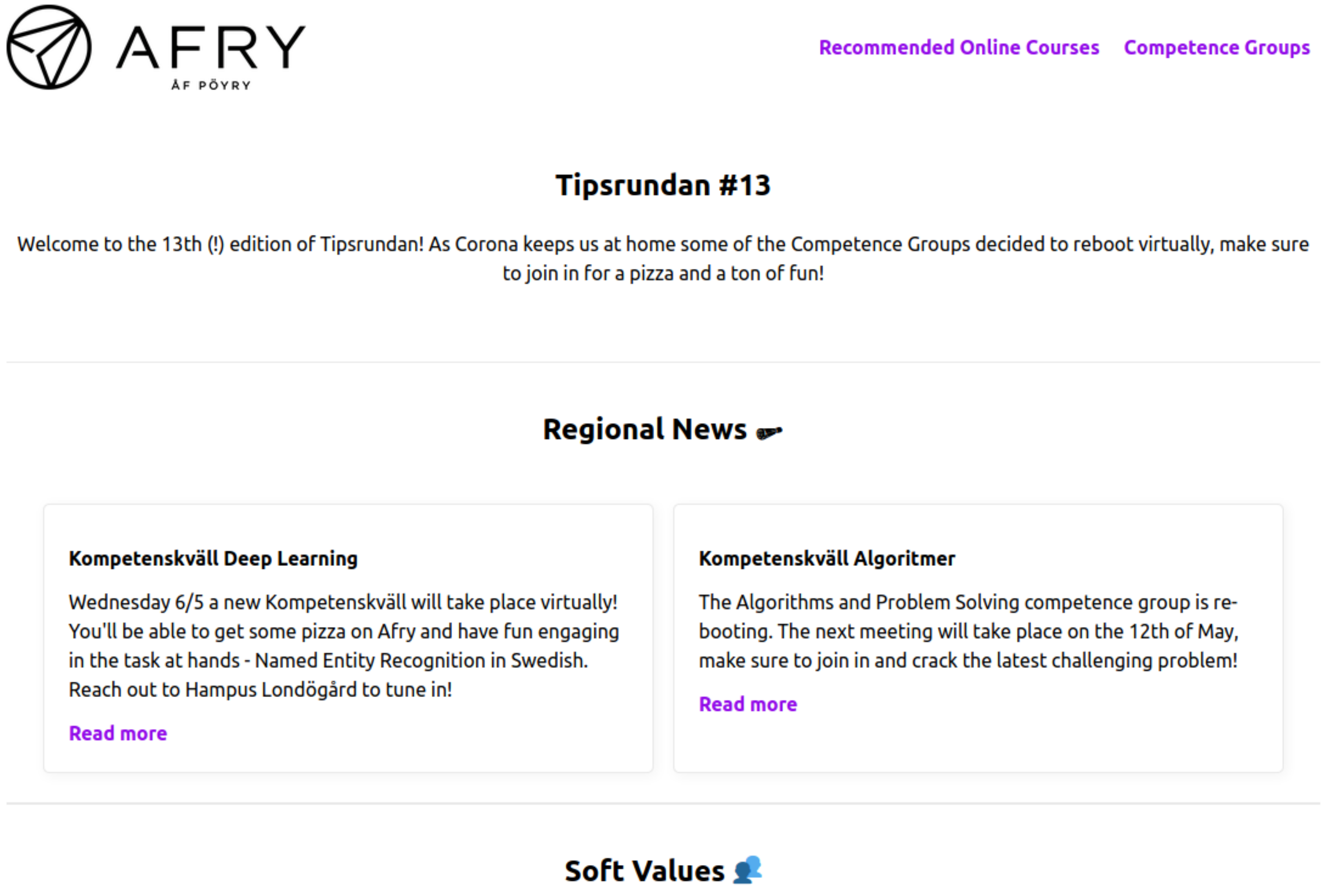
WHENEVER I LEARN A NEW SKILL I CONCOCT ELABORATE FANTASY SCENARIOS WHERE IT LETS ME SAVE THE DAY.
OH NO! THE KILLER MUST HAVE FOLLOWED HER ON VACATION!
BUT TO FIND THEM WE'D HAVE TO SEARCH THROUGH 200 MB OF EMAILS LOOKING FOR SOMETHING FORMATTED LIKE AN ADDRESS!
IT'S HOPELESS!
EVERYBODY STAND BACK.
I KNOW REGULAR EXPRESSIONS.
PERL!
TAP TAP

http://xkcd.com/208/

Afry IT South Competence

- Tipsrundan



Tipsrundan #13

Welcome to the 13th (!) edition of Tipsrundan! As Corona keeps us at home some of the Competence Groups decided to reboot virtually, make sure to join in for a pizza and a ton of fun!

Regional News 🔭

Kompetenskväll Deep Learning

Wednesday 6/5 a new Kompetenskväll will take place virtually! You'll be able to get some pizza on Afry and have fun engaging in the task at hands - Named Entity Recognition in Swedish. Reach out to Hampus Londögård to tune in!

Read more

Kompetenskväll Algoritmer

The Algorithms and Problem Solving competence group is re-booting. The next meeting will take place on the 12th of May, make sure to join in and crack the latest challenging problem!

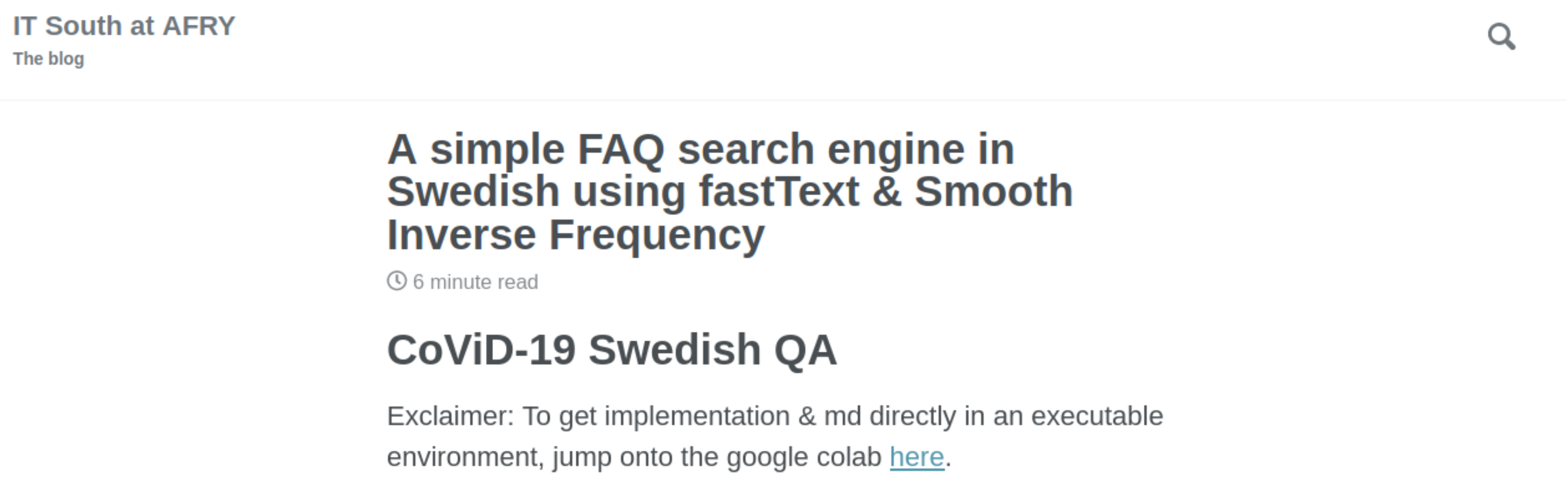
Read more

Soft Values 👥

buitsouth.slack.com

Afry IT South Competence

- Tipsrundan
- Afry IT South Blog - afry-south.github.io



IT South at AFRY

The blog

A simple FAQ search engine in Swedish using fastText & Smooth Inverse Frequency

6 minute read

CoViD-19 Swedish QA

Exclaimer: To get implementation & md directly in an executable environment, jump onto the google colab here.

buitsouth.slack.com

Afry IT South Competence

- Tipsrundan
- Afry IT South Blog - afry-south.github.io
- Kompetenskvällar
    - Machine Learning, Algorithms & Front-End (currently)

Afry IT South Competence

- Tipsrundan
- Afry IT South Blog - afry-south.github.io
- Kompetenskvällar
  - Machine Learning, Algorithms & Front-End (currently)
- Kompetenslunch / Competence Sharing Lunch

buitsouth.slack.com

Afry IT South Competence

- Tipsrundan
- Afry IT South Blog - afry-south.github.io
- Kompetenskvällar
  - Machine Learning, Algorithms & Front-End (currently)
- Kompetenslunch / Competence Sharing Lunch
- & more

Responsible: Hampus Londögård & Hassan Ftouni, & all of you.

buitsouth.slack.com

Don't forget about Slack!

AFRY BU IT South
buitsouth.slack.com

Regex 101

Taking search to the next level

Hampus Londögård

AFRY

Searching

Searching

- Find personal numbers

Searching

- Find personal numbers
- Validate passwords

Searching

- Find personal numbers
- Validate passwords
- Normalize input

Searching

- Find personal numbers
- Validate passwords
- Normalize input
- Extracting data from logs

What is Regex?

Regular Expression

Regular Expression

- A **Domain Specific Language**

Regular Expression

- A **Domain Specific Language**
- Sequence of character that define a **search pattern**

Regular Expression

A **DOMAIN SPECIFIC LANGUAGE**

Regular Expression

A **DOMAIN SPECIFIC LANGUAGE**

SEQUENCE OF CHARACTER THAT DEFINE A **SEARCH PATTERN**

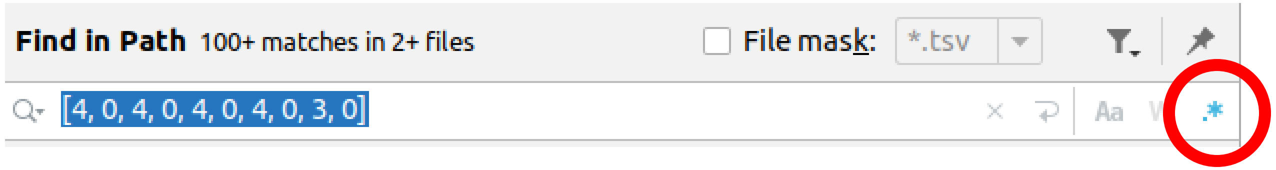
Regular Expression

A **DOMAIN SPECIFIC LANGUAGE**

SEQUENCE OF CHARACTER THAT DEFINE A **SEARCH PATTERN**

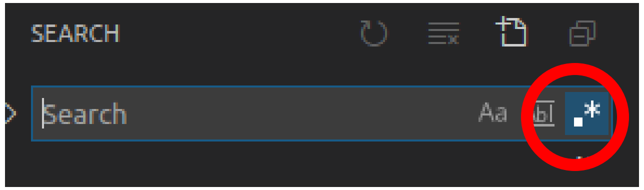
WIDELY USED

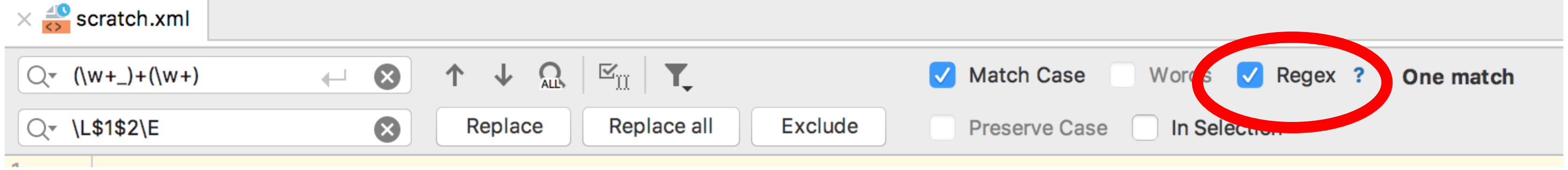
Most of todays editors support Regex



□ File mask: *.tsv

Search: [4, 0, 4, 0, 4, 0, 4, 0, 3, 0]

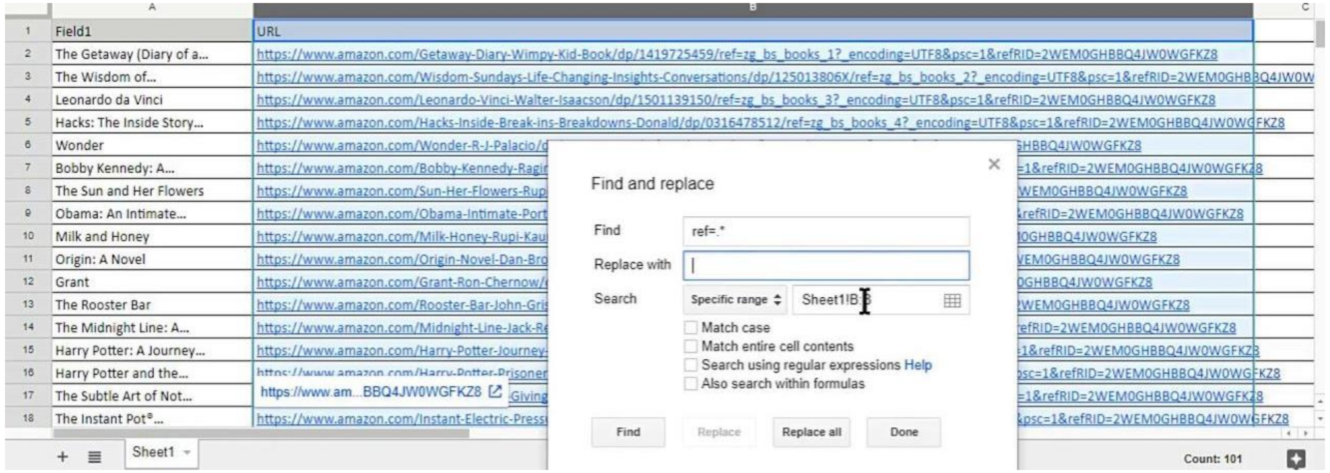




Find: (\w+_)+(\w+)

Replace: \L$1$2\E

☑ Match Case □ Words ☑ Regex ? □ Preserve Case □ In Selection



	A	B
1	Field1	URL
2	The Getaway (Diary of a...	https://www.amazon.com/Getaway-Diary-Wimpy-Kid-Book/dp/1419725459/ref=zg_bs_books_1?_encoding=UTF8&psc=1&refRID=2WEM0GHBBQ4JW0WGFKZ8
3	The Wisdom of...	https://www.amazon.com/Wisdom-Sundays-Life-Changing-Insights-Conversations/dp/125013806X/ref=zg_bs_books_2?_encoding=UTF8&psc=1&refRID=2WEM0GHBBQ4JW0W
4	Leonardo da Vinci	https://www.amazon.com/Leonardo-Vinci-Walter-Isaacson/dp/1501139150/ref=zg_bs_books_3?_encoding=UTF8&psc=1&refRID=2WEM0GHBBQ4JW0WGFKZ8
5	Hacks: The Inside Story...	https://www.amazon.com/Hacks-Inside-Break-ins-Breakdowns-Donald/dp/0316478512/ref=zg_bs_books_4?_encoding=UTF8&psc=1&refRID=2WEM0GHBBQ4JW0WGFKZ8
6	Wonder	https://www.amazon.com/Wonder-R-J-Palacio/d GHBBQ4JW0WGFKZ8
7	Bobby Kennedy: A...	https://www.amazon.com/Bobby-Kennedy-Ragi =1&refRID=2WEM0GHBBQ4JW0WGFKZ8
8	The Sun and Her Flowers	https://www.amazon.com/Sun-Her-Flowers-Rup WEM0GHBBQ4JW0WGFKZ8
9	Obama: An Intimate...	https://www.amazon.com/Obama-Intimate-Port &refRID=2WEM0GHBBQ4JW0WGFKZ8
10	Milk and Honey	https://www.amazon.com/Milk-Honey-Rupi-Kau 0GHBBQ4JW0WGFKZ8
11	Origin: A Novel	https://www.amazon.com/Origin-Novel-Dan-Bro VEM0GHBBQ4JW0WGFKZ8
12	Grant	https://www.amazon.com/Grant-Ron-Chernow/ 0GHBBQ4JW0WGFKZ8
13	The Rooster Bar	https://www.amazon.com/Rooster-Bar-John-Gri 2WEM0GHBBQ4JW0WGFKZ8
14	The Midnight Line: A...	https://www.amazon.com/Midnight-Line-Jack-Re efRID=2WEM0GHBBQ4JW0WGFKZ8
15	Harry Potter: A Journey...	https://www.amazon.com/Harry-Potter-Journey- 1&refRID=2WEM0GHBBQ4JW0WGFKZ8
16	Harry Potter and the...	https://www.amazon.com/Harry-Potter-Prisoner psc=1&refRID=2WEM0GHBBQ4JW0WGFKZ8
17	The Subtle Art of Not...	https://www.am...BBQ4JW0WGFKZ8 -Giving =1&refRID=2WEM0GHBBQ4JW0WGFKZ8
18	The Instant Pot®...	https://www.amazon.com/Instant-Electric-Press &psc=1&refRID=2WEM0GHBBQ4JW0WGFKZ8

Find and replace

Find: ref=.*

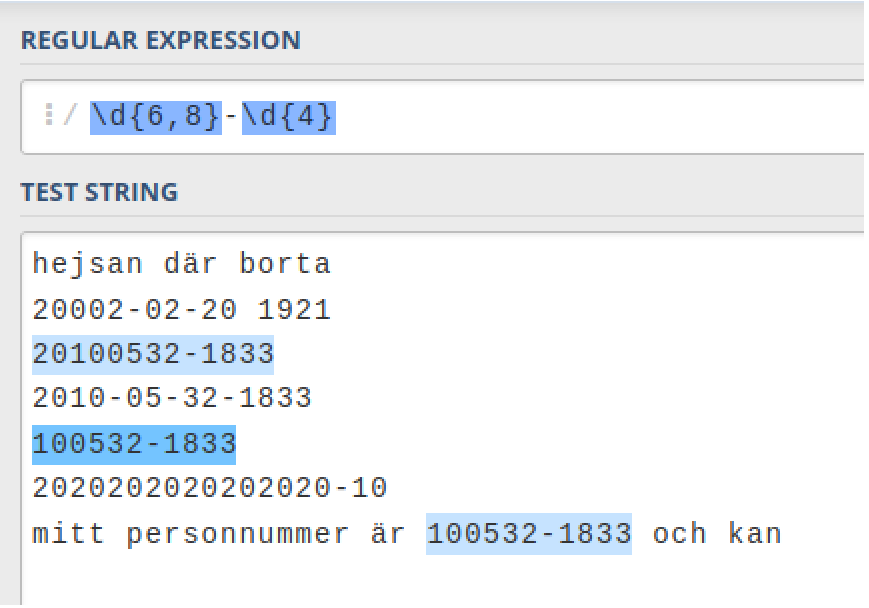
Replace with:

Search: Specific range Sheet1!B:B

□ Match case
□ Match entire cell contents
□ Search using regular expressions Help
□ Also search within formulas

Show me some regex!

Starting with some simple techniques



REGULAR EXPRESSION

```
/ \d{6,8}-\d{4}
```

TEST STRING

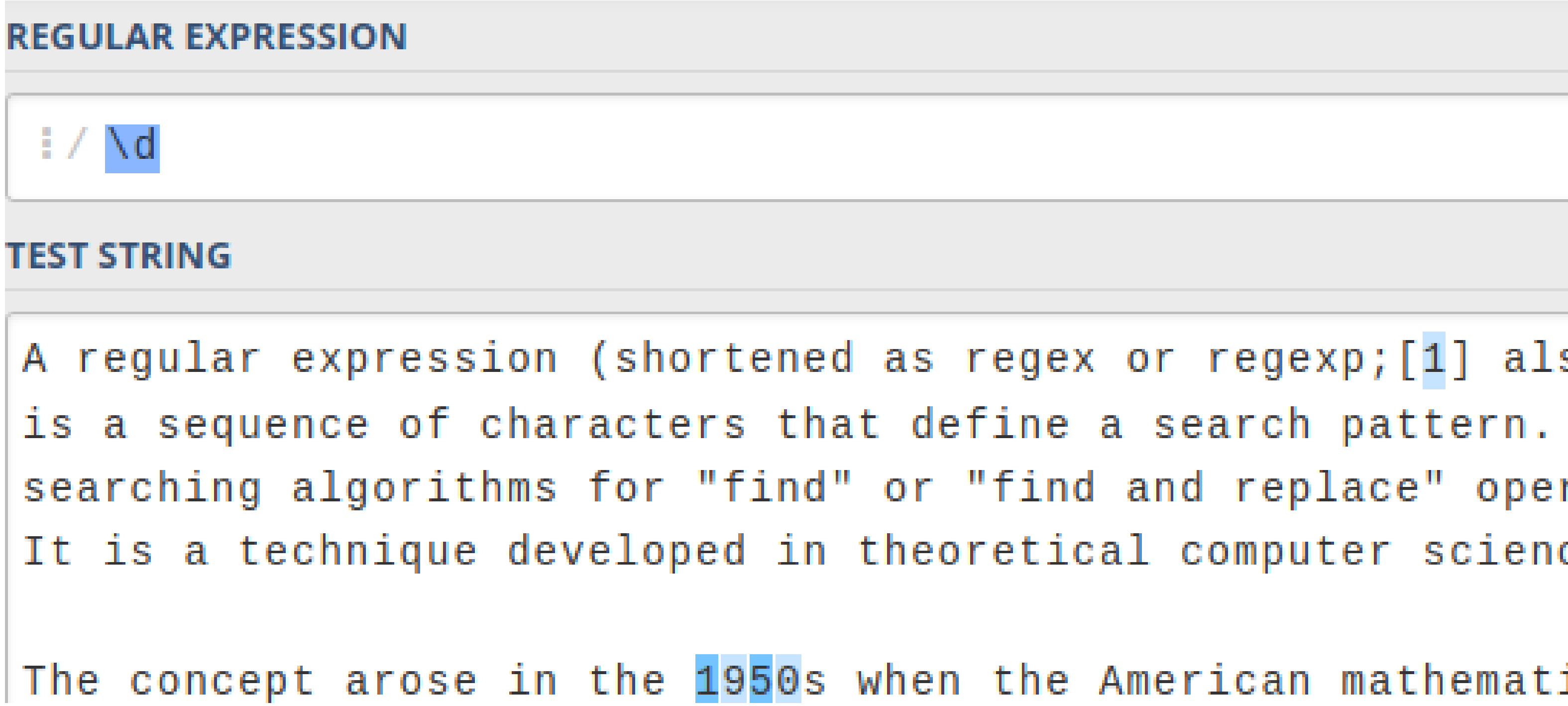
```
hejsan där borta
20002-02-20 1921
20100532-1833
2010-05-32-1833
100532-1833
2020202020202020-10
mitt personnummer är 100532-1833 och kan
```

The basics

Different types

\d digit

\D non-digit



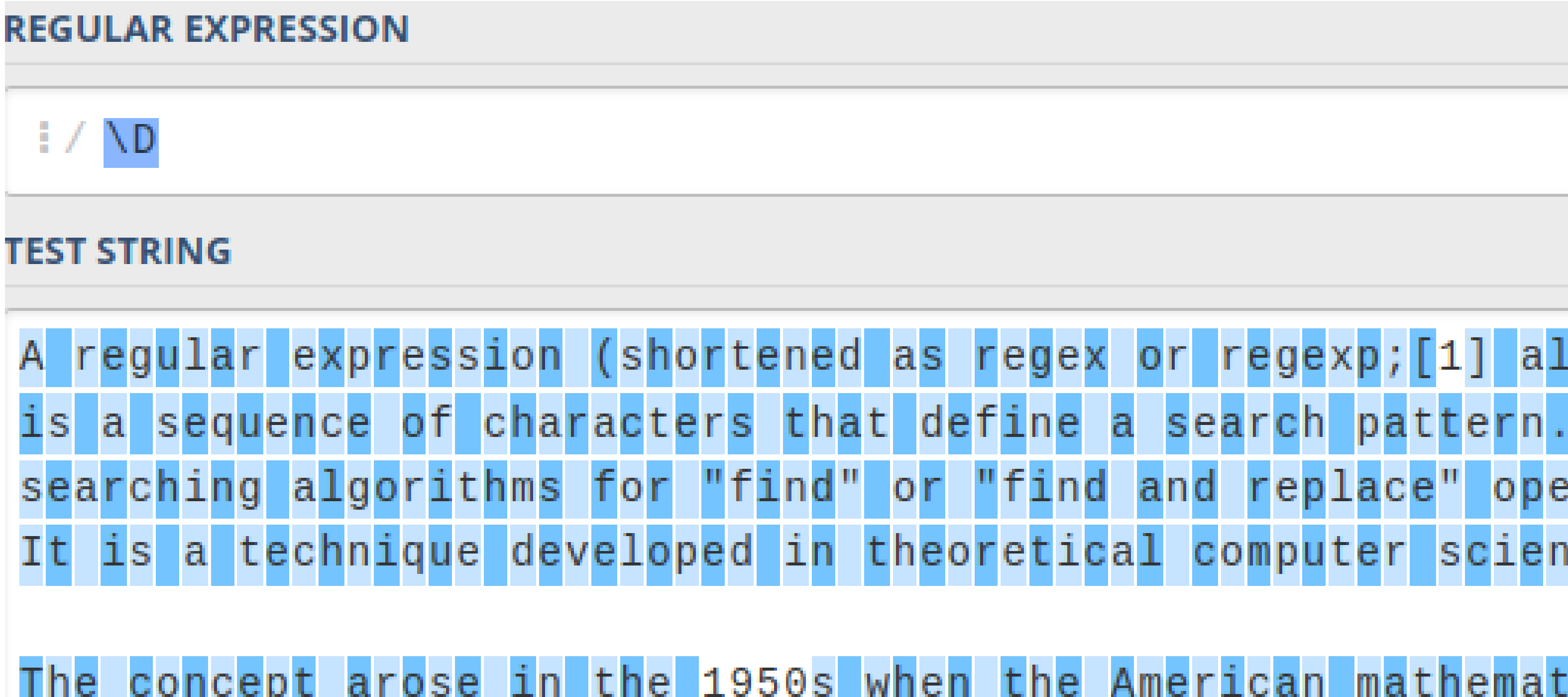
REGULAR EXPRESSION

```
/ \d
```

TEST STRING

```
A regular expression (shortened as regex or regexp;[1] als
is a sequence of characters that define a search pattern.
searching algorithms for "find" or "find and replace" oper
It is a technique developed in theoretical computer scienc

The concept arose in the 1950s when the American mathemati
```



REGULAR EXPRESSION

```
/ \D
```

TEST STRING

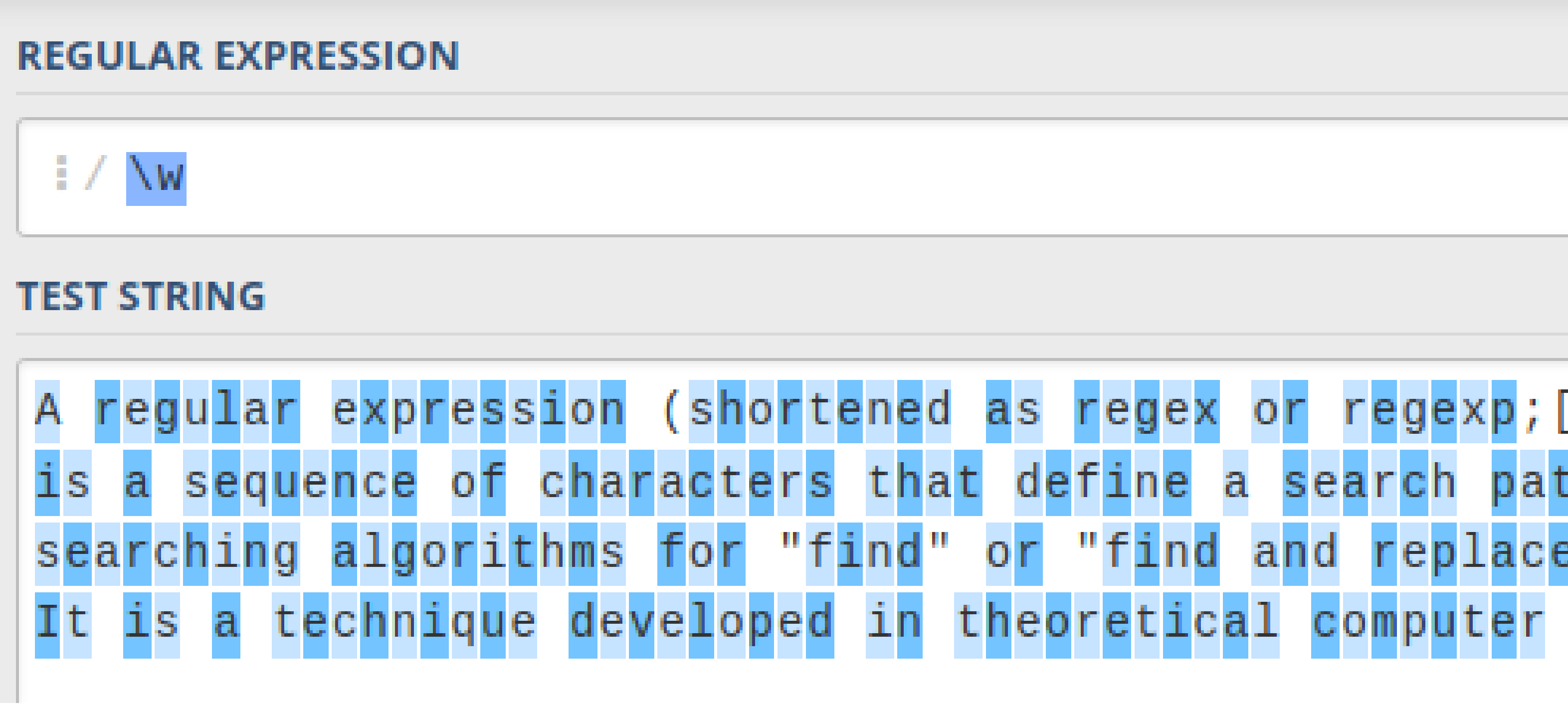
```
A regular expression (shortened as regex or regexp;[1] al
is a sequence of characters that define a search pattern.
searching algorithms for "find" or "find and replace" ope
It is a technique developed in theoretical computer scien

The concept arose in the 1950s when the American mathemat
```

Different types

\w word

\W non-word

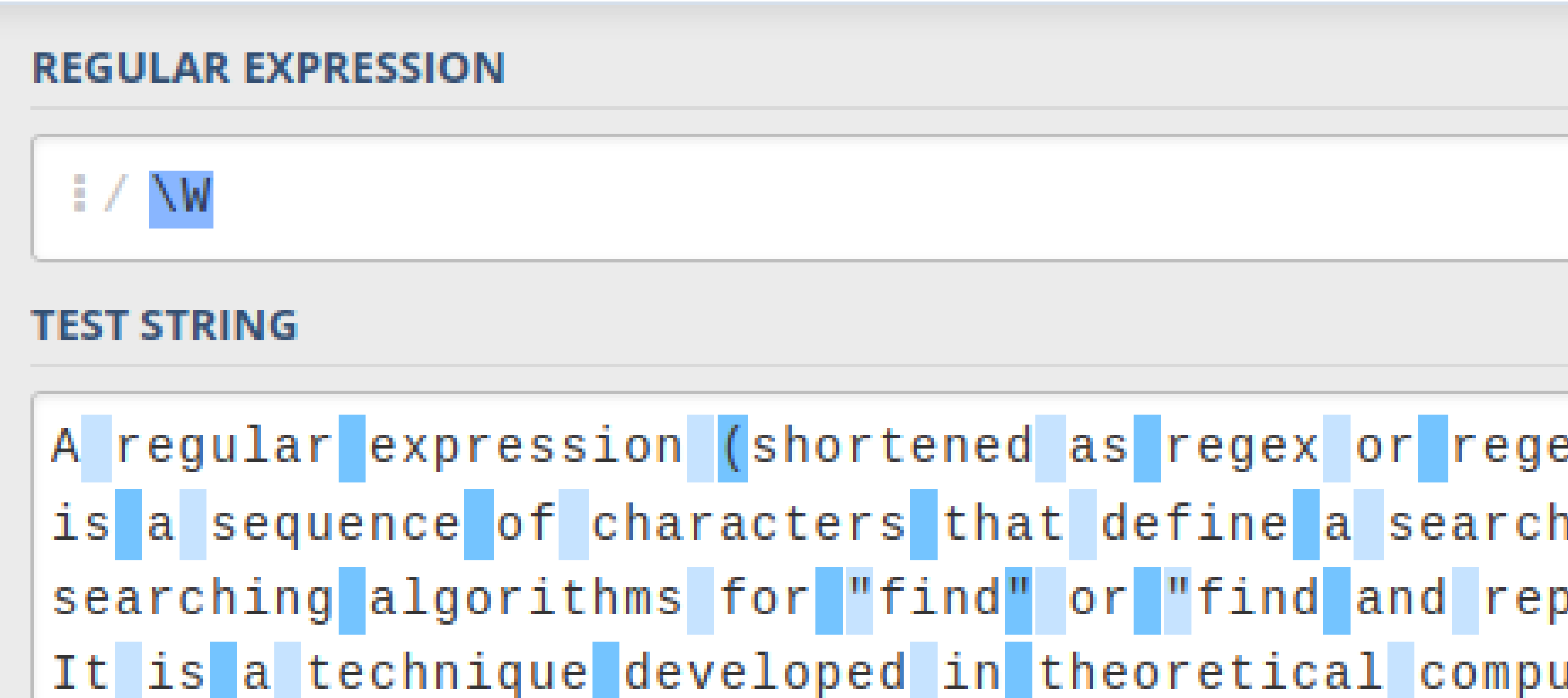


REGULAR EXPRESSION

```
/ \w
```

TEST STRING

```
A regular expression (shortened as regex or regexp;[
is a sequence of characters that define a search pat
searching algorithms for "find" or "find and replace
It is a technique developed in theoretical computer
```



REGULAR EXPRESSION

```
/ \W
```

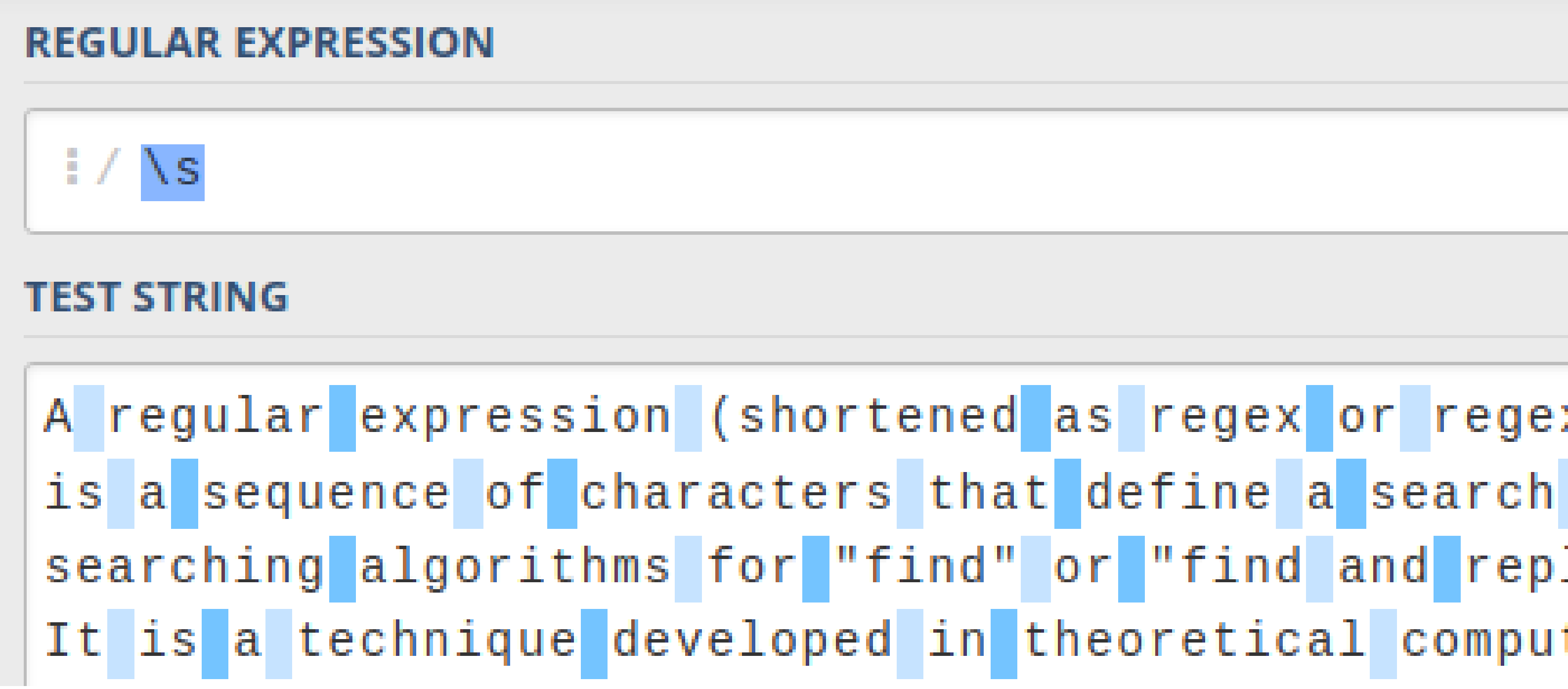
TEST STRING

```
A regular expression (shortened as regex or rege
is a sequence of characters that define a search
searching algorithms for "find" or "find and rep
It is a technique developed in theoretical compu
```

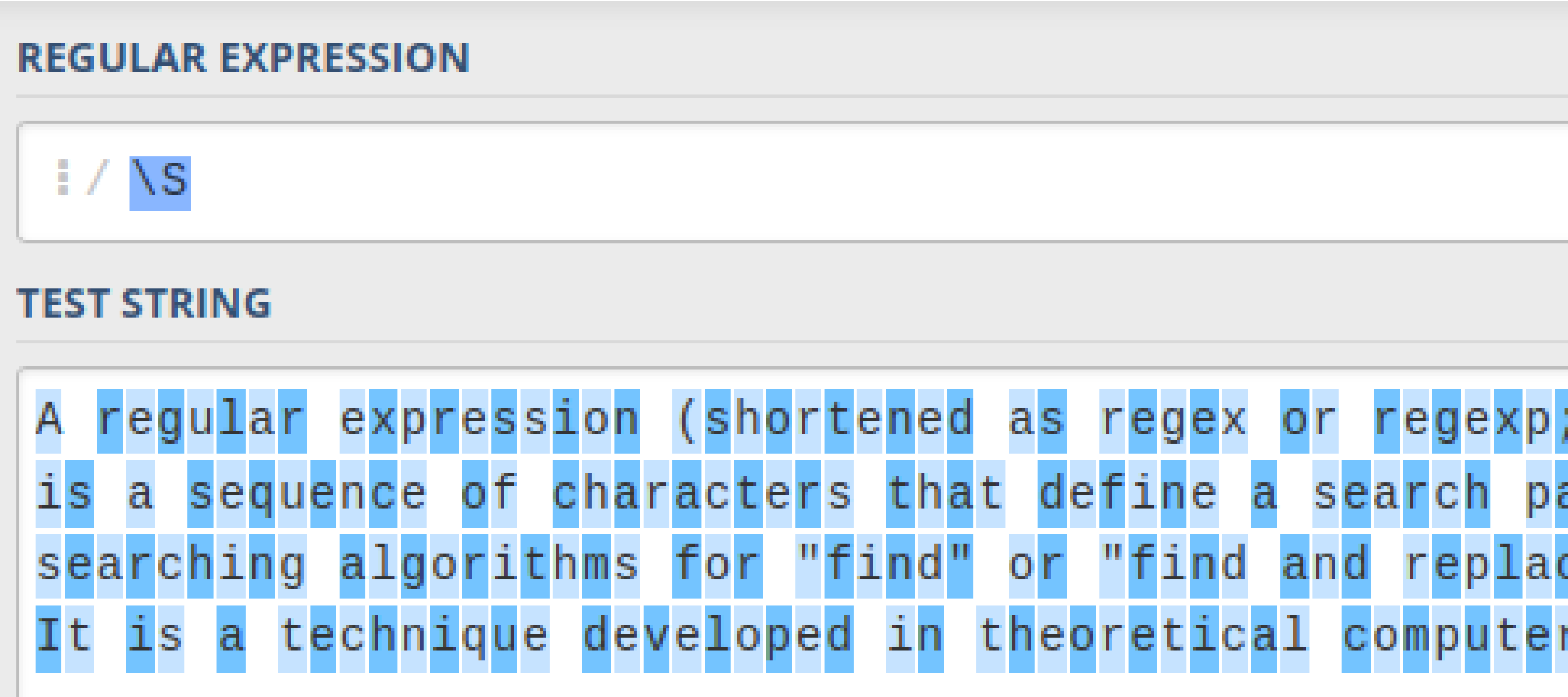
Different types

\s = whitespace

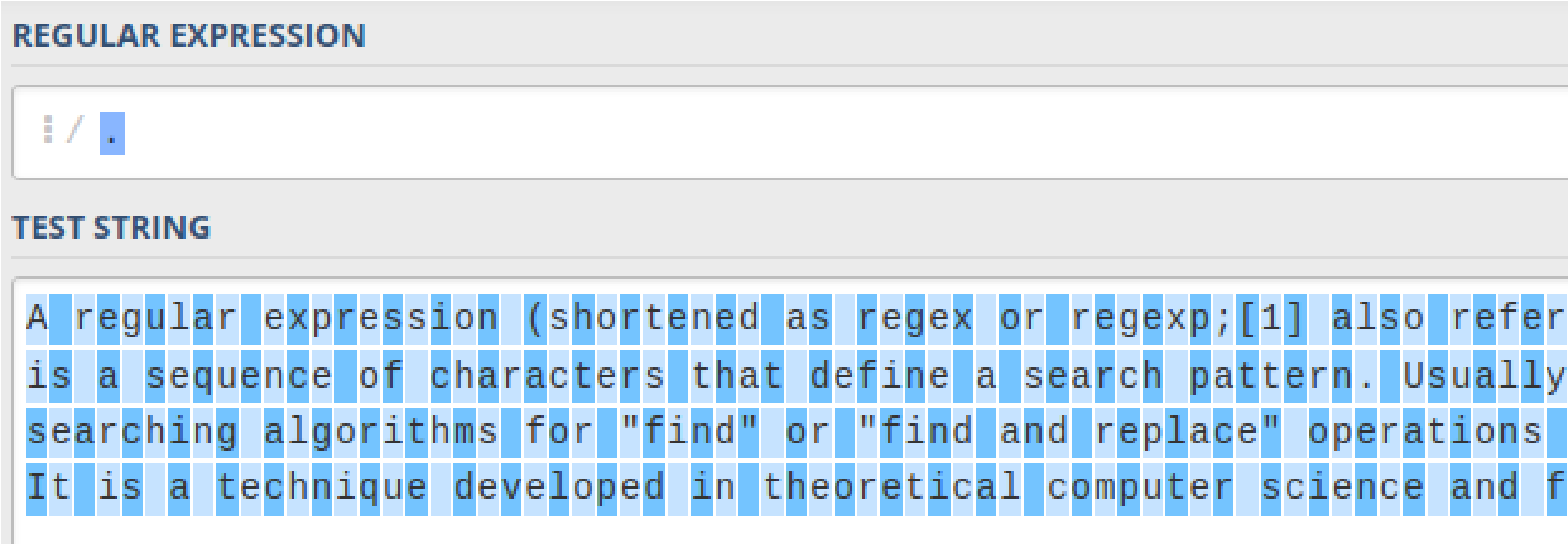
\S = non-whitespace



```
A regular expression (shortened as regex or rege
is a sequence of characters that define a search
searching algorithms for "find" or "find and rep
It is a technique developed in theoretical compu
```



```
A regular expression (shortened as regex or regexp
is a sequence of characters that define a search pa
searching algorithms for "find" or "find and replac
It is a technique developed in theoretical computer
```



REGULAR EXPRESSION

```
/ .
```

TEST STRING

```
A regular expression (shortened as regex or regexp;[1] also refer
is a sequence of characters that define a search pattern. Usually
searching algorithms for "find" or "find and replace" operations
It is a technique developed in theoretical computer science and f
```

Different types

. anything

Different types

\d (\D)

\w (\W)

\s (\S)

.

Custom type

[]

-

^

|

Boundaries

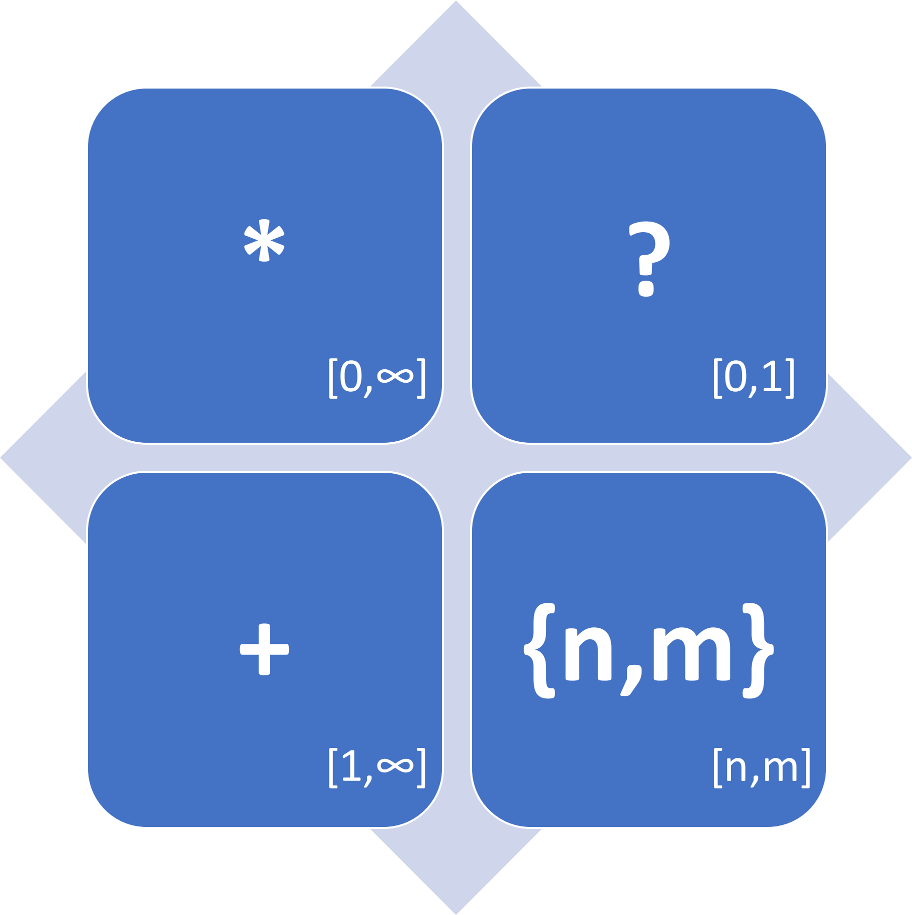
\b

\B

^

$

Different ways of matching on a type



*

[0,∞]

?

[0,1]

+

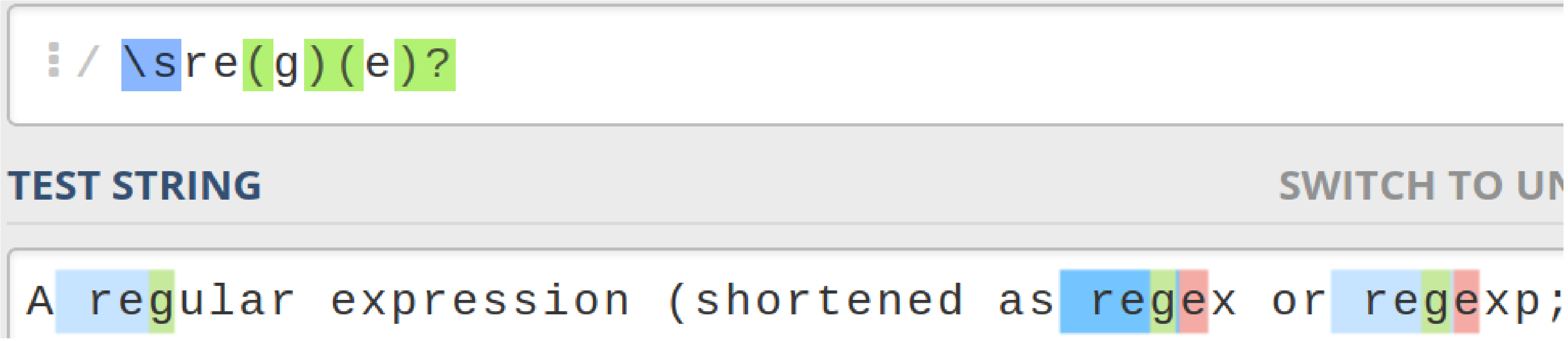
[1,∞]

{n,m}

[n,m]

Grouping and backreferences

- Grouping is done through the use of paranthesis
  - Apply multipliers
  - Backreference (match & replace, match & extract)



```
/ \sre(g)(e)?
```

TEST STRING

```
A regular expression (shortened as regex or regexp;
```

Pros & Cons

Pros

- Versatile
- Exists almost everywhere

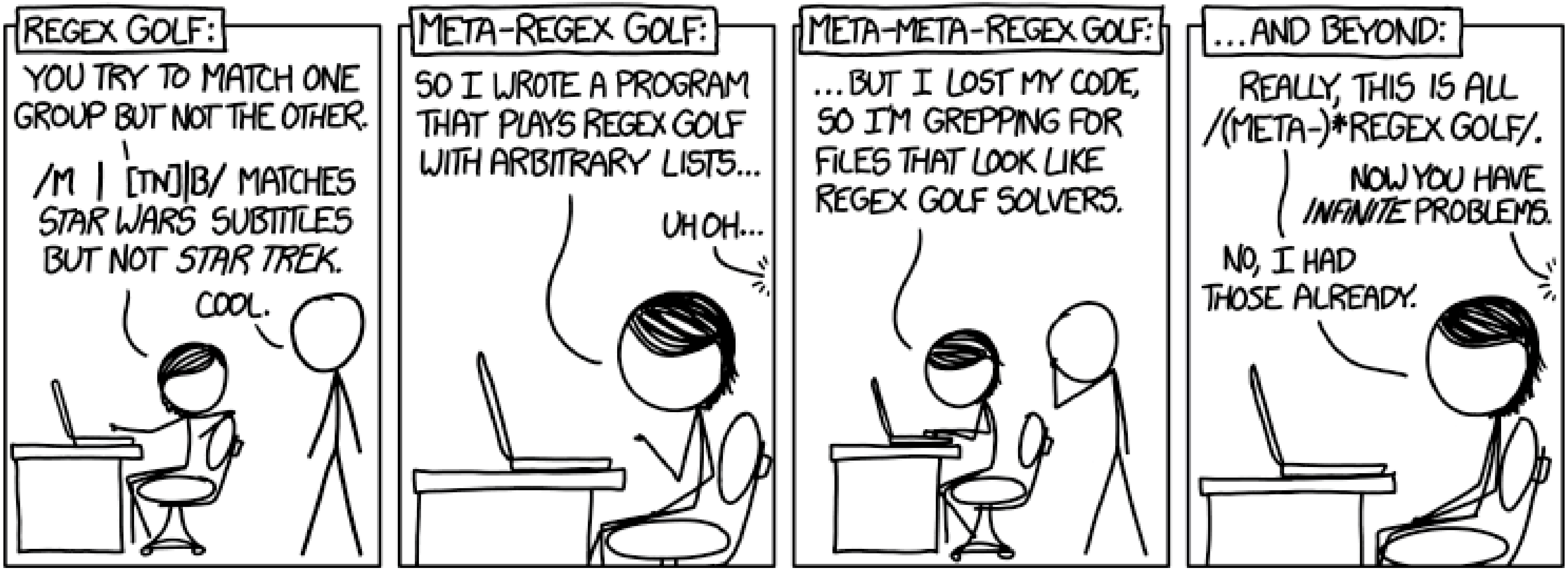
Cons

- A little bit slow
- Easy to "abuse"

Example: Password

Scenario:

- 6 to 12 characters in length
- Must have at least one uppercase letter
- Must have at least one lower case letter
- Must have at least one digit
- Should contain other characters


REGEX GOLF:
YOU TRY TO MATCH ONE GROUP BUT NOT THE OTHER.
/M | [TN]|B/ MATCHES *STAR WARS* SUBTITLES BUT NOT *STAR TREK*.
COOL.
META-REGEX GOLF:
SO I WROTE A PROGRAM THAT PLAYS REGEX GOLF WITH ARBITRARY LISTS...
UH OH...
META-META-REGEX GOLF:
...BUT I LOST MY CODE, SO I'M GREPPING FOR FILES THAT LOOK LIKE REGEX GOLF SOLVERS.
...AND BEYOND:
REALLY, THIS IS ALL /(META-)*REGEX GOLF/.
NOW YOU HAVE *INFINITE* PROBLEMS.
NO, I HAD THOSE ALREADY.

Resources

- Regex101: https://regex101.com
- Regexgolf: https://alf.nu/RegexGolf
- **Challenge:** https://regexone.com/
- Language Docs
- Cheat Sheet: http://web.mit.edu/hackl/www/lab/turkshop/slides/regex-cheatsheet.pdf

- Pro-tip; Fuzzy Search
- Google searching pro-tip
  - https://support.google.com/websearch/answer/2466433