# Properties of Regular Expressions

Let $R$ and $S$ be regular expressions over a finite alphabet $\Sigma$.

1. Associativity w.r.t $\cup$: $(R \cup S) \cup T = R \cup (S \cup T)$
2. Identity w.r.t. $\cup$: $R \cup \emptyset = R$.
3. Commutativity w.r.t. $\cup$: $R \cup S = S \cup R$.
4. Idempotence w.r.t. $\cup$: $R \cup R = R$.
5. Associativity w.r.t $\circ$: $(R \circ S) \circ T = R \circ (S \circ T)$
6. Identity w.r.t $\circ$: $R \circ \epsilon = R$.
7. Nullity w.r.t. $\circ$: $R \circ \emptyset = \emptyset$.
8. Distributivity:
   (a) $R \circ (S \cup T) = (R \circ S) \cup (R \circ T)$.
   (b) $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$.
9. $(R^*)^* = R^*$.
10. $\emptyset^* = \epsilon$
11. $\epsilon^* = \epsilon$

## DFA to Regular Expressions
## (Ref: "Automata and Complexity" by Dexter Kozen)

- Let $M = (Q, \Sigma, \delta, s, F)$ be a DFA.
- Let $p, q \in Q$ and $X \subseteq Q$.
- Let $R_{pq}^X$ denote a regular expression for the set of strings $w$ over $\Sigma$ such that:
  - there is a path from $p$ to $q$ in $M$ with label $w$, and
  - all *intermediate* states along this path are from the set $X$.
- The union over all final states $f \in F$ of $R_{sf}^Q$ represents $L(M)$: $L(M) = \bigcup_{f \in F} R_{sf}^Q$.
- Construct $R_{pq}^X$ using induction on $X$.
  - Basis: $X = \emptyset$.
    * Let $p \neq q$.
    * If no transition from $p$ to $q$ in $M$:
    $$R_{pq}^{\emptyset} = \emptyset$$
    * If $a_1, \cdots, a_k$ are the symbols in $\Sigma$ from which there are transitions from $p$ to $q$ in $M$:
    $$R_{pq}^{\emptyset} = a_1 \cup a_2 \cup \cdots \cup a_k$$
    * Let $p = q$.
    * If no transition from $p$ to $q$ in $M$:
    $$R_{pq}^{\emptyset} = \epsilon$$
    * If $a_1, \cdots, a_k$ are the symbols in $\Sigma$ from which there are transitions from $p$ to $q$ in $M$:
    $$R_{pq}^{\emptyset} = a_1 \cup a_2 \cup \cdots \cup a_k \cup \epsilon$$

## DFA to Regular Expressions (continued)

- Induction: $X \neq \emptyset$.
  - Let $P$ be any path in $M$ from $p$ to $q$ with all intermediate states in $X$.
  - Let $r \in X$.
  - Either $P$ does not visit $r$. Then,

$$R_{pq}^{X} = R_{pq}^{X - \{r\}}$$

  - Or $P$ visits $r$ at least once. In this case, $P$ can be split into the following pieces:
    - The first piece is from $p$ to $r$ without visiting $r$, the middle pieces go from $r$ to $r$ without visiting $r$ in between, and the last piece goes from $r$ to $q$ without visiting $r$. Then,

$$R_{pq}^{X} = R_{pr}^{X - \{r\}} \circ (R_{rr}^{X - \{r\}})^{*} \circ R_{rq}^{X - \{r\}}$$

  - So, the resulting expression is:

$$R_{pq}^{X} = R_{pq}^{X - \{r\}} \cup (R_{pr}^{X - \{r\}} \circ (R_{rr}^{X - \{r\}})^{*} \circ R_{rq}^{X - \{r\}})$$

- A bottom-up algorithm:
  - Let $|Q| = n$. Let the states in $Q$ be $\{1, 2, \cdots n\}$.
  - The basis case is as above.
  - Assume that $R_{pq}^{X}$ has been constructed for $X = \{1, 2, \cdots, k - 1\}$.
  - To construct $R_{pq}^{X}$ for $X = \{1, 2, \cdots, k\}$.
  - Choose $r$ in the above proof to be the state $k$.
  - Then, for all $1 \leq i, j \leq n$, we have

$$R_{ij}^{X} = R_{ij}^{X - \{k\}} \cup (R_{ik}^{X - \{k\}} \circ (R_{kk}^{X - \{k\}})^{*} \circ R_{kj}^{X - \{k\}}).$$