

## Assignment 2 - Estimating the size of the web

Zubair Abid (20171076)

1. **What is the number of people with access to the internet today? Do all of them have equal and unrestricted access to the internet?**

Around *4.57 billion people* across the world <sup>1</sup> have access to the internet today. Other statistics point it at being around *50% of the world's population*<sup>2 3</sup>, which would bring the number a little below 4 billion people. This includes people accessing the internet from any device, both mobile phones and desktop devices.

Not everyone has equal access - broadband subscription percentage is only *40%* in countries like South Korea and Germany <sup>4</sup>. Some countries also restrict access to the internet, such as China's great firewall <sup>5 6</sup>.

2. **How many pages/websites exist as part of the internet? What percent of them belong to the surface web and the Deep/Dark web?**

*Indexed pages:* A 2017 report <sup>7</sup> mentions that Google's index contains over *130 trillion* pages. The listed source, Google's howsearchworks page, currently places a vague figure of "hundreds of billions"

*Deep web:* Often confused with the dark web, it's much larger than that. Commonly cited research from 2001 estimates that the Deep Web as a whole is *400 to 500 times the size of the "surface web"* <sup>8</sup>. Assuming the estimates to carry over, that would put it at around *650,000 trillion* pages.

*Dark web:* It's a bit hard to get exact figures on the size of the Dark Web, but a reasonable proxy is Tor traffic percentages. One of the founders of the Tor project has said <sup>9</sup> that only around *3% of Tor traffic* was used

---

<sup>1</sup><https://www.statista.com/statistics/617136/digital-population-worldwide/>

<sup>2</sup><https://ourworldindata.org/internet#internet-access>

<sup>3</sup><https://data.worldbank.org/indicator/it.net.user.zs>

<sup>4</sup><https://ourworldindata.org/internet#broadband-access>

<sup>5</sup>[https://www.washingtonpost.com/world/asia\\_pacific/chinas-scary-lesson-to-the-world-censoring-the-internet-works/2016/05/23/413afe78-fff3-11e5-8bb1-f124a43f84dc\\_story.html](https://www.washingtonpost.com/world/asia_pacific/chinas-scary-lesson-to-the-world-censoring-the-internet-works/2016/05/23/413afe78-fff3-11e5-8bb1-f124a43f84dc_story.html)

<sup>6</sup><https://www.dezshira.com/library/legal/cyber-security-law-china-8013.html>

<sup>7</sup><https://searchengineland.com/googles-search-indexes-hits-130-trillion-pages-documents-263378>

<sup>8</sup><https://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104>

<sup>9</sup>[https://www.theregister.com/2017/07/29/tor\\_dark\\_web/](https://www.theregister.com/2017/07/29/tor_dark_web/)

to connect to “Hidden services”. This puts the estimate at around *19,500 trillion* pages.

**3. What is the volume of internet traffic that flows through the internet daily?**

Based on Cisco predictions of monthly data usages in 2020 <sup>10</sup>,

- Fixed internet traffic: *3.8 EB/day*
- Mobile internet traffic: *1.36 EB/day*

**4. What is the size of the internet in total data today? (Include all media content + text as well as other data in your estimation)**

As of 2019 <sup>11</sup>, around 16 Zettabytes (ZB) of data has been generated till date. This is an aggregate, for both text and media content.

**5. What are the most popular 3-4 web server frameworks for hosting websites? What is their market share?**

Server Name	Percentage market share
Apache	36.3%
Nginx	32.4%
Cloudfare server	15.9%
Microsoft-IIS	7.8%
LiteSpeed	7.1%

Sourced from w3techs <sup>12</sup>

<sup>10</sup><https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>

<sup>11</sup><https://pingiiiit.org/2019/06/the-forgotten-century/>

<sup>12</sup>[https://w3techs.com/technologies/overview/web\\_server](https://w3techs.com/technologies/overview/web_server)