Language Definitions and Notations

See Lecture 2

Here are a bunch of basic definitions that we will use all semester. There is nothing exciting here but you won't be able to follow much until you get these in your head.

Σ is a finite set of symbols called our *alphabet*. This could be the set {0,1} of binary digits, or the set of lower-case letters 'a' to 'z'. Don't let the term "alphabet" confuse you. Σ could also be the set of valid Java keywords and identifiers up to length 64 (so it is finite). Any finite set of atomic elements will do.

A *string* or *word* over Σ is any finite sequence of elements of Σ.

ε represents the *empty string*: the string of length 0

Σ^n is the set of strings over Σ of length n (exactly n).

Σ^* is the set of *all* strings over Σ, including the empty string.

Σ^+ is the set of all strings with positive length over Σ.

Obviously, $\Sigma^* = \Sigma^+ \cup \{\varepsilon\}$

Kozen avoids this terminology, but most people say that a *language* over Σ is any subset of Σ^*.

Question 1: How big is Σ^*?

Well, if Σ is the empty set then Σ^* is $\{\varepsilon\}$. If Σ is not empty then Σ^* is countable -- it is a countable union of finite sets.

Question 2: How many languages are there over Σ?

If Σ is empty there are two, both trivial: $\{\}$ and $\{\varepsilon\}$.

If Σ is not empty there are uncountably many languages over it (for if you could number the subsets of Σ^* you could create a new subset that wasn't in any of them.