Wildcards on Steroids

What Are Regular Expressions ?

And Why Are They Useful to You ?

Regular Expression:

A sequence of characters that define a search pattern.

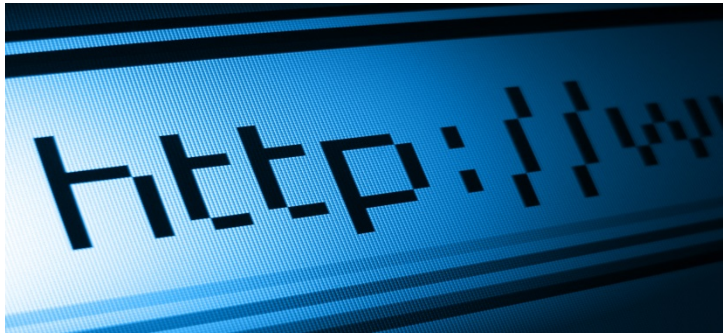
Pattern Recognition

Patterns

PIN
1
2
3
4
5
6
7
8
9
*
0
#

JPG
PNG
PSD
TIFF
BMP

SLAVE


http://w

RCA
AM
5:41

775-AB
SV 03-081
VIC
KPY·07
LJ 149-26
B43-76
VICTORIA-GARDEN STA
3423
90-32 LJ
SXK·87
SA · The Festival Stat
126 12

EURO
EYPO
EURO
BCE ECB EZB EKT EKP 2002

Patterns

TSA PRE INTL GT4PJ2 39H 132
UA 898 07E217

PNA89241
BEIJING TO WASHINGTON

UA898 GATE BOARD TIME SEAT
PEK-IAD E29 5:40P 39H BOARDING GROUP
THU MAY 1 2014 GATE MAY CHANGE DEPARTS: 6:25 PM AISLE 5
ARRIVES: 8:05 PM ECONOMY

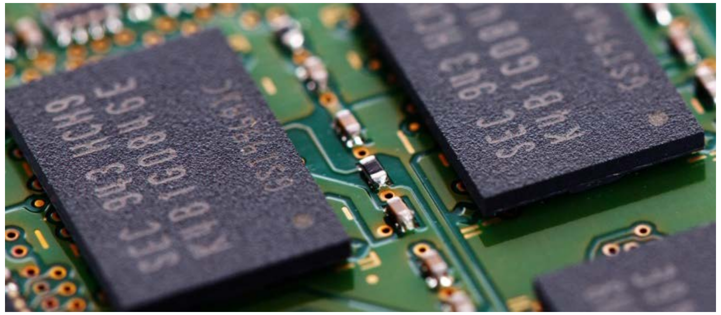
CONFIRMATION: GT4PJ2
TICKET: 0164793692676

STAPLE HERE
INSERT

De Slegte
610
De last van de wereld
De Slegte Leiden
** **
Totaal 1 artikel
Euro 17,50
EUR 17,50
0,00% Exclusief 17,50 BTW 0,00

Patterns

ISBN 1 86197 271 7



Typical Uses for Regular Expressions

- Input validation
- Search (and replace)
- String parsing
- Data scraping
- Syntax highlighting
- Data mapping

Users of Regular Expressions

Developers

Working with strings

Data Professionals

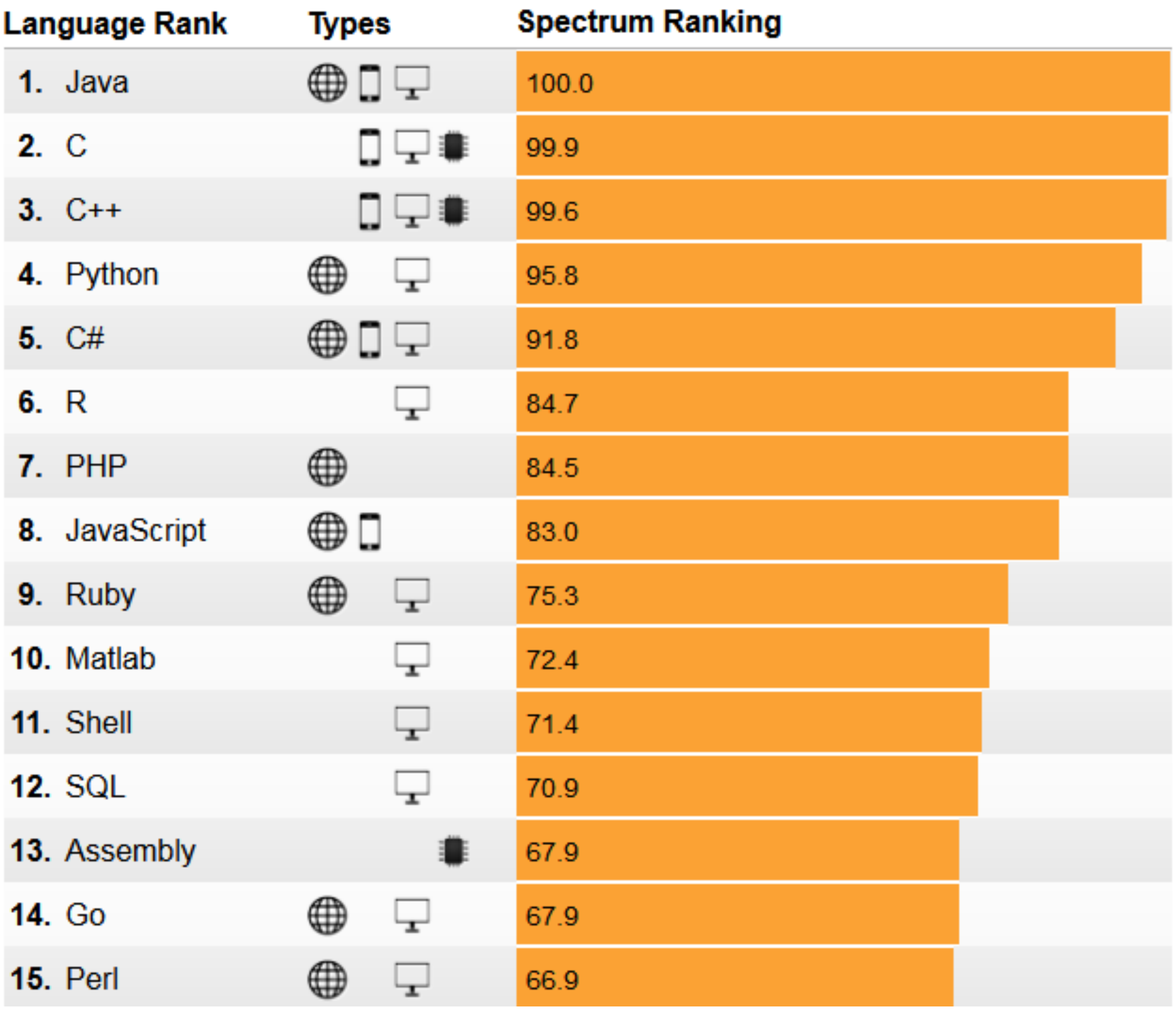
Query data

(Sys-)Admins

File system

Server directives

Top 15 Programming Languages



Language Rank	Types	Spectrum Ranking	
1. Java	🌐 📱 🖥	100.0	✓
2. C	📱 🖥 ▪	99.9	✓
3. C++	📱 🖥 ▪	99.6	✓
4. Python	🌐 🖥	95.8	✓
5. C#	🌐 📱 🖥	91.8	✓
6. R	🖥	84.7	✓
7. PHP	🌐	84.5	✓
8. JavaScript	🌐 📱	83.0	✓
9. Ruby	🌐 🖥	75.3	✓
10. Matlab	🖥	72.4	✓
11. Shell	🖥	71.4	✓
12. SQL	🖥	70.9	✓
13. Assembly	▪	67.9	✗
14. Go	🌐 🖥	67.9	✓
15. Perl	🌐 🖥	66.9	✓

Source: IEEE, July 2015

grep

g/re/p

```
g/<regular expression>/p
```

g/<regular expression>/p

[G]lobal search

[P]rint result

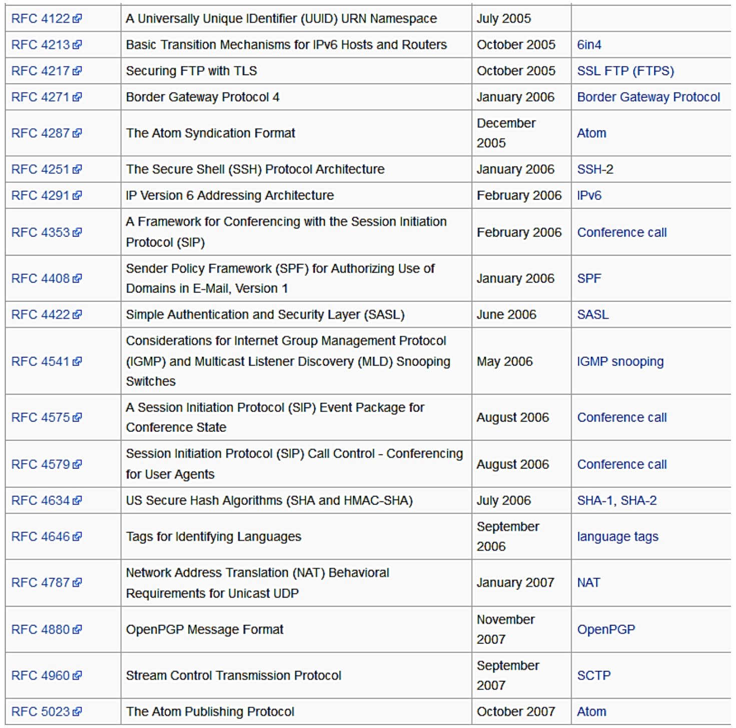
How It Works

How It Works

Pattern

Determining Pattern Rules

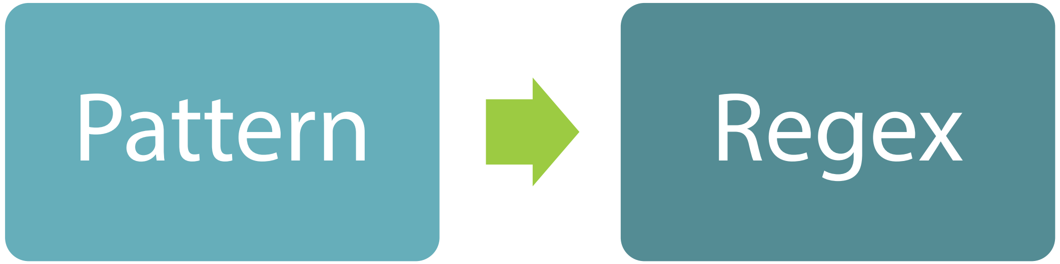
Photo by Squaio



RFC 4122	A Universally Unique IDentifier (UUID) URN Namespace	July 2005	
RFC 4213	Basic Transition Mechanisms for IPv6 Hosts and Routers	October 2005	6in4
RFC 4217	Securing FTP with TLS	October 2005	SSL FTP (FTPS)
RFC 4271	Border Gateway Protocol 4	January 2006	Border Gateway Protocol
RFC 4287	The Atom Syndication Format	December 2005	Atom
RFC 4251	The Secure Shell (SSH) Protocol Architecture	January 2006	SSH-2
RFC 4291	IP Version 6 Addressing Architecture	February 2006	IPv6
RFC 4353	A Framework for Conferencing with the Session Initiation Protocol (SIP)	February 2006	Conference call
RFC 4408	Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1	January 2006	SPF
RFC 4422	Simple Authentication and Security Layer (SASL)	June 2006	SASL
RFC 4541	Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches	May 2006	IGMP snooping
RFC 4575	A Session Initiation Protocol (SIP) Event Package for Conference State	August 2006	Conference call
RFC 4579	Session Initiation Protocol (SIP) Call Control - Conferencing for User Agents	August 2006	Conference call
RFC 4634	US Secure Hash Algorithms (SHA and HMAC-SHA)	July 2006	SHA-1, SHA-2
RFC 4646	Tags for Identifying Languages	September 2006	language tags
RFC 4787	Network Address Translation (NAT) Behavioral Requirements for Unicast UDP	January 2007	NAT
RFC 4880	OpenPGP Message Format	November 2007	OpenPGP
RFC 4960	Stream Control Transmission Protocol	September 2007	SCTP
RFC 5023	The Atom Publishing Protocol	October 2007	Atom

Photo by Succo

How It Works



Pattern

Regex

```
/^(?!(?:(?:\x22?\x5C[\x00-\x7E]\x22?)|(?:\x22?[^\x5C\x22]\x22?)){255,})(?!(?:
(?:\x22?\x5C[\x00-\x7E]\x22?)|(?:\x22?[^\x5C\x22]\x22?)){65,}@)(?:(?:[\x21
\x23-\x27\x2A\x2B\x2D\x2F-\x39\x3D\x3F\x5E-\x7E]+)|(?:\x22(?:[\x01-\x08\x0B
\x0C\x0E-\x1F\x21\x23-\x5B\x5D-\x7F]|(?:\x5C[\x00-\x7F]))*\x22))(?:\.(?:(?:
[\x21\x23-\x27\x2A\x2B\x2D\x2F-\x39\x3D\x3F\x5E-\x7E]+)|(?:\x22(?:[\x01-\x08
\x0B\x0C\x0E-\x1F\x21\x23-\x5B\x5D-\x7F]|(?:\x5C[\x00-\x7F]))*\x22)))*@(?:(?:
(?!.*[^.]{64,})(?:(?:(?:xn--)?[a-z0-9]+(?:-+[a-z0-9]+)*\.){1,126}){1,}(?:(?:
[a-z][a-z0-9]*)|(?:(?:xn--)[a-z0-9]+))(?:-+[a-z0-9]+)*)|(?:\[(?:(?:IPv6:(?:(?:
[a-f0-9]{1,4}(?::[a-f0-9]{1,4}){7})|(?:(?!(?:.*[a-f0-9][:\]]){7,})(?:[a-f0-9]
{1,4}(?::[a-f0-9]{1,4}){0,5})?::(?:[a-f0-9]{1,4}(?::[a-f0-9]{1,4}){0,5})?)))|
(?:(?:IPv6:(?:(?:[a-f0-9]{1,4}(?::[a-f0-9]{1,4}){5}:)|(?:(?!(?:.*[a-f0-9]:)
{5,})(?:[a-f0-9]{1,4}(?::[a-f0-9]{1,4}){0,3})?::(?:[a-f0-9]{1,4}(?::[a-f0-9]
{1,4}){0,3}:)?)))?(?:(?:25[0-5])|(?:2[0-4][0-9])|(?:1[0-9]{2})|(?:[1-9]?[0-9]
))(?:\.(?:(?:25[0-5])|(?:2[0-4][0-9])|(?:1[0-9]{2})|(?:[1-9]?[0-
9]))){3}))\]))\z/i
```

© Michael Rushton & The PHP Group

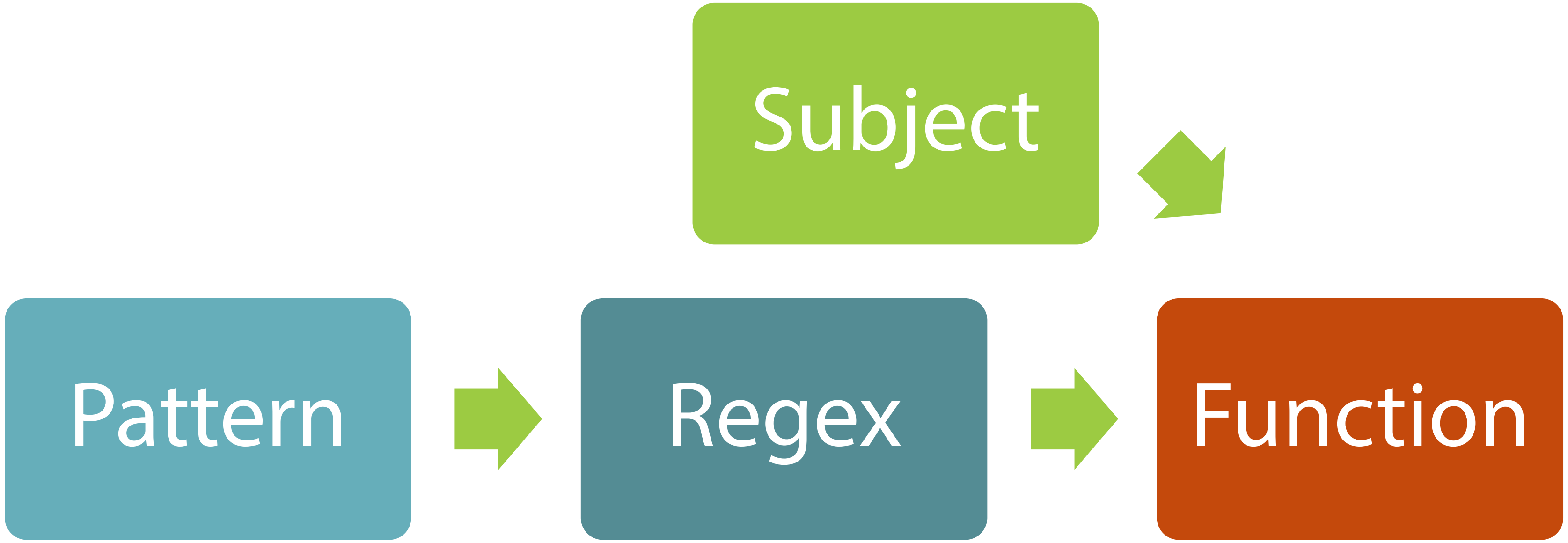
Email According to RFC 5321

```
/^[a-z0-9!#$%&'*+/=?^_`{|}~-]+(?:\.[a-z0-9!#$%&'*+/=?^_`{|}~-]+)*@(?:[a-z0-9](?:[a-z0-9-]*[a-z0-9])?\.)+[a-z0-9](?:[a-z0-9-]*[a-z0-9])?\z/i
```

Basic Email Validation

- does not allow for i8n domains
- is flexible towards new top-level domain additions

How It Works



Subject

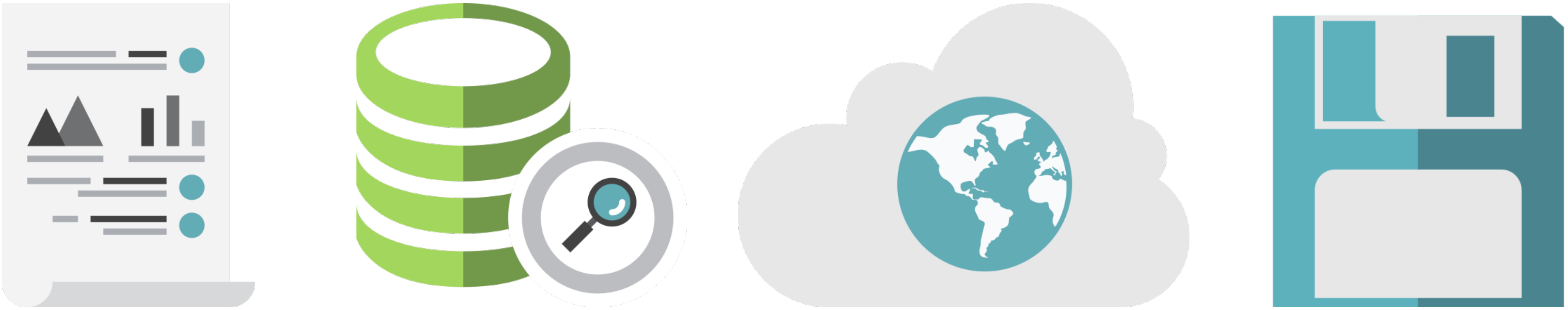
Pattern

Regex

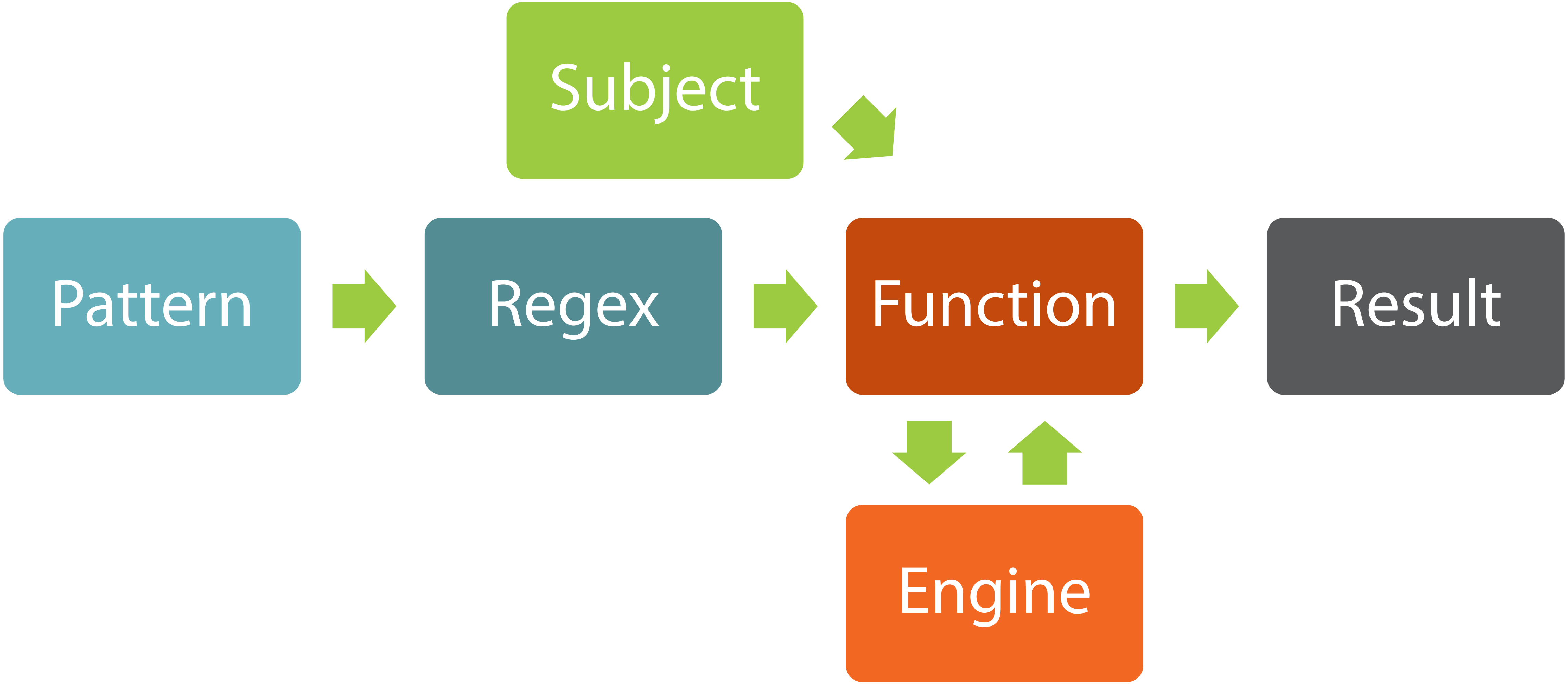
Function

Subject String

Subject



How It Works



Subject

Pattern

Regex

Function

Result

Engine

Result Types

Does it match ?	How many matches have been found ?	What are the matches ?
Boolean or similar	Integer	Array