Theory: Units of information

7 minutes 5 / 8 problems solved

Start practicing

Throughout life, we always find something to measure: the amount of food we need to cook for the family, the length and width of that couch you want to put in the room, our weight and height. The latter is especially exciting: it's really cool to learn that in just one year you grew by the full 2 inches!

Each measurement requires an instrument and its own **unit of measurement**. For example, bodyweight is measured with scales in kilograms (or pounds), time is measured with clocks in seconds, etc. But how does one measure information?

§1. Bit: the smallest unit of information

The information entered into the computer should be specific and unambiguous. For a long time, people have used ciphers. The simplest and most convenient of them were digital. Any information from the name of the flowers to the days of the week can be presented in the form of numbers. When processed with a conventional computer, the data is encoded by numbers. They are represented by the electrical signals that the computer works with. For the convenience of distinguishing, signals of two levels are used in classic electronic computers. One of them corresponds to the number `1`, and the other to `0`. Any letter, sound, or image in the computer is represented by a set of numbers. The numbers `1` and `0` are called **binary**. These are the symbols that make up the language understood and used by the computer. Any information on the computer is represented by binary digits: `1`, meaning "**there is a signal**" and `0`, meaning "**no signal**".

> The smallest unit of information is the **bit (b)**.

Each digit of the machine binary code carries the amount of information equal to one bit. It can take only one of two values: either `1` or `0`. It is very inconvenient to measure information in bits because the numbers come out too big.

§2. Byte: a sequence of eight bits

Since people do not consider the mass of ships in grams, larger and hence more convenient units were invented for measuring information as well.

The processing of the information takes place in the processor. This is a device that can work with several bits at once (8, 16, 32, 64, ...). The more bits of information that are processed simultaneously, the faster the computer operation is. The first computers processed 8 bits of information simultaneously, so we needed a new unit of measurement which was called a **byte (B)** that means **8 bits**.

> Bit marks are easily confused with byte marks. Note that the abbreviations for bit numbers use the lowercase letter "**b**" while the bytes are capital "**B**".

§3. Large units of information

There are larger units of information since modern computers process huge amounts of information significantly exceeding bytes.

The computer industry has historically used the units **kilobyte**, **megabyte**, and **gigabyte** in at least two slightly different measurement systems which are slightly contradictory to each other.

- The first one is a decimal-based system, which uses bytes in the powers of ten: **kilobyte** (10^3 bytes), **megabyte** (10^6 bytes), and **gigabyte**

Current topic:
Units of information (Stage 1)
Topic is required for:
Sizes and ranges (Stage 1)
Unicode
The classification of basic types

(10^9 bytes) and so on. These units are used by the International System of Units (SI).
- The second one is a binary-based system which uses bytes in the powers of two: **kilobyte** (2^10 bytes), **megabyte** (2^20 bytes), **gigabyte** (2^30 bytes) and so on. This system was actively used to describe computer memory.

To resolve this confusion, the International Electrotechnical Commission (IEC) suggested to use prefixes **kilo**, **mega** and **giga** only for the decimal-based system and to use new prefixes **kibi**, **mebi**, **gibi** for the binary-based system. Here **bi** means **bi**nary: **kibibyte** is **ki**lo **bi**nary **byte**.

Here is a table with commonly used units of information according to the modern international standards.

SI metric	Symbol	Powers of ten	IEC metric	Symbol	Powers of two
Kilobyte	kB	10^3 B (1000 B)	Kibibyte	KiB	2^10 B (or 1024 B)
Megabyte	MB	10^6 B (1000 kB)	Mebibyte	MiB	2^20 B (or 1024 KiB)
Gigabyte	GB	10^9 B (1000 MB)	Gibibyte	GiB	2^30 B (or 1024 MiB)
Terabyte	TB	10^12 B (1000 GB)	Tebibyte	TiB	2^40 B (or 1024 GiB)
Petabyte	PB	10^15 B (1000 TB)	Pebibyte	PiB	2^50 B (or 1024 TiB)

Of course, not all units of measurement are listed here. We hope, this classification will not cause you any difficulties. It is good that a byte is always 8 bit :) But even this did not come immediately.

Note, that some people and organizations still prefer **kilo**, **mega** and **giga** to describe powers of two. In this course, we follow recommendations of IEC and use modern prefixes **kibi**, **mebi**, **gibi**.

§4. Measurement units conversion

To strengthen your newly obtained knowledge, let's look at the solution of a rather typical problem where you need to convert **1 GiB** to **KiB**. When we convert bigger units into smaller ones, we need to resort to an arithmetic operation called multiplication:

```
1 GiB = 1 * 1024 * 1024 = 1048576 KiB
```

Accordingly, when you need to convert small units into big ones, you use the division. Let's try to convert **16384** bits to **KiB**:

```
16384 bits = (16384 / 8) / 1024 = 2 KiB
```

If you want to convert **1 GB** to **kB**, you should multiply the number by a thousand twice:

```
1 GB = 1 * 1000 * 1000 = 1000000 kB
```

Congratulations, now you have studied one of the basic topics of computer science and are ready to storm new dizzying heights of knowledge.

Report a typo

1768 users liked this theory. **72** didn't like it. **What about you?**

Start practicing

Table of contents:

Comments (33) Hints (0) Useful links (0) Show discussion