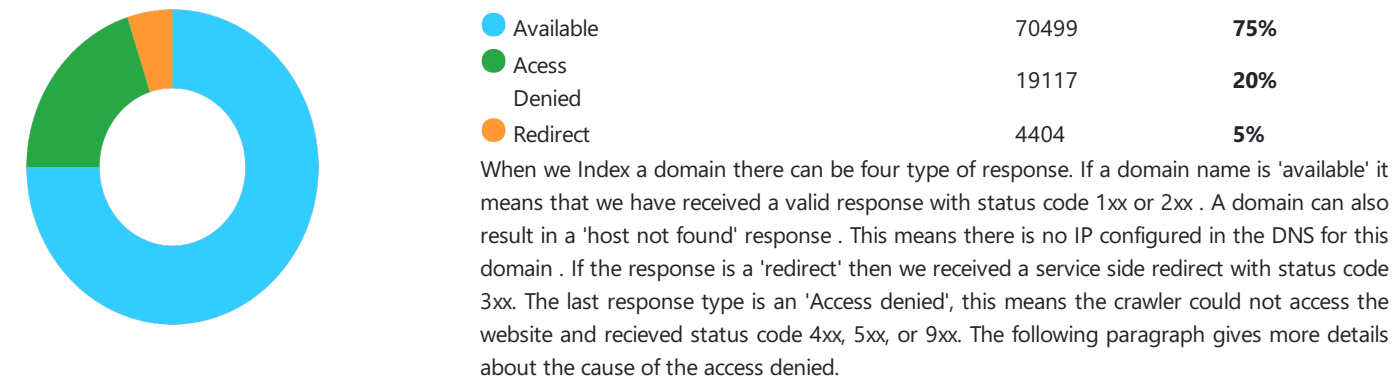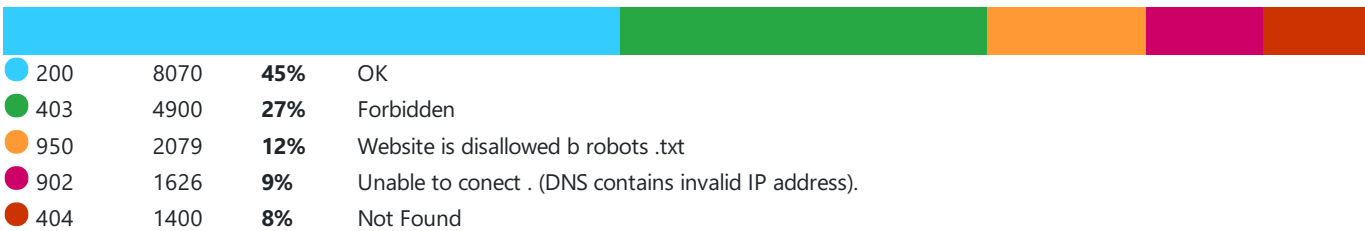# Zone file response

In November we've analyzed 94,020 domains with our crawlers, The results from that crawl are used to generate this report. Not every domain contains a website. This chapter gives insights into the responses of the domains.

Response

🔵 Available		70499	**75%**
🟢 Acess Denied		19117	**20%**
🟠 Redirect		4404	**5%**

When we Index a domain there can be four type of response. If a domain name is 'available' it means that we have received a valid response with status code 1xx or 2xx . A domain can also result in a 'host not found' response . This means there is no IP configured in the DNS for this domain . If the response is a 'redirect' then we received a service side redirect with status code 3xx. The last response type is an 'Access denied', this means the crawler could not access the website and recieved status code 4xx, 5xx, or 9xx. The following paragraph gives more details about the cause of the access denied.
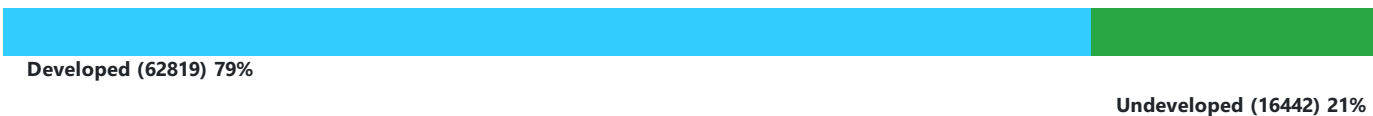
Access Denied

An 'Access denied' means that the crawler can't access the website. This can occur when the DNS is not configured, the server is unavailable or access is not allowed. in most cases there is no website (DNS is not configured) but sometimes there is, in that case the hosting provider, Webmaster or CMS of the website doesn't allow the crawler to visit the website. if a domain result in an Access denied the 'Status Code' explains why access was denied. In the following chart you can see the top 5 reasons why some domains resulted in an access denied.

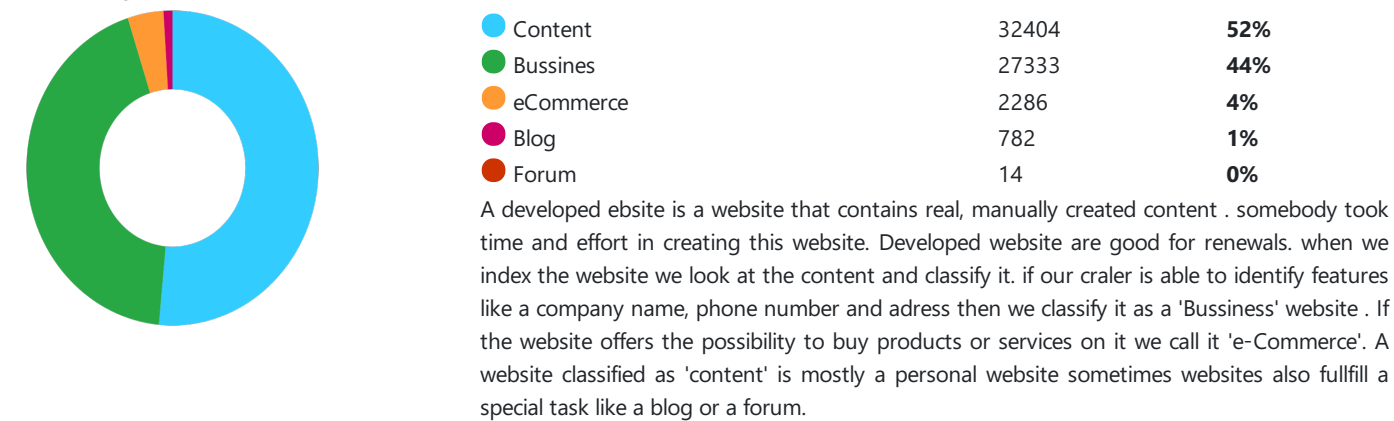|				
🔵 200	8070	**45%**	OK	
🟢 403	4900	**27%**	Forbidden	
🟠 950	2079	**12%**	Website is disallowed b robots .txt	
🟣 902	1626	**9%**	Unable to conect . (DNS contains invalid IP address).	
🔴 404	1400	**8%**	Not Found	

Available

When a domain result in an actual website we say the response is available. We have not looked at the actual content et. The content of a domain is important for reneals. A domain that is available can be 'developed' or 'undeveloped' . A developed website is a ebsite that contains real, manually created content. Somebody took time and effort to create this website . The opposite is an undeveloped website . This is q website that contain a placeholder (default page from the registar), is paked . Contains almost no content or shows the content from another website(frame).

Developed (62819) 79%

Undeveloped (16442) 21%

Developed

🔵 Content		32404	**52%**
🟢 Bussines		27333	**44%**
🟠 eCommerce		2286	**4%**
🟣 Blog		782	**1%**
🔴 Forum		14	**0%**

A developed ebsite is a website that contains real, manually created content . somebody took time and effort in creating this website. Developed website are good for renewals. when we index the website we look at the content and classify it. if our craler is able to identify features like a company name, phone number and adress then we classify it as a 'Bussiness' website . If the website offers the possibility to buy products or services on it we call it 'e-Commerce'. A website classified as 'content' is mostly a personal website sometimes websites also fullfill a special task like a blog or a forum.

Undeveloped

● Placeholder	8469	**52%**
● Low content	7475	**45%**
● Framed	468	**3%**
● Parking	30	**0%**

An undevelop website is a website that contains a placeholder (default page from the registrar), is parked , has almost no content or shows the content of another website (frame). An undeveloped website is not good for renewals . if people didn't take the effort to put a website on the domain, then the chance of a domain being renewad is lower.

Website redirect

A redirect is a technique which can be used to redirect a domain to or URL, A redirect can be implemented 'server side' or 'client side' . A server side redirect is done b using a 3xx status code followed by the new destination . A client side redirect is done by using a piece of javascript or a META refresh . The crawler cannot execute javascript so onl detects server side redirect . The following chart show what TLD's are redirected b domains:
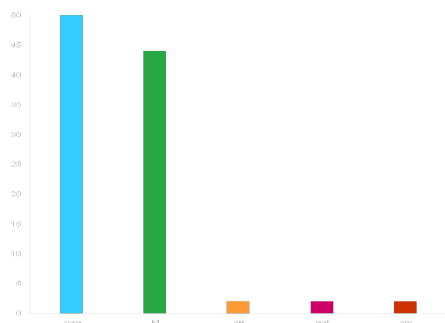
● .com	2058	**50%**
● .id	1811	**44%**
● .eu	91	**2%**
● .net	81	**2%**
● .org	80	**2%**

Top 15 regitrars

In the following chart we display the top 15 registrars . it's not only the number of domains that is important to know but also how these domains are being used . The chart shows for each registrar how many domains the have and how man are actually avilable (contain a website) and developed (higher is better).

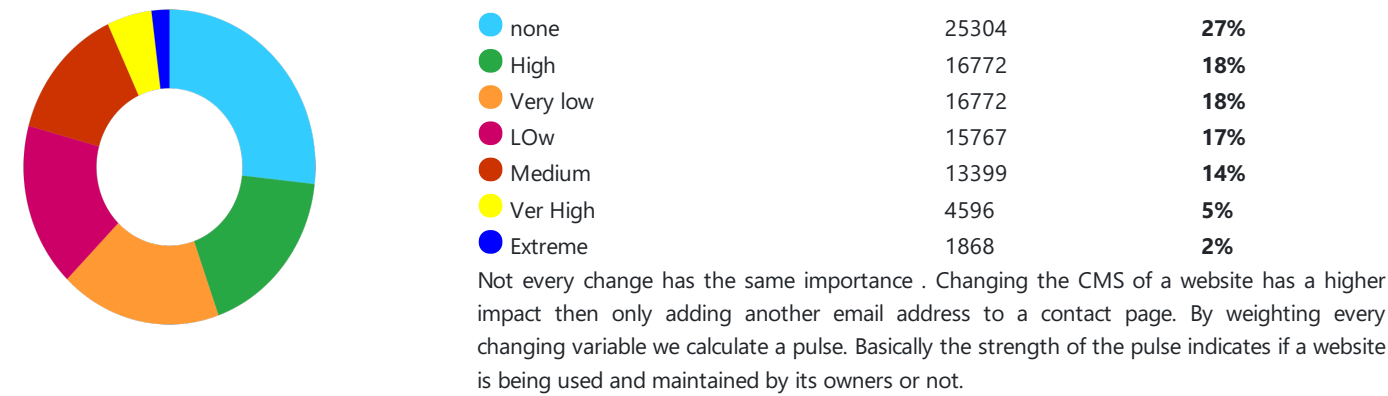| Hosting	● Domains	● Available	● Developed
Digital registra	41068	30833	24620
Reseller.co.id	12828	9019	7123
Cbn registrar	9946	7163	4923
Pt registrasi nama domain	6380	5365	4446
Kementrian komunikasi	3202	2448	2242
indoreg	3118	2335	1919
Domain cloud	2675	2090	1654
Pt qwords company international	2355	1893	1388
Satujuta domain	2037	1917	1853
Core mediatech	1804	1413	1167

Heartbeat

Every month we index all the website we track again and update all the varaiables . Besides updating we also archive all the result and keep track of all the changes. The monthly number of weighted changes result in a heartbeat.

Heartbeat

Active (757) 76%

Not active (243) 24%

Pulse

🔵 none		25304	**27%**
🟢 High		16772	**18%**
🟠 Very low		16772	**18%**
🟣 LOw		15767	**17%**
🔴 Medium		13399	**14%**
🟡 Ver High		4596	**5%**
🔵 Extreme		1868	**2%**

Not every change has the same importance . Changing the CMS of a website has a higher impact then only adding another email address to a contact page. By weighting every changing variable we calculate a pulse. Basically the strength of the pulse indicates if a website is being used and maintained by its owners or not.

| TOP 10 reistrars		Domains	🔵 Active	🟢 Not active
Digital registra		41068	🔵 75 %	🟢 25%
Reseller.co.id		12828	🔵 72 %	🟢 28%
Cbn registrar		9946	🔵 70 %	🟢 30%
Pt registrasi nama domain		6380	🔵 77 %	🟢 23%
Kementrian komunikasi		3202	🔵 84 %	🟢 16%
indoreg		3118	🔵 79 %	🟢 21%
Domain cloud		2675	🔵 73 %	🟢 27%
Pt qwords company international		2355	🔵 68 %	🟢 32%
Satujuta domain		2037	🔵 94 %	🟢 6%
Core mediatech		1804	🔵 65 %	🟢 35%

Renewal probability

At Dataprovider.com we see milions of domains come online and go offline. we took features of these domains like their content, age, popularit and activity and created a model to predict if a domain will get renewed or not. We call this score the 'Reneal probability'.

Renewal probability

- 🔵 LOw   0   **0%**
- 🟢 Medium   16046   **17%**
- 🟠 High   77974   **83%**

If a domain has a low reneal probability, then it is very likely the domain will drop within a year . Domain with a high renewal probability will most likely renew.
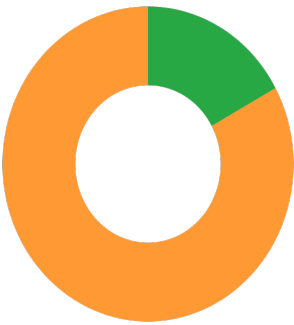
Top 15 registrars with high renewal risk

In the following chart we display the top 15 registrar and their renewal probability . It's not only the number of domains that is important to know, but also how many of these domains get renewad . The chart shows for each registrar how many domains they have and how many domains are probably not going to get renewed (higher is better).

Top 15 registrars	🔵	Domains	🟢	Low renewal est.
-Digital registra	🔵	41068	🟢	
-Reseller.co.id	🔵	12828	🟢	
-Cbn registrar	🔵	9946	🟢	
-Pt registrasi nama domain	🔵	6380	🟢	
-Kementrian komunikasi	🔵	3202	🟢	
-indoreg	🔵	3118	🟢	
-Domain cloud	🔵	2675	🟢	
-Pt qwords company international	🔵	2355	🟢	
-Satujuta domain	🔵	2037	🟢	
-Core mediatech	🔵	1804	🟢	

Hosting

If a domain results in an actual website, parking page or placeholder then this website is hosted at a hosting company. This chapter gives insight in what companies these websites are hosted at, and in what countries.
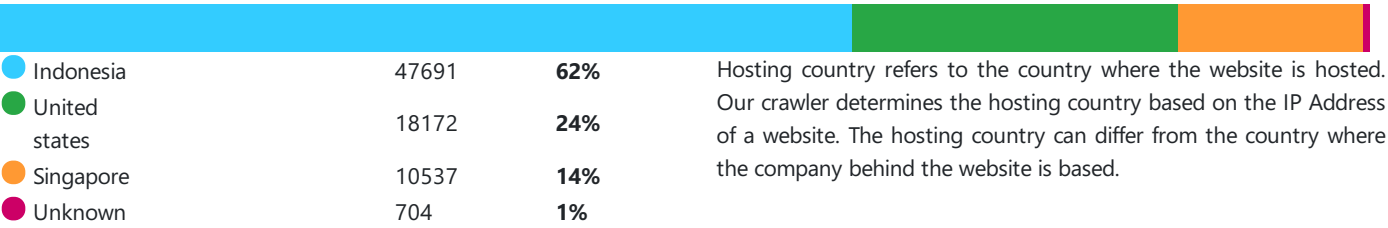
Hosting company

The hosting company is also known as the AS company. This company is responsible for the network on which the website is hosted and owns the IP address the website is on. The hosting company corresponds with the assigned AS number . The following chart shows were most website are hosted and how many are developed.

Company	🔵	Domain	🟢	Developed	🟠	Undeveloped
pt master web network	🔵	6976	🟢	5123	🟠	1853
cloudflare, inc	🔵	6010	🟢	4034	🟠	1976

rumah web indonesia	🔵	5775	🟢	3016	🟠	2759
hostinger international limit	🔵	4331	🟢	2617	🟠	1714
pt beon intermedia	🔵	4076	🟢	3333	🟠	743
Google.inc	🔵	3359	🟢	3083	🟠	276
Dtpnet nap	🔵	3175	🟢	2497	🟠	678
pt telekomunikasi indonesia	🔵	2893	🟢	2394	🟠	499
pt suraloka digital kreatif	🔵	1932	🟢	1455	🟠	477
Choopa, llc	🔵	1917	🟢	1656	🟠	261
softlaer technologies inc.	🔵	1893	🟢	1406	🟠	487
indosatm2 asn	🔵	1802	🟢	1018	🟠	784
Biznet network	🔵	1685	🟢	1098	🟠	587
pt qwords company internet	🔵	1565	🟢	1045	🟠	520
Global media teknologi pt	🔵	1557	🟢	577	🟠	980

Hosting country

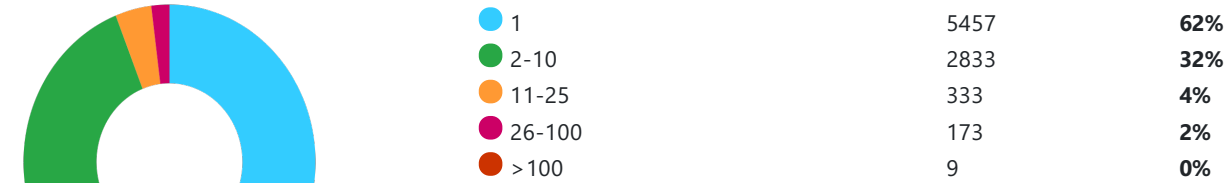|			
🔵 Indonesia	47691	**62%**	Hosting country refers to the country where the website is hosted. Our crawler determines the hosting country based on the IP Address of a website. The hosting country can differ from the country where the company behind the website is based.
🟢 United states	18172	**24%**	
🟠 Singapore	10537	**14%**	
🔴 Unknown	704	**1%**	

Subdomains

The subdomain refers to the first part of a domain. The most used subdomain in domain names is 'www' but there are also other subdomains, such as 'shop' or 'wiki'. The following chart shows how man domains use subdomains (other than www).
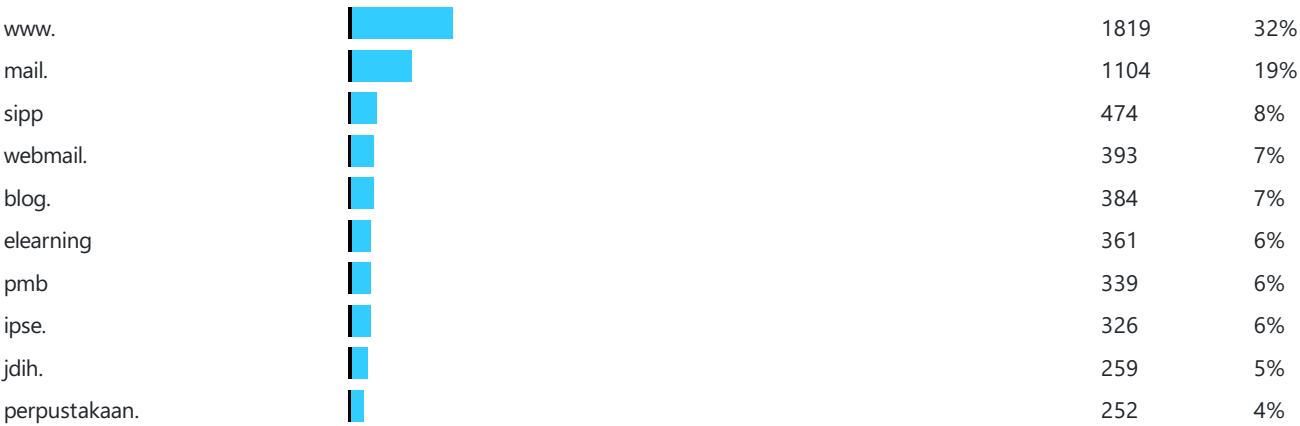
Yes (62,819) 79,3%

No (16.442) 20,7%

Distribution

🔵 1	5457	**62%**	
🟢 2-10	2833	**32%**	
🟠 11-25	333	**4%**	
🔴 26-100	173	**2%**	
🔴 >100	9	**0%**	
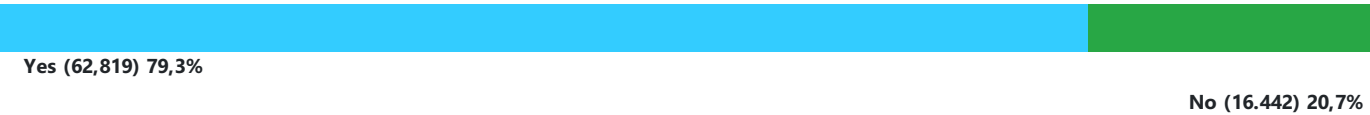
The number of subdomains used by a website tells something about the size of the website. Big companies tend to have multiple subdomains. The following charts show how many subdomains website have if they use subdomains and what names the used most often for a subdomain.

Top 10 subdomains

In the following chart we display the top 10 subdomains and how many domains use that subdomain. The subdomain refers to the first part of a domain.

www.		1819	32%
mail.		1104	19%
sipp		474	8%
webmail.		393	7%
blog.		384	7%
elearning		361	6%
pmb		339	6%
ipse.		326	6%
jdih.		259	5%
perpustakaan.		252	4%

DNS

The domain Name System (DNS) is a naming system for doamin names. It translate domain names to numerical IP address.

DNS type

●	DNS TXT	48337	**22%**
●	DNS MX	81529	**37%**
●	DNS NS	89174	**40%**
●	DNS SEC	2840	**1%**
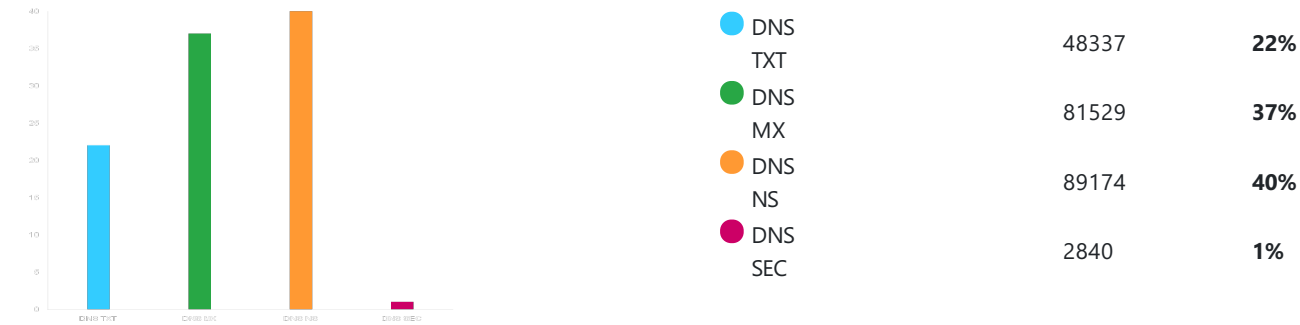
DNS TXT

Dataprovider.com does not only index the website, we also index the DNS records. For each domain we collect the A(IPv4), AAAA coverage of these records in the zone file.

●	SPF	1	**0%**
●	Verification	4065	**100%**
●	DKIM	5	**0%**

Top 5 MX domains

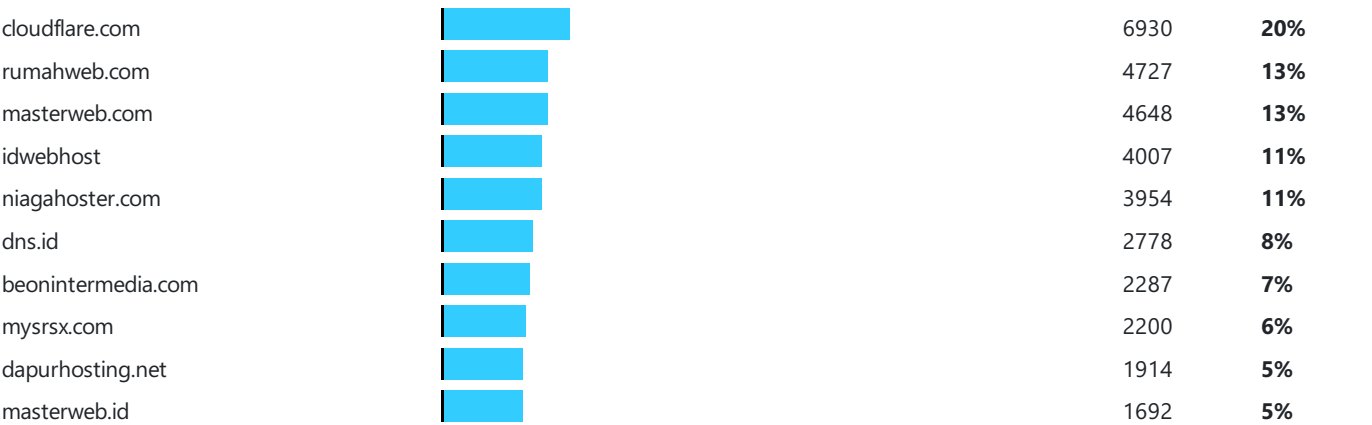|			
google.com		8094	**62%**
zoho.com		1911	**15%**
mail.my.id		1425	**11%**
vhosta.com		1126	**9%**
outlook.com		574	**4%**

DNS AAAA records IPV6

Every device on the internet is assigned a unique IP. with the rapid growth of the internet in the 1990s, it became evident that far more address would be needed to connect devices than the IPV4 address spaced had available. IPV6 was to deal with the longanticipated problem of IPv4 address exhaustion . IPv6 is intented to replace IPv4. Not many devices make use of IPv6 yet. The following chart show the availibility of IPv6.

Yes (5,292) 5.6%

No (88,728) 94.4%

DNS NS(Nameserver)

The NS record (nameserver) is a computer that is permanently connected to the internet and translates domain names into IP address. it contains the DNS of domain with all IP address that belong to that domain. Mostly the NS record contains a hostname of the register were the domain is registered . The following chart shows the most used nameserver (registars).

cloudflare.com		6930	**20%**
rumahweb.com		4727	**13%**
masterweb.com		4648	**13%**
idwebhost		4007	**11%**
niagahoster.com		3954	**11%**
dns.id		2778	**8%**
beonintermedia.com		2287	**7%**
mysrsx.com		2200	**6%**
dapurhosting.net		1914	**5%**
masterweb.id		1692	**5%**

WHOIS

The WHOIS (~who is) database contains the registration information for internet resources such as domain names. WHOIS informataion has limited availability via registrars and registries. Dataprovider.com has the WHOIS records for many domains, but not all of them.

WHOIS coverage

Yes (91,582) 97,4%

No (2,438) 2.6%

- Yes (980) **1%**
- No (93,040) **99%**
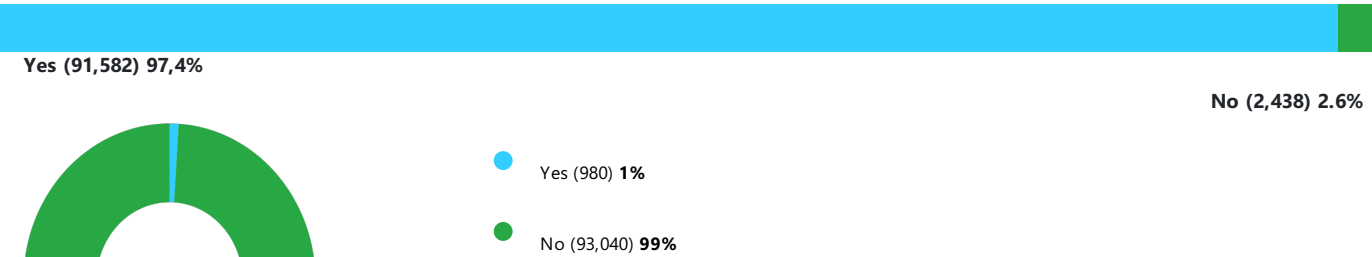
Most domain names registered through a registrar require the collection of essential contact information, such as the registrant's name, email adress, mailing address, phone number etc. This information is used in case dispute arises about the domain name. This information is displayed in the global whois system which is public and can be viewed by anyone. if you would like to keep your contact information private for a domain then ou can use a WHOIS privacy service like WhoisGuard, Domains by Proxy or Whois agent. Not every registrant fills out all his information . The following chart shows the completeness of the available WHOIS information.

Completeness

Name		91429	**91%**
Organization		89264	**89%**
Adress		91241	**91%**
Zip		89826	**90%**

|			
City		91226	**91%**
Country		91415	**91%**
Phone		90445	**90%**
Email		91582	**92%**

SSL

SSL (Secure Sockets Layers) is a security technology that encrypts communication between a browser and a server. You can recognize if a website uses SSL by checking if there is a small green lock in the address bar.

SSl certificates are utilized by millions of online bussinesses and individuals to decrease the risk of sensitive information being stolen or tempered with by hackers or identify thieves.

SSL enabled

Yes (52950) 56,3%

No (27922) 43,7%
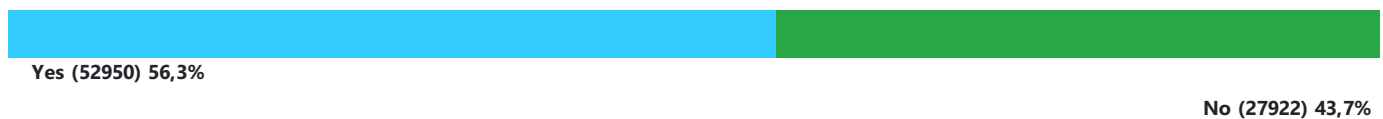
When we index a website we always start by resolving the hostname. we check if there is a valid response and obtain the IP address. We use this IP address to setup an SSL connection betwen the crawler and the server. After authentication the spider retrieves all the SSL certificate information such as the SSL type, issuer organization, and expiration dates.
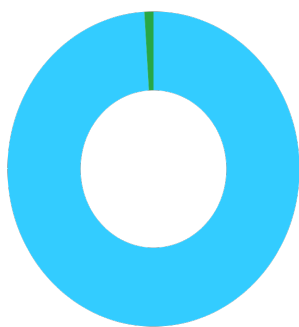
●	Domain Validation	45017	**99%**
●	Organization Validation	419	**1%**
●	Extended Validation	96	**0%**

There are 3 type of certificate available: Domain validation, Organization Validation, and Extended validation . Domain validation is default.

Privacy sensitive websites using SSL

privacy sensitive websites are websites that store personal information such as contact or payment details. Websites that offer the possibility to create an account, log into a system, fill in a contact form or process a payment are examples of privacy sensitive websites. these websites should have an SSL certificate installed and transmit the information via https. The following chart shows how many privacy sensitive websites have SSL installed.

Yes (62,819) 79,3%

No (16.442) 20,7%
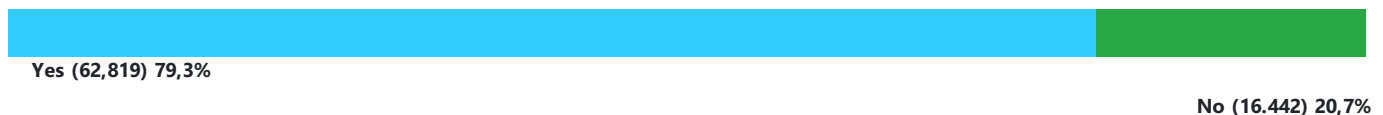
SSL issuer organization

The certificate issuer organization is responsible for delivering the certificate to a website . An SSL certificate is a digital certificate that authenticates the identity of a website and encrypts information sent. These certificates can only be issued by a certificate Authority . The following chart shows the top issuer organization.

cPanel		31676	**70%**
Let's Encrpt		7697	**17%**
COMODO CA Limited		5107	**11%**
GeoTrust inc.		347	**1%**
GlobalSign nv-sa		278	**1%**
DigiCert inc.		124	**0%**
GoDadd.com		56	**0%**

Symantec Corporation		56	**0%**
Amazon		34	**0%**
CloudFlare		32	**0%**

Content

if there is a website available on the domain the Dataprovider.com indexes 10-20 pages of each website . Using this content we can identity language country, size of the website and the use of social media.
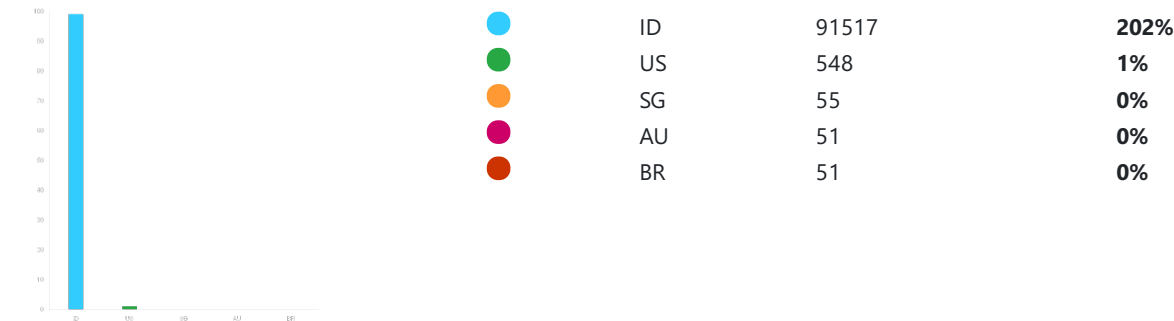
Language

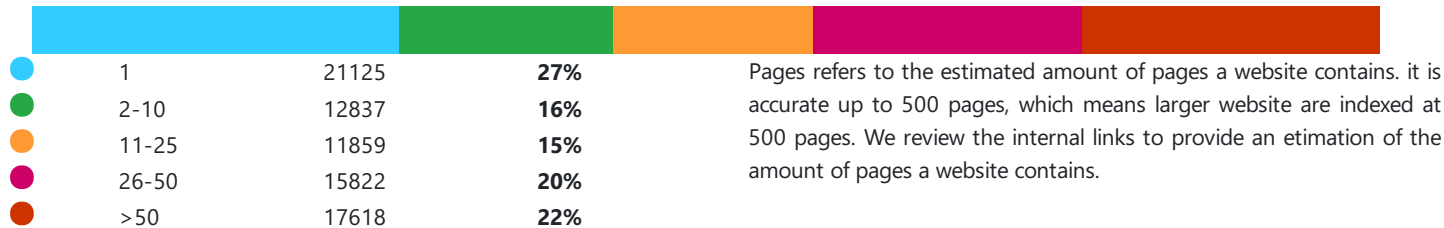|			
🔵	Indonesian	36923	**81%**
🟢	English	19913	**44%**
🟠	Italian	1541	**3%**
🔴	German	105	**0%**
🔴	Portuguese	54	**0%**
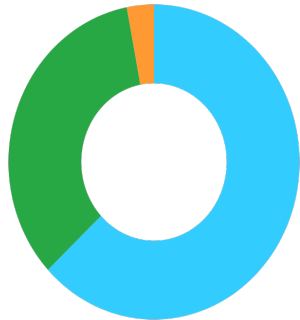
We currently recognize languages from 50 different countries. Our crawler determines the language of a website by using an n-gram model. An n-gram is a contiguous sequence of n items from a given sequence of text or speech. An n-gram model models sequences, notably natural languages, using the statistical properties of n-gram. The chart above shows the top there detected languages.

Countries

🔵	ID	91517	**202%**
🟢	US	548	**1%**
🟠	SG	55	**0%**
🔴	AU	51	**0%**
🔴	BR	51	**0%**

Country refers to the land where the website is located or where the website originated from. We determine the country of a website based on multiple variables such as hosting country. language on the website , top-level domain and contact details.

Pages per website

🔵	1	21125	**27%**
🟢	2-10	12837	**16%**
🟠	11-25	11859	**15%**
🔴	26-50	15822	**20%**
🔴	>50	17618	**22%**

Pages refers to the estimated amount of pages a website contains. it is accurate up to 500 pages, which means larger website are indexed at 500 pages. We review the internal links to provide an etimation of the amount of pages a website contains.
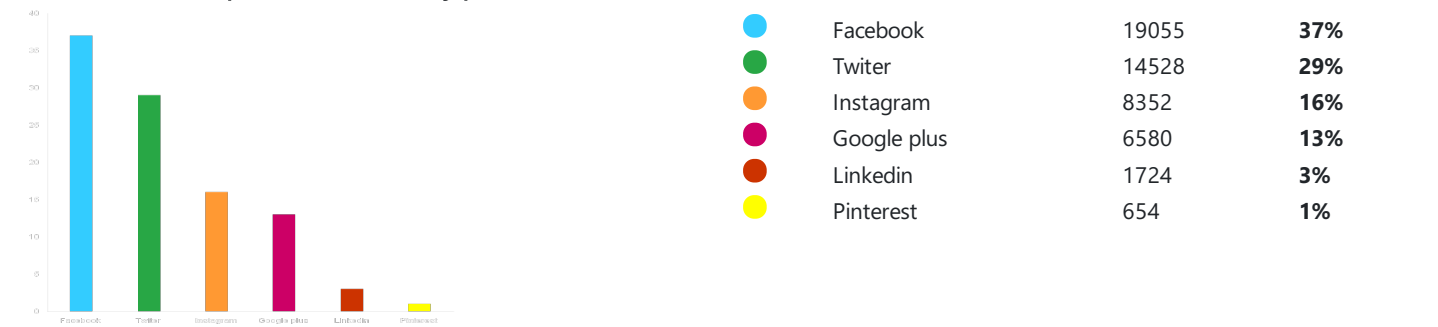
Social Media

Many people and companies use social media these days. Social media allows individuals to interact with one another, exchanging details about their lives such as biographical data, professional information, personal photos and up-to-the-minute thoghts. Website use social media to interact with their (potential) customers.

Yes (22118)

No (71902)

Social media per website type

🔵	Facebook	19055	**37%**
🟢	Twiter	14528	**29%**
🟠	Instagram	8352	**16%**
🔴	Google plus	6580	**13%**
🔴	Linkedin	1724	**3%**
🟡	Pinterest	654	**1%**

The usage of social media differs per website type. Bussines and eCommerce websites tend to have a higher penetration of social media because they use social media to engage with their audience. The chart above shows the penetration of different platforms for each website type.

Social media platformss

🔵	Content	9391	**43%**
🟢	Business	11636	**53%**
🟠	eCommerce	524	**2%**
🔴	Blog	258	**1%**
🔴	Forum	1	**0%**
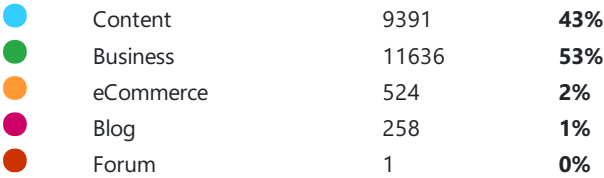
This chart shows the available social media platforms that are found on the homepages of available website. There are many social media platforms available like Facebook, Twiter, and Linkedin. we determine the social media platform based upon the domain name in the social media profile.