Oggi introdurremo l'analisi lessicale.
Vediamo innanzitutto le **espressioni regolari**. Le seguenti regole definiscono le *espressioni regolari di un alfabeto* Σ.

Sia Σ un alfabeto, cioè un insieme di simboli. Da Σ si definiscono in maniera induttiva le espressioni regolari. Si dice cioè quali sono le minime espressioni regolari, gli operatori che possiamo usare fra tali espressioni; utilizzando espressioni regolari e operatori possiamo comporre espressioni più grandi.
La specifica di un'espressione regolare è un'esempio di definizione ricorsiva, in quanto fa uso di base e passo induttivi:

Base:

1 ε è un'espressione regolare, che denota il linguaggio $\{\varepsilon\}$.

2 Se a è un simbolo di Σ, allora a è un'espressione regolare che denota il linguaggio $\{a\}$.

Passo:

Siano r, s due espressioni regolari, che rispettivamente denotano i linguaggi $L(r)$ e $L(s)$

Allora valgono:

1. $r|s$ è un'espressione regolare che denota $L(r) \cup L(s)$.
2. rs (meno comunemente, $r \cdot s$) è un'espressione regolare che denota $L(r)L(s) := \{w = w_1 w_2 \ \& \ w_1 \in L(r) \ \& \ w_2 \in L(s)\}$.
3. r^* è un'espressione regolare che denota $L(r^*) = \{\varepsilon\} \cup \{w_1 \dots w_k \mid \forall i = 1, \dots, k, \ w_i \in L(r_1)\}$.
4. (r) è un'espressione regolare che denota $L(r)$.

   Esempio: $L(a) = \{a\}$, $L((a)) = \{a\}$.

Le parentesi vengono utilizzate per meglio evidenziare l'ordine di precedenza e associatività.
Convenzioni su precedenza e associatività degli operatori:

- $*$ Ha la precedenza più alta
- $\cdot$ Ha precedenza inferiore a $*$
- $|$ Ha precedenza inferiore a $\cdot$

associativi a sinistra.

Esempio: $a|b^*c$ 1. $a|(b^*)c \rightarrow$ 2. $a|((b^*)c)$

Sia $r = a|b^*c$. $L(r) = L(a) \cup L(b^*c)$ $\cdot \ L(b^*c) = \{w \mid w = w_1 w_2 \ \& \ w_1 \in L(b^*) \ \& \ w_2 \in L(c)\}$

$\cdot \ L(c) = \{c\} \Rightarrow w_2 = c$

$= \{wc \mid w \in L(b^*)\}$

$= \{a\} \cup \{b^n c \mid n \geq 0\}$

Facciamo alcuni esempi

r	L(r)
$a\|b$	$\{a, b\}$
$ab\|b$	$\{ab, b\}$
$a(b\|c)$	$\{ab, ac\}$
a^*b^*	$\{a^n b^j \mid n, j \geq 0\}$ → si può fare di meglio: $\{a^n \mid n \geq 0\}$
$a\|a^*b$	$\{a\} \cup \{a^n b \mid n \geq 0\}$

Ora proviamo il contrario.

L	r
Identificatori, cioè stringhe, che cominciano con una lettera dell'alfabeto, cui seguono lettere o numeri.	$(a\|\dots\|z)(a\|\dots\|z\|0\|\dots\|9)^*$
Numeri binari multipli di 2	✓ $(0\|1)^*0$ E' possibile che cominci con 0, ma per la prof vale "full mark" ✓ $1(0\|1)^*0$ Va bene... dipende se vogliamo considerare zero un multiplo di 2 o no. Volessimo aggiungere la possibilità "0": $(1(0\|1)^*)\|0$
Stringhe di 'a' e di 'b' con 'a' consecutive	✓ $(a\|b)^*aa(a\|b)^*$
" senza "	MANCA ✗ $(ab^*\|b)^*$ NO, "AA" $\notin L(r)$