Regular Expressions

A *regular expression* A is a string (or pattern) formed from the following 6 pieces of information: $a \in \Sigma$, ϵ, $\emptyset$, and the operations: $+$, ., and $*$.

We think of a regular expression as pattern which can be matched by strings from Σ. The language of A, $L(A)$ is equal to all those strings which match A, $L(A) = \{x \in \Sigma^* | x \text{ matches } A\}$.

For any $a \in \Sigma$, $L(a) = a$.

$L(\epsilon) = \{\epsilon\}$

$L(\emptyset) = \emptyset$

$+$ functions as an **or**, $L(A + B) = L(A) \cup L(B)$.

. creates a product structure, $L(AB) = L(A)L(B)$.

* denotes concatenation, $L(A^*) = \{x_1 x_2 \ldots x_n \,|\, x_i \in L(A) \text{ and } n \geq 0\}$

Example The regular expression $(ab)^*$ matches the set of strings: $\{\epsilon$, ab,$abab$, $ababab$, $abababab, \ldots\}$.

Example The regular expression $(aa)^*$ matches the set of strings on one letter which have even length.

Example The regular expression $(aaa)^* + (aaaaa)^*$ matches the set of strings of length equal to a multiple of 3 or 5.

We have seen that NFAs seem to be more compact than DFAs, but now we give a family of examples where NFAs are exponentially smaller than any DFA expressing the same language. Consider strings with the nth bit from the right equal to 1. We can represent this with an NFA on $n + 1$ states (see Figure 1). However, if we try to represent such strings with a DFA, we must use 2^n states. (why?)

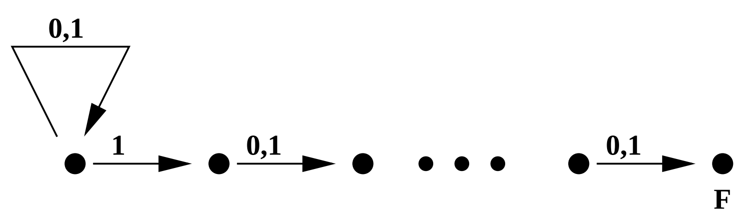

Figure 1: An NFA with n+1 states.

Next we will see that regular expressions are equivlalent to DFAs (and hence NFAs) in that they express the same languages.

Theorem. *For any NFA N there is a regular expression r s.t. $L(N) = L(r)$. Conversely, for any regular expression r there exists an NFA N s.t. $L(r) = L(N)$.*

Example Figure 2 shows an NFA for the following regular expression:

$(11 + 0)^*(00 + 1)^*$.

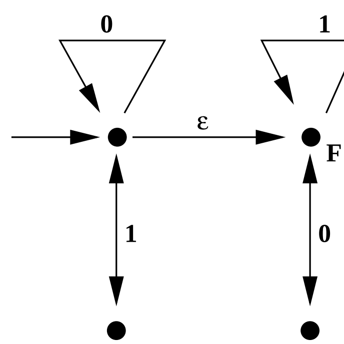

Figure 2: An NFA using an ϵ transition.

Suppose we are working over an alphabet of just one element, $\Sigma = \{a\}$. Then a collection A is regular iff the set $\{n|a^n \in A\}$ is such that after some value k, any $n \geq k$ $n \in A$ iff $n + t \in A$. Namely the set becomes periodic. The equivalent NFA has the form of Figure 3

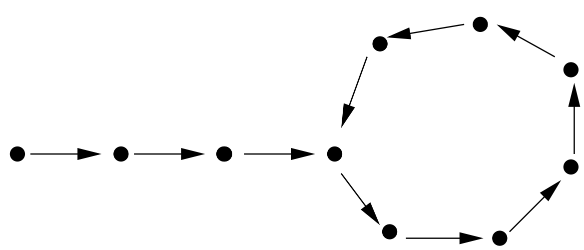

Figure 3: