Informatica per le Digital Humanities
Lecture 6
March 24 2025
Texts

Slides 2-15
In the previous lectures we have highlighted the role of encoding in digital technology, showing how computers elaborate texts, images and sounds after these phenomena are put in correspondence with numbers. Is it all then just about numbers?

Slides 16-19
The answer is yes and no. No, because not everything happening inside a computer is numbers: there are also the operations that are made on those numbers (like addition, for example), and their execution is a transformation of numbers, that is not a number in itself. Yes, because the description of those operations (called instructions when they are inside a computer program) is inevitably based on an encoding, and so even the description of the operation of addition will exist inside a computer in the form of a number. However, this may be cause for confusion: how can we distinguishing, among the numbers inside a computer, those who are actual numbers to be processed (we call them data) from those that are representing the operations to be executed onto those numbers (also known as instructions)?

Slides 20-30
Enter John Von Neumann, a genius-level polymath who emigrated from Hungary to the USA (after studying in Swizterland) during the Nazi takeover of Europe before and during World War II. His actual name was different, and it is bizarre that he wanted to change his name into something that sounds like coming from a noble German family, given his hatred for that country and its regime back then. In the USA, he was part of the Manhattan project (he is also depicted in the "Oppenheimer" movie by Christopher Nolan) and was very keen to apply his mathematical skills to design missiles and increase their deadly impact. He wanted to bomb Kyoto, the cultural capital of Japan, but the then Secretary of War had very nice memories of his honeymoon there, so Hiroshima and Nagasaki were selected instead. By the way, experiments with nuclear bombs went on also after the end of World War II, and bikinis (scant swimsuits) are named after the islands in the Pacific Ocean where some experiments were conducted, because of their "explosive" effects. Von Neumann died young because of a cancer very likely caused by his excessive exposure to radioactive material. His most famous contribution to digital technology and computers, whose merit he has to share with some contemporaries like Turing, who worked on very similar ideas at the same time, and Maulchy and Eckert, who were part of his research team in the USA, is called "the stored program".

Slides 31-35
The stored program is a paradigm, a way to organize and manage how numbers are stored inside a computer. Its own name explains it: both the numbers representing operands (the data) and the numbers representing operations (the instructions, the program) are going to be stored in the same place. The paradigm is all about taking care of the fact that those two kinds of numbers do not get mixed inside the storage. For an easier visualization for us humans (computer obviously do not have any perceptions of colors, despite enabling their depiction on digital screens) we will use a color scheme in these slides to draw distinctions: data/operand numbers will be depicted in blue, instructions/operation numbers will be depicted in red. Green will be used to indicate the places where those numbers will be stored.

Slides 36-43
The matrix is again useful: not the movie, obviously, but the mathematical device that organizes numbers in a rectangular structure with rows and columns. We can imagine storing the numbers inside a computer inside a memory device that is organized like a matrix.

Slides 44-50
Each row of this matrix-like structure in which numbers are stores is called “word” in the IT jargon. The length of a word is one of the characteristics that determines how a computer’s storage works. We must not forget that the only “numbers” inside a computer are binary digits (or “bits” in short), because for each position in the matrix we will have a high or low level of electrical tension, which gets interpreted as a 1 or a 0, respectively. Computer nowadays have words that are 32 bits or 64 bits long. Since 8 bits form what we call a byte (bits are “b” with a lowercase b, while bytes are “B” with an uppercase B), a word is typically 4 or 8 bytes long. The proportion between bits and bytes is a remnant of a design choice made by computer chip designers in the 1970s.

Slides 51-55
The stored program prescribes the following, to keep operations and operands neatly separated: all words in the storage should be used with a division; the bits before the division are operation bits, and the ones after the division are operand bits. Again, the position of the division is an arbitrary design choice.

Slides 56-58
The stored program is, thus, a way to organize operations and operands in the form of bits inside storage words.

Slides 59-62
A great advantage in having both operands and operations stored as bits is that we can apply the same processing techniques to both. Hence, in the same way as we modify operands with operations, we can also modify the operations themselves. This is why today you can easily update your apps on your mobile phones, for instance: it is just about bits inside the storage being changed accordingly.

Slides 63-69
We introduced the green color code to refer to numbers dealing with the places where operands and operations are stored. The storage is enriched with an address system, which associates a number (an address) to each word in the matrix. Of course, addresses are also expressed in terms of bits that can be stored themselves in the memory.

Slides 70-74
The stored program paradigm, thus, enables us to store and manipulate all sorts of numbers: data, software, and also the addresses indicating where the data and the software are stored. Managing addresses is a fundamental part of the system that enables transfers of data from one part of the computer to another, and also between computers.

Slides 75-84
Typically, we are used to transferring files (from your computer to a USB drive, or from your computer to another computer via email as an attachment, or from your smartphone to your computer via Bluetooth, etc.). Files are conceptually treated as units (a document, an app, a song, etc. visually and textually distinguished by the so-called “file extension” at the end of

their name), but they actually are a sequence of bits inside a computer's storage system. As with physical file cabinets in real life, we need a manager for taking care of the system. In IT terms, this is what we call a "file system". A file system keeps track of all the addresses of all the files inside a computer, to be able to find them and work with them.

Slides 85-89

Files are "logical" unit, meaning that they are treated as a whole, but in the physical reality of the storage unit inside a computer, they may be widespread throughout the storage in different pieces. It is a task of the file system to keep track of the addresses of all these pieces to retrieve them and compose them and use them as a unit when needed. One particular kind of file is a file containing not data (like "my_thesis.docx"), not instructions (like Microsoft Word), but addresses. These are called "folders" and they are presented to you, the end users, with a visual metaphor depicting a folder containing other files, but it is actually just a bunch of addresses pointing at those files.

Slides 90-114

I've been calling the place where bits are stored inside a computer "storage", but its most common name is obviously "memory". I've tried to avoid use this name too often, because its name derives from a metaphor that compares the sotrage of bits inside a computer to the memory that humans use to remember the past. There are some differences and some analogies that deserve further analysis. Whether we remember with our minds or with our hearts, it is obvious that a computer has neither, so the comparison must be drawn with care. One funamental component of memory is time: we remember the past and not the future; we are able to remember a past event because we have somem sort of description of that event that we keep and access to reminisce. Here the subjective experience of a human and the objective, mathematical models used by computers show their difference once again. We could digitize descriptions of past events only in those aspects that can be digitized: images, sounds, and texts, but not tastes or smells.

Slides 115-126

When it comes to memories in terms of a person who remembers and a memory support that helps such act of remembering, there are two important relations that deserve further analysis. The first one is between the person who remembers and the event that is remembered: it is a relation that defines the act of remembering; in other words, just like a tree that falls on a deserted island does not produce a sound (because a sound is a relation between the soundwaves produced by the fall and a functioning ear that catches them) so a person must have attended an event for them to be able to remember it.

Slides 127-133

The other relation is between the person who remembers and the memory support that helps them remember. It is not about the content of the description of a past event, but about the container. Technological issues rise here: of how much use can a faded Polaroid be in helping someone remember about a past event? Memory devices like disks, USB keys, hard drives are supposed to keep the bits that describe past events (in terms of texts, images, sounds) for a very long time, but there is not yet a guarantee that they can stand the test of time.

Slides 134-143

We must not forget that all these devices are digital memory devices, which means that all texts, images, and sounds are encoded into digits first, and then stored in them. This means that not only we are facing the challenge of building devices that can stand the test of time by not breaking down and keeping on functioning for the years, centuries, millenia to come. We must also make sure that the computers of the future that we will use to access these memory devices are going to be compatible with them. Imagine: will there be a USB port in the eyewear/computer of the year 3500 (if we ever get there)? All digitization processes are based on agreements on encodings (like ASCII, JPG, MP3 etc.) that need to be kept in the future, if we want files encoded in the present to be still readable in a 1000 years from now.

Slides 144-147
This is why memory devices that are meant to last for many millennia are conceived in a way that does not depend on any encoding standard or digital technology. Nuclear waste repositories around the worlds contain material that will be harmful for humans for at least 10,000 years. If we put signs in various languages to warn people to stay away, we are assuming that those languages are going to be understood in 10,000 years, which cannot be taken for granted (think of how much we understand of ancient languages of a few thousand years ago). The collective memory of the danger of nuclear waste will likely be lost. Thus, to scare future people away from those dangerous sites, many think that communication based on signs and encodings won't be possible, but rather we will have to rely on a physical barrier with characteristics that are intrinsically repellent to people, because of how they can do to their bodies rather than of what message they carry.