

Set standard di stringhe

- A^n = Insieme di tutte le stringhe di A di lunghezza n
- A^+ = Insieme di tutte le stringhe di A di lunghezza maggiore di 0
- A^* = Insieme di tutte le stringhe di A

Provable facts

- $A^0 = \{\epsilon\}$
- $A^+ = \bigcup_{n>0} A^n$
- $A^* = \bigcup_{n>0} A^n = A^0 \cup A^+$

Linguaggio Formale

Definizione: Un linguaggio L creato a partire da un alfabeto A è sottoinsieme di A^* .

Operatori:

- **Concatenazione:** $L1 \bullet L2$
è ogni possibile combinazione delle stringhe dei due insiemi
- **Unione:** $L1 \cup L2$
unisce le stringhe dei due insiemi in una sola stringa

Concatenazione per n volte

L^n L concatenato con se stesso per n volte.

esempio:

$$L^2 = L \bullet L$$

$$L^0 = \emptyset$$

L^+ concatenazione qualsiasi di stringhe nell'alfabeto A, con lunghezza 1

L^* concatenazione qualsiasi di stringhe nell'alfabeto A

Espressioni regolari

Formalismo usato comunemente per definire linguaggi semplici; le espressioni regolari (E.R.) **definiscono un linguaggio sintatticamente**.

Riconoscimento delle E.R.

Casi base:

- L'insieme vuoto è una E.R.
- Una stringa vuota è una E.R.
- Ogni stringa di lunghezza 1 è una E.R.

Casi induttivi:

- L'**Unione** di due E.R. è una E.R.
- La **Concatenazione** di due E.R. è una E.R.
- La **Kleene Star** di una E.R. è una E.R.
 - se x è una E.R. allora x^* è una E.R.
- **Parentesi**:
 - se x è una E.R. allora (x) è una E.R.

Semantica di un'Espressione Regolare

- **empty set**: \emptyset is a regular expression over A
- **singleton set with the empty string**: ϵ is a regular expression over A
- **singleton set with a string of length 1**:
for all $\sigma \in A$, σ is a regular expression over A

Regole di Precedenza tra Operatori

Dall'operatore con massima a quello con minima precedenza:

1. Parentesi
2. Kleen Star
3. Concatenazione
4. Unione
5. Tra due concatenazioni o due unioni viene eseguita prima quella a sinistra per la proprietà di **associatività a sinistra**.

Caratteri Speciali

- `\t`: tab
- `\n`: newline (=line feed)
- `\s`: any white space character
- `\S`: any non-white space character
- `\d`: any digit character, `\d` an abbreviation of `[0-9]`
- `\D`: any non-digit character, `\D` an abbreviation of `[^\d]`
- `\w`: any word character
`\w` an abbreviation of `[a-zA-Z0-9_]` (underscore `_` is allowed)
- `\W`: any non-word character, `\W` an abbreviation of `[^\w]`