و المشكلة تكون في البحث عن نصوص ما لكن بصيغة معينة ، اي يكون لها شكل محدد ، مثل رقم تليفون - 605-542-5233

فهذا الرقم يتكون من 3 ارقام ثم - ثم 4 ارقام ، ونفس الامر يتكرر في صيغ مختلفة مثل الايميل او الرقم البريدي

فباستخدام مكتبة `re` يمكن البحث في اطار شكل معين

و يمكن استخدام هذا الموقع للتجريب عليه

<https://www.regexpal.com/>

و اهم القواعد الاساسية هي :

- تستخدم [ ] للتعبير داخلها عن اي حروف معينة
- اذا تم كتابة عدد من الحروف داخلها فهذا يعني كذا او كذا [as] يعني a او s
- لذا فهذا [0123456789] تعني اي رقم
- تستخدم ^ لحذف حرف معين
- تستخدم ? لجعل الحرف السابق لها اختياري قد يأتي او لا
- تستخدم \* للتعبير ان الحرف السابق لها قد يأتي مكررا او لا يأتي اصلا
- تستخدم + للتعبير ان الحرف السابق لها قد يأتي اكثر من مرة ( تختلف عن \* ان + تعني وجود الحرف مرة واحدة على الاقل )
- تستخدم . للتعبير ان في هذا المكان قد يأتي اي حرف او رقم
- تستخدم \$ مع حروف هنا [ ] للتعبير ان المقصود بها في اخر الكلمة

\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*

- Letters inside square brackets []

Pattern	Matches
[wW]oodchuck	Woodchuck, woodchuck
[1234567890]	Any digit

- Ranges [A-Z]

Pattern	Matches
[A-Z]	An upper case letter
[a-z]	A lower case letter
[0-9]	A single digit

Drenched Blossoms  
my beans were impatient  
Chapter 1: Down the Rabbit Hole

- Negations [ ^Ss ]

- Carat means negation only when first in []

Pattern	Matches	
[ ^A-Z ]	Not an upper case letter	Oyfn pripetchik
[ ^Ss ]	Neither 'S' nor 's'	I have no exquisite reason"
[ ^e^ ]	Neither e nor ^	Look here
a^b	The pattern a carat b	Look up <u>a^b</u> now

## Pattern

## Matches

groundhog | woodchuck

yours  
mine

a | b | c

= [abc]

[ gG ] roundhog | [ Ww ] oodchuck

Pattern	Matches		
colou?r	Optional previous char	<u>color</u>	<u>colour</u>
oo*h!	0 or more of previous char	<u>oh!</u> <u>ooh!</u>	<u>ooo!</u> <u>ooooh!</u>
o+h!	1 or more of previous char	<u>oh!</u> <u>ooh!</u>	<u>ooo!</u> <u>ooooh!</u>
baa+		<u>baa</u> <u>baaa</u>	<u>baaaa</u> <u>baaaaa</u>
beg.n		<u>begin</u> <u>begun</u>	<u>begun</u> <u>beg3n</u>

Pattern	Matches
<code>^ [ A-Z ]</code>	<u>Palo Alto</u>
<code>^ [ ^A-Za-z ]</code>	<u>1</u> <u>"Hello"</u>
<code>\. \$</code>	The end <u>.</u>
<code>. \$</code>	The end <u>?</u> The end <u>!</u>

و هنا امثلة متنوعة على نفس الموقف . بهذه الاختصارات

[a]

[A]

[Ww]

[a-z]

[B-S]

[A-Za-z]

[ !]

[1-4]

[0-9]

[^a-l]

[^ ]

[Tt]he

of|for

[Hh]as|[Hh]ave

[d\$]

[^A-Za-z]for[^A-Za-z]

Regular Expression

/[a]/g

Javascript flags

128 matches

Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

Substitution



Ads by Google

[Stop seeing this ad](#) [Why this ad? ▾](#)

Save this Regex

Name:

Description:

Save

Top Regular Expressions

Url checker with or without http:// or https://

Only letters and numbers

date format (yyyy-mm-dd)

Url Validation Regex | Regular Expression - Taha

match whole word

nginx test

special characters check

Extract String Between Two STRINGS

Match anything enclosed by square brackets.

Blocking site with unblocked games

Simple date dd/mm/yyyy

Find Substring within a string that begins and ends with parenthesis

Match anything after the specified

all except word

10-digit phone number with hyphens

Match if doesn't start with string

Java Variable

RegEx for Json

Find any word in a list of words

Regular Expression For Decimal Validation | Taha

Cheat Sheet

Character classes

. any character except newline

\w \d \s word, digit, whitespace

\W \D \S not word, digit, whitespace

[abc] any of a, b, or c

[^abc] not a, b, or c

[a-g] character between a & g

Anchors

**Regular Expression**

/[A]/g

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

← +

Ads by Google

Stop seeing this ad Why this ad? ▶

**Save this Regex**

Name:

Description:

Save

**Top Regular Expressions**

- [Url checker with or without http:// or https://](#)
- [Only letters and numbers](#)
- [date format \(yyyy-mm-dd\)](#)
- [Url Validation Regex | Regular Expression - Taha](#)
- [match whole word](#)
- [nginx test](#)
- [special characters check](#)
- [Extract String Between Two STRINGS](#)
- [Match anything enclosed by square brackets.](#)
- [Blocking site with unblocked games](#)
- [Simple date dd/mm/yyyy](#)
- [Find Substring within a string that begins and ends with parenthesis](#)
- [Match anything after the specified](#)
- [all except word](#)
- [10-digit phone number with hyphens](#)
- [Match if doesn't start with string](#)
- [Java Variable](#)
- [RegEx for Json](#)
- [Find any word in a list of words](#)

[Regular Expression For Decimal Validation | Taha](#)**Cheat Sheet****Character classes**

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

**Anchors**

**Regular Expression**

/ [Ww] /g

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?  
Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions  
What has the reaction been?  
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

←

Ads by Google

Stop seeing this ad Why this ad? ▶

**Save this Regex**

Name:

Description:

Save

**Top Regular Expressions**

- Url checker with or without http:// or https://
- Only letters and numbers
- date format (yyyy-mm-dd)
- Url Validation Regex | Regular Expression - Taha
- match whole word
- nginx test
- special characters check
- Extract String Between Two STRINGS
- Match anything enclosed by square brackets.
- Blocking site with unblocked games
- Simple date dd/mm/yyyy
- Find Substring within a string that begins and ends with parenthesis
- Match anything after the specified
- all except word
- 10-digit phone number with hyphens
- Match if doesn't start with string
- Java Variable
- RegEx for Json
- Find any word in a list of words
- Regular Expression For Decimal Validation | Taha

**Cheat Sheet****Character classes**

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

**Anchors**

Regular Expression

/[a-z]/g

Javascript flags

1686 matches

Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media caption Thousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

Substitution

Ads by Google

Stop seeing this ad

Why this ad? ▶

Save this Regex

Name:

Description:

Save

Top Regular Expressions

Url checker with or without http:// or https://

Only letters and numbers

Date format (yyyy-mm-dd)

Url Validation Regex | Regular Expression - Taha

Match whole word

Nginx test

Special characters check

Extract String Between Two STRINGS

Match anything enclosed by square brackets

Blocking site with unblocked games

Simple date dd/mm/yyyy

Find Substring within a string that begins and ends with parenthesis

Match anything after the specified

All except word

10-digit phone number with hyphens

Match if doesn't start with string

Java Variable

RegEx for JSON

Find any word in a list of words

Regular Expression For Decimal Validation | Taha

Cheat Sheet

Character classes

. any character except newline

\w \d \s word, digit, whitespace

\W \D \S not word, digit, whitespace

[abc] any of a, b, or c

[^abc] not a, b, or c

[a-g] character between a & g

Anchors

## Regular Expression

`/[B-S]/g`

Javascript flags

47 matches

## Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

## Substitution

←

Ads by Google

Stop seeing this ad

Why this ad? ▶

## Save this Regex

Name:

Description:

Save

## Top Regular Expressions

[Url checker with or without http:// or https://](#)[Only letters and numbers](#)[date format \(yyyy-mm-dd\)](#)[Url Validation Regex | Regular Expression - Taha](#)[match whole word](#)[nginx test](#)[special characters check](#)[Extract String Between Two STRINGS](#)[Match anything enclosed by square brackets.](#)[Blocking site with unblocked games](#)[Simple date dd/mm/yyyy](#)[Find Substring within a string that begins and ends with parenthesis](#)[Match anything after the specified](#)[all except word](#)[10-digit phone number with hyphens](#)[Match if doesn't start with string](#)[Java Variable](#)[RegEx for Json](#)[Find any word in a list of words](#)[Regular Expression For Decimal Validation | Taha](#)

## Cheat Sheet

## Character classes

[. any character except newline](#)[\w \d \s word, digit, whitespace](#)[\W \D \S not word, digit, whitespace](#)[\[abc\] any of a, b, or c](#)[\[^abc\] not a, b, or c](#)[\[a-g\] character between a & g](#)

## Anchors

Regular Expression

/[A-Za-z]/g

Javascript flags

1746 matches

Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

Substitution



Ads by Google

[Stop seeing this ad](#) Why this ad? ▶



Save this Regex

Name:

Description:

Save

Top Regular Expressions

Url checker with or without http:// or https://

Only letters and numbers

Date format (yyyy-mm-dd)

Url Validation Regex | Regular Expression - Taha

Match whole word

Nginx test

Special characters check

Extract String Between Two STRINGS

Match anything enclosed by square brackets.

Blocking site with unblocked games

Simple date dd/mm/yyyy

Find Substring within a string that begins and ends with parenthesis

Match anything after the specified

All except word

10-digit phone number with hyphens

Match if doesn't start with string

Java Variable

RegEx for Json

Find any word in a list of words

Regular Expression For Decimal Validation | Taha

Cheat Sheet

Character classes

. any character except newline

\w \d \s word, digit, whitespace

\W \D \S not word, digit, whitespace

[abc] any of a, b, or c

[^abc] not a, b, or c

[a-g] character between a & g

Anchors

**Regular Expression**

/[ !]/g

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag? Demonstrators bearing the flag of former Imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media caption Thousands took to the streets of Berlin to rally against coronavirus restrictions  
What has the reaction been?  
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

← Ads by Google Stop seeing this ad Why this ad? ▶

**Save this Regex**

Name:

Description:

Save

**Top Regular Expressions**

- [Url checker with or without http:// or https://](#)
- [Only letters and numbers](#)
- [date format \(yyyy-mm-dd\)](#)
- [Url Validation Regex | Regular Expression - Taha](#)
- [match whole word](#)
- [nginx test](#)
- [special characters check](#)
- [Extract String Between Two STRINGS](#)
- [Match anything enclosed by square brackets.](#)
- [Blocking site with unblocked games](#)
- [Simple date dd/mm/yyyy](#)
- [Find Substring within a string that begins and ends with parenthesis](#)
- [Match anything after the specified](#)
- [all except word](#)
- [10-digit phone number with hyphens](#)
- [Match if doesn't start with string](#)
- [Java Variable](#)
- [RegEx for Json](#)
- [Find any word in a list of words](#)
- [Regular Expression For Decimal Validation | Taha](#)

**Cheat Sheet****Character classes**

- `.` any character except newline
- `\w \d \s` word, digit, whitespace
- `\W \D \S` not word, digit, whitespace
- `[abc]` any of a, b, or c
- `[^abc]` not a, b, or c
- `[a-g]` character between a & g

**Anchors**

**Regular Expression**

Javascript
flags
2 matches

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?  
Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions  
What has the reaction been?  
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

← +

Ads by Google

[Stop seeing this ad](#) [Why this ad? ▶](#)

### Save this Regex

Name:

Description:

Save

### Top Regular Expressions

- [Url checker with or without http:// or https://](#)
- [Only letters and numbers](#)
- [date format \(yyyy-mm-dd\)](#)
- [Url Validation Regex | Regular Expression - Taha](#)
- [match whole word](#)
- [nginx test](#)
- [special characters check](#)
- [Extract String Between Two STRINGS](#)
- [Match anything enclosed by square brackets.](#)
- [Blocking site with unblocked games](#)
- [Simple date dd/mm/yyyy](#)
- [Find Substring within a string that begins and ends with parenthesis](#)
- [Match anything after the specified](#)
- [all except word](#)
- [10-digit phone number with hyphens](#)
- [Match if doesn't start with string](#)
- [Java Variable](#)
- [RegEx for Json](#)
- [Find any word in a list of words](#)
- [Regular Expression For Decimal Validation | Taha](#)

### Cheat Sheet

#### Character classes

- `.` any character except newline
- `\w \d \s` word, digit, whitespace
- `\W \D \S` not word, digit, whitespace
- `[abc]` any of a, b, or c
- `[^abc]` not a, b, or c
- `[a-g]` character between a & g

#### Anchors

## Regular Expression

`/[0-9]/g`

Javascript flags

7 matches

## Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

## Substitution



Ads by Google

Stop seeing this ad

Why this ad?

## Save this Regex

Name:

Description:

Save

## Top Regular Expressions

[Url checker with or without http:// or https://](#)[Only letters and numbers](#)[date format \(yyyy-mm-dd\)](#)[Url Validation Regex | Regular Expression - Taha](#)[match whole word](#)[nginx test](#)[special characters check](#)[Extract String Between Two STRINGS](#)[Match anything enclosed by square brackets.](#)[Blocking site with unblocked games](#)[Simple date dd/mm/yyyy](#)[Find Substring within a string that begins and ends with parenthesis](#)[Match anything after the specified](#)[all except word](#)[10-digit phone number with hyphens](#)[Match if doesn't start with string](#)[Java Variable](#)[RegEx for Json](#)[Find any word in a list of words](#)[Regular Expression For Decimal Validation | Taha](#)

## Cheat Sheet

## Character classes

. any character except newline

\w \d \s word, digit, whitespace

\W \D \S not word, digit, whitespace

[abc] any of a, b, or c

[^abc] not a, b, or c

[a-g] character between a & g

## Anchors

**RegEx Pal**  
From Dan's Tools

Regular Expression  Javascript  flags 1301 matches

**Test String**

```
An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?
Demonstrators bearing the flag of former imperial Germany, used by the Reichsbürger (Reich Citizens) far-right group, overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions
What has the reaction been?
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that
```

**Substitution**  +

Ads by Google  
[Stop seeing this ad](#) [Why this ad? ▶](#)

**Save this Regex**

Name:   
 Description:

**Top Regular Expressions**

- Url checker with or without http:// or https://
- Only letters and numbers
- date format (yyyy-mm-dd)
- Url Validation Regex | Regular Expression - Taha
- match whole word
- nginx test
- special characters check
- Extract String Between Two STRINGS
- Match anything enclosed by square brackets.
- Blocking site with unblocked games
- Simple date dd/mm/yyyy
- Find Substring within a string that begins and ends with parenthesis
- Match anything after the specified
- all except word
- 10-digit phone number with hyphens
- Match if doesn't start with string
- Java Variable
- RegEx for Json
- Find any word in a list of words
- Regular Expression For Decimal Validation | Taha

**Cheat Sheet**

**Character classes**

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

**Anchors**

**Regular Expression**

```
/[^ ]/g
```

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany – used by the Reichsbürger (Reich Citizens) far-right group – overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

←

Ads by Google

Stop seeing this ad Why this ad? ▶

**Save this Regex**

Name:

Description:

Save

**Top Regular Expressions**

- Url checker with or without http:// or https://
- Only letters and numbers
- date format (yyyy-mm-dd)
- Url Validation Regex | Regular Expression - Taha
- match whole word
- nginx test
- special characters check
- Extract String Between Two STRINGS
- Match anything enclosed by square brackets.
- Blocking site with unblocked games
- Simple date dd/mm/yyyy
- Find Substring within a string that begins and ends with parenthesis
- Match anything after the specified
- all except word
- 10-digit phone number with hyphens
- Match if doesn't start with string
- Java Variable
- RegEx for Json
- Find any word in a list of words
- Regular Expression For Decimal Validation | Taha

**Cheat Sheet****Character classes**

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

**Anchors**

## Regular Expression

`/[Tt]he/g`

Javascript flags

37 matches

## Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

## Substitution



Ads by Google

Stop seeing this ad

Why this ad? ▶

23

## Save this Regex

Name:

Description:

Save

## Top Regular Expressions

[Url checker with or without http:// or https://](#)[Only letters and numbers](#)[date format \(yyyy-mm-dd\)](#)[Url Validation Regex | Regular Expression - Taha](#)[match whole word](#)[nginx test](#)[special characters check](#)[Extract String Between Two STRINGS](#)[Match anything enclosed by square brackets.](#)[Blocking site with unblocked games](#)[Simple date dd/mm/yyyy](#)[Find Substring within a string that begins and ends with parenthesis](#)[Match anything after the specified](#)[all except word](#)[10-digit phone number with hyphens](#)[Match if doesn't start with string](#)[Java Variable](#)[RegEx for Json](#)[Find any word in a list of words](#)[Regular Expression For Decimal Validation | Taha](#)

## Cheat Sheet

## Character classes

[. any character except newline](#)[\w \d \s word, digit, whitespace](#)[\W \D \S not word, digit, whitespace](#)[\[abc\] any of a, b, or c](#)[\[^abc\] not a, b, or c](#)[\[a-g\] character between a & g](#)

## Anchors

**Regular Expression**

/of|for/g
Javascript
flags

**Test String**

```
An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely peaceful Berlin demonstration.

What happened at the Reichstag?
Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions
What has the reaction been?
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that
```

**Substitution**

← Ads by Google

Stop seeing this ad Why this ad? ▶

### Save this Regex

Name:

Description:

Save

### Top Regular Expressions

- Url checker with or without http:// or https://
- Only letters and numbers
- date format (yyyy-mm-dd)
- Url Validation Regex | Regular Expression - Taha
- match whole word
- nginx test
- special characters check
- Extract String Between Two STRINGS
- Match anything enclosed by square brackets.
- Blocking site with unblocked games
- Simple date dd/mm/yyyy
- Find Substring within a string that begins and ends with parenthesis
- Match anything after the specified
- all except word
- 10-digit phone number with hyphens
- Match if doesn't start with string
- Java Variable
- RegEx for Json
- Find any word in a list of words
- Regular Expression For Decimal Validation | Taha

### Cheat Sheet

#### Character classes

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

#### Anchors

Regular Expression

/ [Hh]as | [Hh]ave /g

Javascript flags

6 matches

Test String

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely have peaceful Berlin demonstration.

What Has happened at the Reichstag?

Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they Have could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions

What has the reaction been?

"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

Substitution

Ads by Google

Stop seeing this ad Why this ad? ▶

Save this Regex

Name:

Description:

Save

Top Regular Expressions

Url checker with or without http:// or https://

Only letters and numbers

Date format (yyyy-mm-dd)

Url Validation Regex | Regular Expression - Taha

Match whole word

Nginx test

Special characters check

Extract String Between Two STRINGS

Match anything enclosed by square brackets.

Blocking site with unblocked games

Simple date dd/mm/yyyy

Find Substring within a string that begins and ends with parenthesis

Match anything after the specified

All except word

10-digit phone number with hyphens

Match if doesn't start with string

Java Variable

RegEx for Json

Find any word in a list of words

Regular Expression For Decimal Validation | Taha

Cheat Sheet

Character classes

. any character except newline

\w \d \s word, digit, whitespace

\W \D \S not word, digit, whitespace

[abc] any of a, b, or c

[^abc] not a, b, or c

[a-g] character between a & g

Anchors

**Regular Expression**

/[d\$]/g

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely have peaceful Berlin demonstration.

What Has happened at the Reichstag?  
Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.

Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they Have could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions  
What has the reaction been?  
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

← +

Ads by Google

Stop seeing this ad Why this ad? ▶

**Save this Regex**

Name:

Description:

Save

**Top Regular Expressions**

- Url checker with or without http:// or https://
- Only letters and numbers
- date format (yyyy-mm-dd)
- Url Validation Regex | Regular Expression - Taha
- match whole word
- nginx test
- special characters check
- Extract String Between Two STRINGS
- Match anything enclosed by square brackets.
- Blocking site with unblocked games
- Simple date dd/mm/yyyy
- Find Substring within a string that begins and ends with parenthesis
- Match anything after the specified
- all except word
- 10-digit phone number with hyphens
- Match if doesn't start with string
- Java Variable
- RegEx for Json
- Find any word in a list of words
- Regular Expression For Decimal Validation | Taha

**Cheat Sheet****Character classes**

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

**Anchors**

**Regular Expression**

/[^A-Za-z]for[^A-Za-z]/g

**Test String**

An attempt to storm Germany's parliament during Saturday's big Berlin protest against Covid-19 restrictions has been condemned by politicians across the political spectrum.

Demonstrators, many with far-right sympathies, broke through a cordon reform and ran up the steps of the Reichstag before police dispersed them.

The interior minister said there should be "zero tolerance" for such behaviour.

Some 38,000 turned out for the wider, largely have peaceful Berlin demonstration.

What Has happened at the Reichstag?  
Demonstrators bearing the flag of former imperial Germany - used by the Reichsbürger (Reich Citizens) far-right group - overcame a handful of police to run to the building entrance.

Police put the number involved at several hundred.  
Scuffles broke out and the protesters were then overcome by police using pepper spray. Several people were arrested.

Police rejected criticism of their small deployment, saying they have could not "be everywhere all the time".

Media captionThousands took to the streets of Berlin to rally against coronavirus restrictions  
What has the reaction been?  
"The Reichstag is the domain of our parliament and the symbolic centre of our democracy. It is unbearable that

**Substitution**

← Ads by Google Stop seeing this ad Why this ad? ▶

**Save this Regex**

Name:

Description:

Save

**Top Regular Expressions**

- [Url checker with or without http:// or https://](#)
- [Only letters and numbers](#)
- [date format \(yyyy-mm-dd\)](#)
- [Url Validation Regex | Regular Expression - Taha](#)
- [match whole word](#)
- [nginx test](#)
- [special characters check](#)
- [Extract String Between Two STRINGS](#)
- [Match anything enclosed by square brackets.](#)
- [Blocking site with unblocked games](#)
- [Simple date dd/mm/yyyy](#)
- [Find Substring within a string that begins and ends with parenthesis](#)
- [Match anything after the specified](#)
- [all except word](#)
- [10-digit phone number with hyphens](#)
- [Match if doesn't start with string](#)
- [Java Variable](#)
- [RegEx for Json](#)
- [Find any word in a list of words](#)
- [Regular Expression For Decimal Validation | Taha](#)

**Cheat Sheet****Character classes**

- . any character except newline
- \w \d \s word, digit, whitespace
- \W \D \S not word, digit, whitespace
- [abc] any of a, b, or c
- [^abc] not a, b, or c
- [a-g] character between a & g

**Anchors**

و هنا نماذج اخرى :

```
import re
```

و يمكن البدء بالاستخدام العادي للمكتبة عبر التعامل مع أمر : `search` والذي يقوم بالبحث في النص عن كلمة معينة

فإذا كان لدينا كلمة معينة هي :

```
pattern = 'phone'
```

و النص الكامل هو :

```
text = "The agent's phone number is 408-555-1234. Call soon!"
```

فيمكن استخدامها هكذا

```
match = re.search(pattern,text)
```

وامر سبان يأتي برقم بداية و نهاية الكلمة

`match.span()`

او يمكن ايجادهما باوامر مختلفة

`match.start()` , `match.end()`

ولو تكررت الكلمة ، فسيأتي بالاولى فقط ، و اذا اردنا كل النتائج نستخدم :

`text = ""`

A telephone is a telecommunications device that permits two or more users to conduct a conversation when they are too far apart to be heard directly. A telephone converts sound, typically and most efficiently the human voice,

into electronic signals that are transmitted via cables and other communication channels to another telephone which

reproduces the sound to the receiving user. The term is derived from Greek: τῆλε (tēle, far) and φωνή (phōnē, voice),

together meaning distant voice. A common short form of the term is phone, which came into use almost immediately after

the first patent was issued.[1] ""

`matches = re.findall(pattern,text)`

فتكون النتيجة هي لیست فيها العناصر نفسها و عبر ايجاد طولها نعرف عدد المرات

`len(matches)`

او قد نستخدم امر `finditer` لعمل فور لها

```
for match in re.finditer("phone",text):  
    print(match.span())
```

و نفس الأمر في اللغة العربية

`pattern = 'نسبة'`

`text = ""`

ورقة أينشتاين العلمية الثالثة كانت عن "نظرية النسبية الخاصة"، اقترحاها ونشرها في 26 سبتمبر 1905 بعنوان "الإلكتروديناميک للأجسام المتحركة".

فتناولت الورقة الزمان، والمكان، والكتلة، والطاقة، وأسهمت نظرية أينشتاين بإزالة الغموض الذي نجم عن التجربة الشهيرة التي أجرتها الأمريكيةان

الفيزيائي ألبرت ميكلسون والكيميائي إدوارد مورلي) أواخر القرن التاسع عشر في عام 1887، فقد أثبتت أينشتاين أن موجات الضوء تستطيع أن تنتشر في الخلاء

دون الحاجة لوجود وسط أو مجال، على خلاف الموجات الأخرى المعروفة التي تحتاج إلى وسط تنتشر فيه كالهواء أو الماء وأن سرعة الضوء هي سرعة ثابتة وليس نسبية مع حركة المراقب (الملاحظ)، وتجر الإشارة إلى أن نظرية أينشتاين تلك تناقضت بشكل كلي مع استنتاجات إسحاق نيوتن". جاءت تسمية النظرية بال خاصة للتferiq بينها وبين نظرية أينشتاين اللاحقة التي سميت بالنسبة العامة"

اعتباراً من اليوم، النسبة الخاصة هي النموذج الأكثر دقة للحركة بأي سرعة عندما تكون تأثيرات الجاذبية ضئيلة. ومع ذلك، لا يزال نموذج ميكانيكا نيوتن مفيداً كقياس سرعات صغيرة بالنسبة لسرعة الضوء، نظراً لبساطته ودقته العالية في نطاقه.

三

```
match = re.search(pattern,text)  
match.span()  
match.start() , match.end()
```

```
matches = re.findall(pattern,text)
len(matches)
for match in re.finditer(pattern,text):
    print(match.span())
```

لكن ما سبق كان بالبحث عن الكلمة معينة ، ماذًا عن البحث عن صيغة معينة بشكل عام

يتم هذا اولاً عبر التعرف على الحروف الخاصة التي تأتي بعد السلاش \ و التي يكون لها معنى محدد

Character	Description	Example Pattern Code	Exammmple Match
\d	A digit	file_\d\d	file_25
\w	Alphanumeric	\w-\w\w\w\w	A-b_1
\s	White space	a\s\bs\sc	a b c
\D	A non digit	\D\D\D	ABC
\W	Non-alphanumeric	\W\W\W\W\W\W\W	*-+=)
\S	Non-whitespace	\S\S\S\S\S\S	Yoyo

Character	Description	Example Pattern Code	Exammmple Match
+	Occurs one or more times	\w-\w+	A-b1_1
{3}	Occurs exactly 3 times	\D{3}	abc
{2,4}	Occurs 2 to 4 times	\d{2,4}	123
{3,}	Occurs 3 or more	\w{3,}	anycharacters
\*	Occurs zero or more times	A\*B\*C*	AAACC
?	Once or none	plurals?	plural

و يتم استخدام الحروف بهذه الطريقة للبحث عن كلمات باي صيغة مطلوبة

فلو كان لدينا نص فيه رقم تليفون غير معروف :

text = "My telephone number is 408-555-1234"

فيمكن البحث عنه بكتابه صيغته

```
phone = re.search(r'\d\d\d-\d\d\d-\d\d\d\d',text)
```

و عبر اختيار ميثود جروب سيقوم باظهاره

```
phone.group()
```

او قد نقوم باستخدام الاختصارات

```
re.search(r'\d{3}-\d{3}-\d{4}',text)
```

و ستكون بنفس النتيجة

و يمكن اثناء البحث ان تكون النتيجة الظاهرة جزء محدد و ليس النتيجة كلها

و هي تتم عبر عمل اقواس () حول الحروف هكذا :

```
phone_pattern =(r'(\d{3})-(\d{3})-(\d{4})')
```

```
results = re.search(phone_pattern,text)
```

و اذا استخدمنا امر جروب سياتي بالنتيجة كلها

```
results.group()
```

اما اذا استخدمنا رقم معين داخل جروب , فسياتي بالجزء الاول او الثاني او الثالث منها

```
results.group(1), results.group(2) , results.group(3)
```

و اذا استخدمنا رقم اكبر سيقوم بعمل خطأ

```
results.group(4)
```

كما يمكن استخدام | للدلالة على ( او )

```
re.search(r"man|woman","This man was here.")
```

```
re.search(r"man|woman","This woman was here.")
```

\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*

ايضا يمكن استخدام النقطة . لتحديد حرف غير معلوم , فهنا نبحث عن كلمة من ثلاث حروف اخر حرفين هما at

```
re.findall(r".at","The cat in the hat sat here.")
```

ولو كانت الكلمة اكبر من 3 حروف فسيقوم بايجاد 3 حروف منها فقط ، لأن هناك نقطة واحدة

```
re.findall(r".at","The bat went splat")
```

و اذا قمنا باعطاءه نقطتين فسيتعامل مع الكلمات ذات الخمس حروف

```
re.findall(r"...at","The bat went splat")
```

و اذا اردت منه اي حدد من الحروف المجهولة , تقوم باستخدام \l للدالة علي حروف ليست سبيس , مع علامة + لتحديد ان اي رقم موجود

```
re.findall(r'\S+at',"The bat went splat")
```

كما يتم استخدام \$ لتحديد ان الحروف المطلوب هو في النهاية ، فإذا استخدمناها مع d فيكون المطلوب هو اخر رقم

```
re.findall(r'\d$','This ends with a number 28')
```

او استخدام <sup>٨</sup> للدلالة على البداية

```
re.findall(r'^\d','1 is the loneliest number.')
```

[ ] و يمكن عمل بحث عكسي ، اي البحث عن اي شئ الا كذا ، عبر استخدام [ ]

```
phrase = "there are 3 numbers 34 inside 5 this sentence." re.findall(r'^\d+',phrase)
```

## فسياتي بكل الحروف عدا الارقام

و يمكن استخدامها لازالة كل حروف الترقيم هكذا :

```
test_phrase = 'This is a string! But it has punctuation. How can we remove it?'
```

```
re.findall('[^!.? ]+',test_phrase)
clean = ''.join(re.findall('[^!.? ]+',test_phrase))
print(clean)
```

مثل تاني للبحث عن اي حروف w ثم شرطة - ثم حروف  
text = 'Only find the hyphen-words in this sentence. But you do not know how long-ish they are'

```
re.findall(r'[\w]+-[\w]+',text)
```

و نفس الأمر في اللغة العربية

رقم هاتفي هو 9965-986-586 و رقم هاتف محمد هو 9785-476-826  
6855-006'

```
phone = re.search(r'\d\d\d-\d\d\d-\d\d\d\d',text)
```

```
phone.group()
```

```
re.search(r'\d{3}-\d{3}-\d{4}',text)
```

```
phone_pattern =(r'(\d{3})-(\d{3})-(\d{4})')
```

```
results = re.search(phone_pattern,text)
```

```
results.group()
```

```
results.group(1), results.group(2) , results.group(3)
```

```
results.group(4)
```

```
text1 = "ذهب محمد على الجامعة"
```

```
text2 = "استيقظت مني من النوم"
```

```
re.search(r"محمد|مني",text1)
```

```
re.search(r"محمد|مني",text2)
```

'ذهبت الطالبات لحضور عدد من المحاضرات في جامعاتهن' = text1

re.findall(r".ات.",text1)

'ذهبت الطالبات لحضور عدد من المحاضرات في جامعاتهن' = text1

re.findall(r"..ات.",text1)

'ذهبت الطالبات لحضور عدد من المحاضرات في جامعاتهن' = text1

re.findall(r"\S+ات",text1)

'قاد جورج مسافة 50 كيلو متر من الساعة 8 الى الساعة 10' = phrase

re.findall(r'^\d+',phrase)

' . ماذا ؟ ؟ ؟ كيف تجرؤ !! ، هل تخيل عاقبة هذا الأمر ؟ ؟ ، ويل له ' = test\_phrase

re.findall('^\?.!.',test\_phrase)

```
clean = ''.join(re.findall('[^؟،!۔]+',test_phrase))
print(clean)
```

text = 'ذهبت الى مدينة نيويورك ، و منها الى سان-فرانسيسكو'

```
re.findall(r'[\w]+-[\\w]+',text)
```