CS310 : Automata Theory 2019

Lecture 8: Properties of regular languages

Instructor: Ashutosh Gupta

IITB, India

Compile date: 2019-01-21

Topic 8.1

Practical matching

Regular expressions example

Example 8.1

Find the following string in the piazza group and paste in `regex101.com`.

```
Listen close to what I gotta say
Cause you know there ain't no other way
Love is the message
You ready?
Let's go
Yeah.. we can make it better
Yeah.. when we come together
Yeah.. all you got is me and
Yeah.. all I got is you!
Ishq se aage?
Kuch nahi kuch nahi kuch...
Ishq se behtar?
Kuch nahi kuch nahi kuch...
Ishq se upar?
Kuch nahi kuch nahi kuch...
Ishq bina hum kuch nahi...
```

How many words are

- there? `\w+`
- longer than 3? `\w{3,}`
- capitalized? `\[A-Z]w*`
- with at least two vowels? `(\w*[aeiou]){2,}\w*`

Exercise : finding patterns in text

Exercise 8.1

Let us continue using the text

- How many words are with at least three consonants?
- Find two consecutive lines that have "we"?
- Find lines that end with dots?
- Find lines that do not have "g"?
- Find lines that do not have "g" but have "t" ?

Regular languages

We have seen several ways of describing regular languages

Now we will look into some properties of

- regular expressions and
- regular languages.

Topic 8.2

Algebraic laws of regular expressions

Regular expression Laws

Recall, we needed simplifications of regular expressions.

We used informal arguments to simplify

However, some of the arguments can be formally written as algebraic laws.

Associativity and commutativity

Let L and M be regular languages

- $L + M = M + L$
- $L + (M + N) = (L + M) + N$

- $LM = ML$ ✗
- $L(MN) = (LM)N$

Identities and annihilators

- $\emptyset + L = L + \emptyset = L$
- $\epsilon L = L \epsilon$
- $\emptyset L = \emptyset$

Union behaves like addition and concatenation behaves like multiplication.

Distributive Laws

- $L(M + N) = LM + LN$
- $(M + N)L = ML + NL$

Laws for closure

- $(L^*)^* = L^*$
  - Consider a word $\underbrace{w_{11}w_{12}...w_{1_{n_1}}}_{L^*} ... \underbrace{w_{m1}w_{m2}...w_{m_{n_m}}}_{L^*} \in (L^*)^*$, where $w_{ij} \in L$.
  - Clearly, the word is also in L^*
- $\emptyset^* = \epsilon$
- $L^+ = LL^* = L^*L$
- $\epsilon + L^* = L^*$
- $L^* = L^+ + \epsilon$

Exercise 8.2

Simplify the following regular expression

$$\begin{aligned} 0 + 01^* &= 0(\epsilon + 1^*) && \textit{(distributive law)} \\ &= 0(1^*) && (\epsilon \subset 1^*) \end{aligned}$$

Proving laws of regular expression

Let R be a regular expression with variables $L_1, .., L_n$.

Let R' be the concrete regular expression by replacing $L_1, ..., L_n$ by fresh symbols $a_1, ..., a_n$ in R.

Theorem 8.1

If there is a word $w_1 w_2 ... w_n \in R$ such that $w_i \in L_{j_i}$, then there is a word $a_{j_1}, ..., a_{j_n} \in R'$.

Proof sketch.

Since R and R' have same structure, the word produced in R must have a counterpart in R'. We can formally prove it via structural induction. □

Due to the above theorem, we only to analyze concrete languages for proving laws of regular expressions.

Example: proving laws of regular expression

Example 8.2

Consider law

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

Let us replace L *by* 0 *and* S *by* 1. *We obtain* $(0 + 1)^* = (0^* 1^*)^*$.

We can see both the expressions are all the words consists of 0 *and* 1.

Therefore, the above law holds.

Exercise 8.3

Prove or disprove $(RS + R)^* R = R(SR + R)^*$

Topic 8.3

Closure properties of regular languages

Notation alert : Closure properties

Consider a set of objects S, which is subset of universe U.

Let $op : U \times U \to U$ be an operator over U.

Definition 8.1

S is closed under op if for each $e_1, e_2 \in S$, $op(e_1, e_2) \in S$.

Example 8.3

Consider the universe of polynomials. The set of linear terms is a subset of the universe.

- *Linear terms are closed under addition operator*
- *Linear terms are not closed under multiplication operator*

Regular language

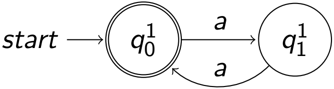
The set of regular languages is subset of all languages.
We ask for the closure properties.

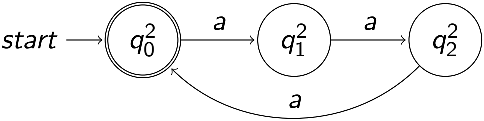
Are they closed under ______?

- Intersection
- Union
- Negation

Example: intersection of DFAs

Example 8.4 *Let $\Sigma = \{a\}$. Consider the following automata*

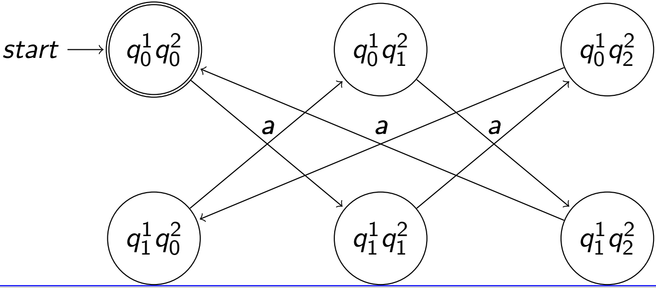




$L(A_1) = \{w \mid |w| \text{ mod } 2 = 0\}$ $\quad$ $L(A_2) = \{w \mid |w| \text{ mod } 3 = 0\}$

Clearly $L(A_1) \cap L(A_2) = \{w \mid |w| \text{ mod } 6 = 0\}$.

Intuitively, running on both automata, i.e., cross product of states.



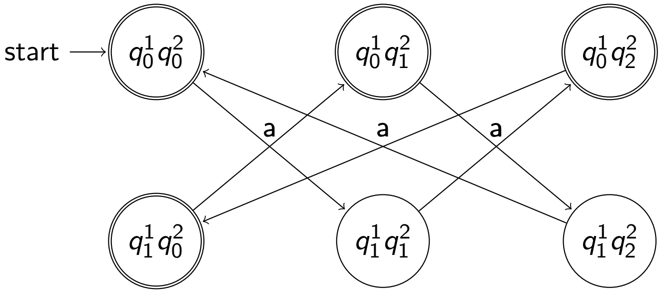
Example: Union

For union, we again run both the automaton.

But, any one of the accepting state of the two automata can accept the word.

Example 8.5

The following accepts $L(A_1) \cup L(A_2)$.



Notation altert : cross product

Definition 8.2

Let S_1 *and* S_2 *be sets.* $S_1 \times S_2 = \{(s_1, s_2) \mid s_1 \in S_1 \wedge s_2 \in S_2\}$.

Regular languages are closed under intersection

Theorem 8.2

If L_1 and L_2 are regular languages, $L_1 \cap L_2$ is also a regular language.

Proof.

Let DFA's $(Q^1, \Sigma, \delta^1, q_0^1, F^1)$ and $(Q^2, \Sigma, \delta^2, q_0^2, F^2)$ recognize L_1 and L_2 respectively.

We construct the following automaton for recognizing $L_1 \cap L_2$.

$$A = (Q^1 \times Q^2, \Sigma, \delta, (q_0^1, q_0^2), F^1 \times F^2)$$

where $\delta((q, q'), a) \triangleq (\delta^1(q, a), \delta^2(q', a))$ for each $q \in Q^1$,$q' \in Q^2$, and $a \in \Sigma$.

We show

$$\hat{\delta}((q_0^1, q_0^2), w) = (\hat{\delta^1}(q_0^1, w), \hat{\delta^2}(q_0^2, w)).$$

...

Regular languages are closed under intersection II

Proof.

base case:

Let $w = \epsilon$.

$\hat{\delta}((q_0^1, q_0^2), \epsilon) = (q_0^1, q_0^2)$ due to the def of extended transitions

$\hat{\delta^1}(q_0^1, \epsilon) = q_0^1$ and $\hat{\delta^2}(q_0^2, \epsilon) = q_0^2$ due to the def of extended transitions

Therefore,

$$\hat{\delta}((q_0^1, q_0^2), \epsilon) = (\hat{\delta^1}(q_0^1, \epsilon), \hat{\delta^2}(q_0^2, \epsilon))$$

...

Regular languages are closed under intersection III

Proof(contd.)

induction step:

Let $w = xa$, where x is a word in Σ^* and x is a letter in Σ.

Due to induction hypothesis, we assume
$\hat{\delta}((q_0^1, q_0^2), x) = (\hat{\delta^1}(q_0^1, x), \hat{\delta^2}(q_0^2, x)) = (q, q')$.

Due to the definition $\hat{\delta}$, $\hat{\delta}((q_0^1, q_0^2), xa) = \delta((q, q'), a) = (\delta^1(q, a), \delta^2(q', a))$.

Due to the definition $\hat{\delta^1}$, $\hat{\delta^1}(q_0^1, xa) = \delta^1(q, a)$.
Due to the definition $\hat{\delta^2}$, $\hat{\delta^2}(q_0^2, xa) = \delta^2(q', a)$.

Therefore, $\hat{\delta}((q_0^1, q_0^2), xa) = (\hat{\delta^1}(q_0^1, xa), \hat{\delta^2}(q_0^2, xa))$. …

Regular languages are closed under intersection IV

Proof(contd.)

claim: $L(A) = L_1 \cap L_2$
w is accepted in A, $\hat{\delta}(q_0, w) = (q, q') \in F^1 \times F^2$.

- Iff, $F^1 \ni q = \hat{\delta^1}(q_0^1, w)$ and $F^2 \ni q' = \hat{\delta^2}(q_0^2, w)$.
- Iff, $w \in L_1$ and $w \in L_2$. $\square$

Regular languages are closed under union

Theorem 8.3

If L_1 and L_2 are regular languages, $L_1 \cup L_2$ is also a regular language.

Proof.

Let DFA's $(Q^1, \Sigma, \delta^1, q_0^1, F^1)$ and $(Q^2, \Sigma, \delta^2, q_0^2, F^2)$ recognize L_1 and L_2 respectively.

We build the following automaton for recognizing $L_1 \cup L_2$.

Only accepting states are different

$$A = (Q^1 \times Q^2, \Sigma, \delta, (q_0^1, q_0^2), \mathbf{F^1} \times \mathbf{Q^2} \cup \mathbf{Q^1} \times \mathbf{F^2})$$

where $\delta((q, q'), a) \triangleq (\delta^1(q, a), \delta^2(q', a))$ for each $q \in Q^1$,$q' \in Q^2$, and $a \in \Sigma$.

The proof follows similar structure as intersection closure. □

Regular languages are closed under complement

Theorem 8.4
If L is a regular language over alphabet Σ, $\Sigma^* - L$ is also a regular language.

Proof.
Let $(Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes L.

We build the following automaton for recognizing $\Sigma^* - L$.

$$A = (Q, \Sigma, \delta, q_0, Q - F)$$

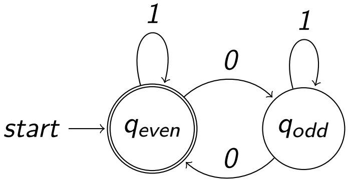
Those words that do not reach the final states F are accepted by A. Clearly, $L(A) = \Sigma^* - L$. □

Exercise 8.4
If L and M are regular languages. $L - M$ is also regular,

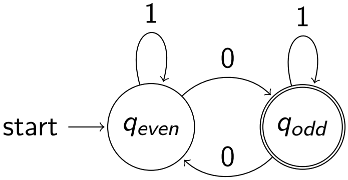
Example: language complement

Example 8.6

Let $\Sigma = \{0, 1\}$. *Let* $L = \{w|w \text{ has even number of 0s}\}$.



$\Sigma^* - L = \{w|w$ has odd number of 0s$\}$.



End of Lecture 8