Lecture 7

Algebra of regular expressions: two constants ε, ϕ, one const a for every $a \in \Sigma$, Two binary ops $+, \cdot$ and one unary op $*$

Laws

1. $R+\phi = \phi+R = R$
2. $R+S = S+R$
3. $R+(S+T) = (R+S)+T$
4. $R+R = R$

(1–4: exactly the laws of union.)

5. $R\cdot\phi = \phi\cdot R = \phi$
6. $R\cdot\varepsilon = \varepsilon\cdot R = R$
7. $R\cdot(S\cdot T) = (R\cdot S)\cdot T$
8. $R\cdot(S+T) = R\cdot S + R\cdot T$
9. $(S+T)\cdot R = S\cdot R + T\cdot R$
10. $\varepsilon + RR^* = \varepsilon + R^*R = R^*$

Proof of (5) $R\cdot S$ ~~mean~~ stands for all words of the form xy with $x \in R$ & $y \in S$. Now if $x \in R\cdot\phi$ then it has to be possible to break x into yz with $y \in R$ & $z \in \phi$. But there is no word in ϕ so no such decomposition is possible so $R\cdot\phi = \phi$.

(7) works because we are working with words & not trees.

(10) suppose $x \in \varepsilon + RR^*$ then either $x \in \varepsilon$ i.e. $x = \varepsilon$ so $x \in R^*$ or $x \in RR^*$ so $x = yz$ with $y \in R$ & $z \in R^*$ but then $z = z_1 \cdots z_k$ with all $z_i \in R$ so $x = yz_1 \cdots z_k$ with y & all $z_i \in R^*$ i.e. $x \in R^*$. Reverse direction is similar.

Other equations $R^{**} = R^*$ clearly $R^* \subseteq R^{**}$

suppose $x \in (R^*)^*$ then $x = x_1 \cdots x_k$ with each $x_i \in R^*$ but then each $x_i = x_i^1 \cdots x_i^{d_i}$ with each $x_i^l \in R$ so we get $x = x_1^1 x_1^2 \cdots x_1^{d_1} x_2^1 x_2^2 \cdots x_2^{d_2} \cdots x_k^{d_k}$ which is just a concatenation of words from R so $x \in R^*$ so $R^{**} \subset R^*$. i.e. $R^{**} = R^*$

$(R^*S)^* R^* = (R+S)^*$

clearly $(R^*S)^* R^* \subseteq (R+S)^*$

Now suppose $\omega \in (R+S)^*$ so $\omega = \omega_1 \dots \omega_n$ with each $\omega_i \in R$ or in S. Let us focus on words from S

$(\omega_1\ \omega_2 \cdots \underline{\omega_i})(\omega_{i+1} \cdots \underline{\underline{\omega_j}}) \dots \omega_n$

So each S word has some number of R words before it since the last S word. We note that each of these packets can be viewed as a word in R^*S, there may be several of them so overall $(R^*S)^*$ and there may be some more R words at the end so finally $\omega \in (R^*S)^* R^*$. Note any of the $*$ things can be absent so we can have a pure R^* word or a pure S^* word.