# CS331 Uncountablity of Languages

**July 24, 2001** **Amol Mane**

In the previous lecture, we discussed the definitions of alphabet, string and language

Let $\Sigma$ be the alpabet $\{0, 1\}$.
Let the set of all strings of countable length that can be formed from the alphabet $\Sigma$ be denoted by $S$.
Thus, $S = \{00,01,1010, \ldots\}$. Note this also includes strings that are countably infinite in length.
Now consider the languages that can be formed from $\Sigma$. These are nothing but subsets of $S$, e.g., $L_1 = \{00, 01\}$, $L_2 = \{1\}$, $L_3 =$ set of countably infinite length strings in which the number of 0's and number of 1's are equal.
The question arises **"How many such languages can there be ?"**
The answer is **Uncountable**. And we'd like to see how...

Recall a set $S$ is said to be *Countable* if

1. Its cardinality is a natural number and hence is bounded from above, or
2. It is as numerous as the set of natural numbers, i.e., there exists a bijection from $S$ to the set of natural numbers. Such sets are also called countably infinite.

To avoid confusion, we will henceforth call the first category of sets *finite*. The understanding is that whenever a set is finite, its cardinality is bounded from above.
Now consider the set $S$ of strings that can be formed from the alphabet $\Sigma$. This includes all strings of finite length and all strings that are countably infinitely long. Here's an informal hand-waving argument why the cardinality of $S$ must be larger than that of the set of natural numbers.
Suppose we restrict the size of strings in $S$ to be at most $k$.
It's easy to see that if
  k = 1 , $|S| = 2^1 = 2$

k = 2 , $|S| = 2^1 + 2^2 = 6$

And in general, $|S| = 2^1 + 2^2 + ......2^k = 2^{k+1} - 2$.

Note, we have not considered strings of 0 length – we'll soon see such strings, but it is not important for the present uncountability argument.

Therefore $|S| \geq 2^k$. However, in our set $S$ of strings, the size of a string is not really bounded, i.e., $k$ is finite but has no upper bound. So, in an informal sense, we want to have $k$ as large as the cardinality of natural nos ($|N|$). In which case, $|S|$ would grow as the cardinality of the power set of N. And it is well known that the power set of N is uncountable. Thus, $S$ is an uncountable set.

The set of languages formed from $\Sigma$ is nothing but the power set of $S$. Since $S$ itself turns out to be uncountable, the no. of languages that can be formed, i.e., the cardinality of the power set of $S$, is uncountable as well (in a sense, more uncountable than $S$ itself, but we'll not dwell on this now).

So with just two symbols in the alphabet, we blew up quite fast! The set $S$ and the set of languages, both turned out to be uncountable.

Let's see what happens when $\Sigma = \{0\}$, i.e., our alphabet has just one symbol (it wouldn't be very interesting if we didn't have even one symbol to compute with).

With one symbol, cardinality of the set $S$ of strings that can be formed $= 1^1 + 1^2 + \ldots$, which is nothing but the cardinality of N. Thus there exists a bijection from $S$ to the set N of natural numbers, and hence $S$ turns out to be countable. Unfortunately, as before, the no. of languages that can be formed = no. of subsets of $S$ = no. of subsets of N = cardinality of power set of N. And that is known to be uncountable!

So even with a single symbol in our alphabet, uncountable languages can be formed.