

Regular Expressions

Prof. (Dr.) K.R. Chowdhary

Former Professor & Head, Department of Computer Sc. & Engineering
MBM Engineering College, Jodhpur

K.R. Chowdhary

Theory of Computation

Automata, Formal Languages, Computation and Complexity

Focuses on pedagogy in its writing, that represents a refreshing approach

Ensures comprehensive and enjoyable learning

Undergone a rigorous classroom testing



© 2025



Get 20% off with this code: **SPRAUT**

Available on Springer Nature Link



[link.springer.com/book/
9789819762347](https://link.springer.com/book/9789819762347)

Please note that promotional coupons are only valid for English-language Springer, Apress, and Palgrave Macmillan books & eBooks and are redeemable on link.springer.com only. Titles affected by fixed book price laws, forthcoming titles and titles temporarily not available on Springer Nature Link are excluded from promotions, as are reference works, handbooks, encyclopedias, subscriptions, or bulk purchases. The currency in which your order will be invoiced depends on the billing address associated with the payment method used, not necessarily your home currency. Regional VAT/tax may apply. Promotional prices may change due to exchange rates. Promotions are valid for individual customers only. Booksellers, book distributors, and institutions such as libraries and corporations please visit springernature.com/contact-us. Promotions do not work in combination with other discounts or gift cards.

Regular Languages

- The set R on an alphabet Σ is regular language, defined as follows:
 $\emptyset \subseteq R$, $\{\epsilon\} \subseteq R$, $\{a\} \subseteq R$ for all $a \in \Sigma$
if $L_1, L_2 \subseteq R$, then $L_1 \cup L_2 \subseteq R$, $L_1 \circ L_2 \subseteq R$, and $L_1^* \subseteq R$.
- The language $\{0, 00, 01, 000, 001, 010, 011, 0000, \dots\}$, which consists all binary strings is regular. It can be constructed using regular expression $\{0\} \circ (\{0\} \circ \{1\})^*$.
- A Regular language is represented by regular expression.

Definition

Following are regular expressions:

- 1 $\emptyset, \epsilon$, and $a \in \Sigma$ are regular expressions.
 - 2 if α, β are regular expressions, the $\alpha \cup \beta$, $\alpha \circ \beta$, and α^* are also regular expressions.
 - 3 Nothing else is regular expression.
- A language is regular *iff* there is regular expression to represent it.

Definition

Every regular expression has functional mapping to a regular language, as follows:

- The regular languages corresponding to regular expressions ϕ, ϵ are: ϕ , and $\{\epsilon\}$, respectively.
 - if α, β are reg. expressions, then corresponding regular languages are: $\mathcal{L}\{\alpha\}$, and $\mathcal{L}\{\beta\}$.
 - Also, $\mathcal{L}\{\alpha \circ \beta\} = \mathcal{L}\{\alpha\} \mathcal{L}\{\beta\}$, $\mathcal{L}\{\alpha \cup \beta\} = \mathcal{L}\{\alpha\} \cup \mathcal{L}\{\beta\}$,
 $\mathcal{L}\{\alpha^*\} = \mathcal{L}\{\alpha\}^*$
-
- If L is regular, then is $\bar{L}$ also regular (to be seen in the study of finite automata)
 - Is the intersection of two regular languages also necessarily regular? (to be seen later)
 - Are all the languages regular? Justify (to be seen later)

Regular Languages

- All the regular expressions can be listed of increasing length. Hence, they can be mapped with set $\mathbb{N}$. Thus regular expressions are **countable**. However, set of all the possible languages on set Σ^* is power set $2^{|\Sigma^*|}$, which is not countable. **Conclusion: There are not enough regular expressions to represent all languages.**
- Language over $\Sigma = \{a, b\}$ for regex $(a + b)^*$ is: $L = \mathcal{L}((a + b)^*) = \{\epsilon, a, b, aa, ab, ba, bb, aaa, \dots\}$, which is countably infinite set. Thus, L is *bijection* to the set $\mathbb{N}$. Thus, L is called **Recursively Enumerable language (RE)**.

Definition

(RE). A language L is RE if there exists an algorithm that, when given an input w , eventually halts *iff* $w \in L$. Equivalently, there is an algorithm that enumerates the members of L . If necessary, this algorithm may run forever. Thus, L is called *semidecidable language*.

Regular Languages

Consider the language $L = \{a, ab, bba, bca, abcd\}$, which has bijection with a set $n \in \mathbb{N}$, such that $|n| = |L|$. The mapping of bijection is $\{(0, a), (1, ab), (2, bba), (3, bca), (4, abcd)\}$. The language, L above is called **recursive.(R)**

Definition

(Recursive). A language L is recursive if there exists an algorithm that, given an input word w , will determine in a finite amount of time if $w \in L$ or not. A recursive language is **decidable**.

- A recursive (**R**) language is enumerable. Hence, $\mathbf{R} \subseteq \mathbf{RE}$.
- A set of programs that do not terminate on some inputs are **not RE**.
- Since regular languages are countable, and for each countable set there are uncountable subsets of languages. Thus, there exists languages which are not regular.
- **Conclusion: All languages are not the regular languages.**



Chowdhary, K.R. (2025). Finite Automata and Regular Expressions.
In: Theory of Computation. Springer, Singapore.
https://doi.org/10.1007/978-981-97-6234-7_2