Handout 1

This course is about processing of strings. Lets start with what we mean by *string*. Strings are lists of characters drawn from an *alphabet*. If nothing else is specified, we usually assume the alphabet are letters a, b, ..., z and A, B, ... Z. Sometimes we explicitly restrict strings to only contain the letters a and b. Then we say the alphabet is the set $\{a, b\}$.

There are many ways how we write string. Since they are lists of characters we might write them as "*hello*" being enclosed by double quotes. This is a shorthand for the list

$$[h, e, l, l, o]$$

The important point is that we can always decompose strings. For example we often consider the first character of a string, say h, and the "rest" of a string "*ello*". There are also some subtleties with the empty string, sometimes written as "" or as the empty list of characters $[]$.

We often need to talk about sets of strings. For example the set of all strings

$$\{"", "a", "b", "c", \ldots, "z", "aa", "ab", "ac", \ldots, "aaa", \ldots\}$$

Any set of strings, not just the set of all strings, is often called a *language*. The idea behind this choice is that if we enumerate, say, all words/strings from a dictionary, like

$$\{"the", "of", "milk", "name", "antidisestablishmentarianism", \ldots\}$$

then we have essentially described the English language, or more precisely all strings that can be used in a sentence of the English language. French would be a different set of string, and so on. In the context of this course, a language might not necessarily make sense from a natural language perspective. For example the set of all strings from above is a language, as is the empty set (of strings). The empty set of strings is often written as $\varnothing$ or $\{\}$. Note that there is a difference between the empty set $\{\}$ and the set that contains the empty string $\{""\}$.