Language

Definition

A *language* over alphabet A is a set $L \subseteq A^*$. Example for $A = \{0,1\}$:

- a finite language like $L = \{1, 10, 1001\}$ or the empty language $\varnothing$
- infinite but very difficult to describe (there are random languages: there exist more languages as subsets of A^* than there are finite descriptions)
- infinite but having some nice structure, where words follow a certain "pattern" that we can describe precisely and check efficiently $\leftarrow$ these are our focus

$L_2 = \{01, 0101, 010101, \ldots\}$ = those non-empty words that are of the form $01\ldots01$ where the block 01 is repeated some finite positive number of times. Using notation $(01)^n$ for a word consisting of block 01 repeated n times, we can write $L_2 = \{(01)^n \mid n \geq 1\}$.

Languages are sets, so we can take their union ($\cup$), intersection ($\cap$), and apply other set operations on languages.

Languages $\varnothing$ and $\{\varepsilon\}$ are very different: $\varnothing$ is a set that contains no words, whereas $\{\varepsilon\}$ contains precisely one word, the word of length zero.

More Language Operations

In addition to set operations, languages support operations defined on their words.

Definition (Language concatenation)

Given $L_1 \subseteq A^*$ and $L_2 \subseteq A^*$, define $L_1 \cdot L_2 = \{w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$

Example: $\{\varepsilon, \mathrm{a}, \mathrm{aa}\} \cdot \{\mathrm{b}, \mathrm{bb}\} = \{\mathrm{b}, \mathrm{bb}, \mathrm{ab}, \mathrm{abb}, \mathrm{aab}, \mathrm{aabb}\}$

In other words, $w \in L_1 L_2$ iff w can be split into $w_1 \in L_1$ and $w_2 \in L_2$ s.t. $w = w_1 w_2$.

More Language Operations

In addition to set operations, languages support operations defined on their words.

Definition (Language concatenation)

Given $L_1 \subseteq A^*$ and $L_2 \subseteq A^*$, define $L_1 \cdot L_2 = \{w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$

Example: $\{\varepsilon, \mathrm{a}, \mathrm{aa}\} \cdot \{\mathrm{b}, \mathrm{bb}\} = \{\mathrm{b}, \mathrm{bb}, \mathrm{ab}, \mathrm{abb}, \mathrm{aab}, \mathrm{aabb}\}$
In other words, $w \in L_1 L_2$ iff w can be split into $w_1 \in L_1$ and $w_2 \in L_2$ s.t. $w = w_1 w_2$.

Definition (Language exponentiation)

Given $L \subseteq A^*$, define

$$L^0 = \{\varepsilon\}$$
$$L^{n+1} = L \cdot L^n$$

More Language Operations

In addition to set operations, languages support operations defined on their words.

Definition (Language concatenation)

Given $L_1 \subseteq A^*$ and $L_2 \subseteq A^*$, define $L_1 \cdot L_2 = \{w_1 \cdot w_2 \mid w_1 \in L_1, w_2 \in L_2\}$

Example: $\{\varepsilon, \mathrm{a}, \mathrm{aa}\} \cdot \{\mathrm{b}, \mathrm{bb}\} = \{\mathrm{b}, \mathrm{bb}, \mathrm{ab}, \mathrm{abb}, \mathrm{aab}, \mathrm{aabb}\}$
In other words, $w \in L_1 L_2$ iff w can be split into $w_1 \in L_1$ and $w_2 \in L_2$ s.t. $w = w_1 w_2$.

Definition (Language exponentiation)

Given $L \subseteq A^*$, define

$$L^0 = \{\varepsilon\}$$
$$L^{n+1} = L \cdot L^n$$

Theorem

Given $L \subseteq A^*$, we have $L^n = \{w_1 \dots w_n \mid w_1, \dots, w_n \in L\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\emptyset = \{w_1 w_2 \mid w_1 \in L, w_2 \in \emptyset\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\} L = L$
- $\varnothing\{\varepsilon\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$?

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\emptyset = \{w_1 w_2 \mid w_1 \in L, w_2 \in \emptyset\} = \emptyset$
- $\emptyset L = \emptyset$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\} L = L$
- $\emptyset\{\varepsilon\} = \emptyset$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n = \{w_1 \dots w_n \mid w_1 \in \{\varepsilon\}, \dots, w_n \in \{\varepsilon\}\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\} L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n = \{w_1 \dots w_n \mid w_1 \in \{\varepsilon\}, \dots, w_n \in \{\varepsilon\}\} = \{\varepsilon^n\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n = \{w_1 \ldots w_n \mid w_1 \in \{\varepsilon\}, \ldots, w_n \in \{\varepsilon\}\} = \{\varepsilon^n\} = \{\varepsilon\}$

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n = \{w_1 \dots w_n \mid w_1 \in \{\varepsilon\}, \dots, w_n \in \{\varepsilon\}\} = \{\varepsilon^n\} = \{\varepsilon\}$

Note the difference between:

- the empty language $\varnothing$, which contains no words
- the language $\{\varepsilon\}$, which contains exactly one word, ε

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n = \{w_1 \dots w_n \mid w_1 \in \{\varepsilon\}, \dots, w_n \in \{\varepsilon\}\} = \{\varepsilon^n\} = \{\varepsilon\}$

Note the difference between:

- the empty language $\varnothing$, which contains no words
- the language $\{\varepsilon\}$, which contains exactly one word, ε

Q: for any L_1 and L_2, does $L_1 L_2 = L_2 L_1$ hold? Can it possibly hold?

Expanding the Definition

If L is an arbitrary language, compute each of the following:

- $L\varnothing = \{w_1 w_2 \mid w_1 \in L, w_2 \in \varnothing\} = \varnothing$
- $\varnothing L = \varnothing$
- $L\{\varepsilon\} = \{w_1 w_2 \mid w_1 \in L, w_2 \in \{\varepsilon\}\} = \{w_1 w_2 \mid w_1 \in L, w_2 = \varepsilon\} = \{w_1 \varepsilon \mid w_1 \in L\} = L$
- $\{\varepsilon\}L = L$
- $\varnothing\{\varepsilon\} = \varnothing$
- Is LL equal to $\{w^2 \mid w \in L\}$? No! The latter only has words of the shape ww.
- $\{\varepsilon\}^n = \{w_1 \dots w_n \mid w_1 \in \{\varepsilon\}, \dots, w_n \in \{\varepsilon\}\} = \{\varepsilon^n\} = \{\varepsilon\}$

Note the difference between:

- the empty language $\varnothing$, which contains no words
- the language $\{\varepsilon\}$, which contains exactly one word, ε

Q: for any L_1 and L_2, does $L_1 L_2 = L_2 L_1$ hold? Can it possibly hold?
Answer: no (ex: $L_1 = \{\text{a}\}$, $L_2 = \{\text{b}\}$) and yes (ex: $L_1, L_2 \in \{\varnothing, \{\varepsilon\}\}$, $L_1 = L_2$, ...).

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element?

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
  $L_1 \cdot (L_2 \cdot L_3) = \{w_1(w_2 w_3) \mid w_1 \in L_1 \wedge (w_2 \in L_2 \wedge w_3 \in L_3)\}$

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
  $L_1 \cdot (L_2 \cdot L_3) = \{w_1(w_2 w_3) \mid w_1 \in L_1 \land (w_2 \in L_2 \land w_3 \in L_3)\}$
  $(L_1 \cdot L_2) \cdot L_3 = \{(w_1 w_2)w_3 \mid (w_1 \in L_1 \land w_2) \in L_2 \land w_3 \in L_3\}$

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
  $L_1 \cdot (L_2 \cdot L_3) = \{w_1(w_2 w_3) \mid w_1 \in L_1 \wedge (w_2 \in L_2 \wedge w_3 \in L_3)\}$
  $(L_1 \cdot L_2) \cdot L_3 = \{(w_1 w_2) w_3 \mid (w_1 \in L_1 \wedge w_2) \in L_2 \wedge w_3 \in L_3\}$

It is a monoid.

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
  $L_1 \cdot (L_2 \cdot L_3) = \{w_1(w_2 w_3) \mid w_1 \in L_1 \wedge (w_2 \in L_2 \wedge w_3 \in L_3)\}$
  $(L_1 \cdot L_2) \cdot L_3 = \{(w_1 w_2) w_3 \mid (w_1 \in L_1 \wedge w_2) \in L_2 \wedge w_3 \in L_3\}$

It is a monoid.

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
  $L_1 \cdot (L_2 \cdot L_3) = \{w_1(w_2 w_3) \mid w_1 \in L_1 \wedge (w_2 \in L_2 \wedge w_3 \in L_3)\}$
  $(L_1 \cdot L_2) \cdot L_3 = \{(w_1 w_2) w_3 \mid (w_1 \in L_1 \wedge w_2) \in L_2 \wedge w_3 \in L_3\}$

It is a monoid.

Does the cancellation law hold? (Reminder: it means $x \star y = z \star y \Longrightarrow x = z$)

Concatenation of Languages

Let A be alphabet. Consider the set of all languages $L \subseteq A^*$

Is this a monoid?

- Is there a neutral element? $L\{\varepsilon\} = L$, $\{\varepsilon\}L = L$, so $\{\varepsilon\}$ is one
- Which law needs to hold? Does it hold?
  Associativity: we need $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3$
  $L_1 \cdot (L_2 \cdot L_3) = \{w_1(w_2 w_3) \mid w_1 \in L_1 \wedge (w_2 \in L_2 \wedge w_3 \in L_3)\}$
  $(L_1 \cdot L_2) \cdot L_3 = \{(w_1 w_2)w_3 \mid (w_1 \in L_1 \wedge w_2) \in L_2 \wedge w_3 \in L_3\}$

It is a monoid.

Does the cancellation law hold? (Reminder: it means $x \star y = z \star y \Longrightarrow x = z$)
No, because there is an *absorbing* element, $\varnothing$.
For all L_1, L_2 we have $L_1\varnothing = \varnothing = L_2\varnothing$, but not necessarily $L_1 = L_2$

Representing Languages Through Programs

In general not possible: some formal languages are not *recursively enumerable* sets.

Reasonably powerful representation: computable characteristic functions.
A language $L \subseteq A^*$ is given by its *characteristic function* $f_L : A^* \longrightarrow \{0,1\}$
defined by $f_L(w) = 1$ for $w \in L$ and $f_L(w) = 0$ for $w \notin L$.

Representing Languages Through Programs

In general not possible: some formal languages are not *recursively enumerable* sets.

Reasonably powerful representation: computable characteristic functions.
A language $L \subseteq A^*$ is given by its *characteristic function* $f_L : A^* \to \{0,1\}$
defined by $f_L(w) = 1$ for $w \in L$ and $f_L(w) = 0$ for $w \notin L$.

```
case class Lang[A](contains: List[A] ⇒ Boolean)
```

Representing Languages Through Programs

In general not possible: some formal languages are not *recursively enumerable* sets.

Reasonably powerful representation: computable characteristic functions.
A language $L \subseteq A^*$ is given by its *characteristic function* $f_L : A^* \to \{0,1\}$
defined by $f_L(w) = 1$ for $w \in L$ and $f_L(w) = 0$ for $w \notin L$.

```
case class Lang[A](contains: List[A] ⇒ Boolean)
```

Example: build the language $L_2 = \{(01)^n \mid n \geq 1\}$.

```
def f(w: List[Int]): Boolean = w match
  case Cons(0, Cons(1, Nil())) ⇒ true
  case Cons(0, Cons(1, wRest)) ⇒ f(wRest)
  case _ ⇒ false
val L2 = Lang(f)
```

```
L2.contains(0::1::0::1::Nil()) // true
```

Representing Language Concatenation

We can use code to express concatenation of computable languages.

```
def concat(L1: Lang[A], L2: Lang[A]): Lang[A] =
   def f(w: List[A]) =
      val n = w.length
      def checkFrom(i: Int) =
         require(0 <= i && i <= n)
         (L1.contains(w.take(i)) && L2.contains(w.drop(i))) ||
         (i < n && checkFrom(i + 1))
      checkFrom(0, w.length)
   Lang(f) // return the language with characteristic function f

// take and drop are defined on lists:
(a :: b :: c :: d :: e :: Nil()).take(2) == a :: b :: Nil()
(a :: b :: c :: d :: e :: Nil()).drop(2) == c :: d :: e :: Nil()
```

Repetition of a Language: Kleene Star

Definition (Kleene star)

Given $L \subseteq A^*$, define

$$L^* = \bigcup_{n \geq 0} L^n$$

Repetition of a Language: Kleene Star

Definition (Kleene star)

Given $L \subseteq A^*$, define

$$L^* = \bigcup_{n \geq 0} L^n$$

Theorem

For $L \subseteq A^*$, for every $w \in A^*$ we have $w \in L^*$ if and only if

$$\exists n \geq 0.\ \exists w_1, \ldots, w_n \in L.\ w = w_1 \ldots w_n$$

Repetition of a Language: Kleene Star

Definition (Kleene star)

Given $L \subseteq A^*$, define

$$L^* = \bigcup_{n \geq 0} L^n$$

Theorem

For $L \subseteq A^*$, for every $w \in A^*$ we have $w \in L^*$ if and only if

$$\exists n \geq 0.\ \exists w_1, \ldots, w_n \in L.\ w = w_1 \ldots w_n$$

$\{a\}^* = \{\varepsilon, a, aa, aaa, \ldots\}$

$\{a, bb\}^* = \{\varepsilon, a, bb, abb, bba, aa, bbbb, aabb, \ldots\}$ (describe this language)

Repetition of a Language: Kleene Star

Definition (Kleene star)

Given $L \subseteq A^*$, define

$$L^* = \bigcup_{n \geq 0} L^n$$

Theorem

For $L \subseteq A^*$, for every $w \in A^*$ we have $w \in L^*$ if and only if

$$\exists n \geq 0.\ \exists w_1, \ldots, w_n \in L.\ w = w_1 \ldots w_n$$

$\{a\}^* = \{\varepsilon, a, aa, aaa, \ldots\}$

$\{a, bb\}^* = \{\varepsilon, a, bb, abb, bba, aa, bbbb, aabb, \ldots\}$ (describe this language)

- words whose all contiguous blocks of b-s have even length

Can L^* be finite for some L? If so, describe all such L

Repetition of a Language: Kleene Star

Definition (Kleene star)

Given $L \subseteq A^*$, define

$$L^* = \bigcup_{n \geq 0} L^n$$

Theorem

For $L \subseteq A^*$, for every $w \in A^*$ we have $w \in L^*$ if and only if

$$\exists n \geq 0.\ \exists w_1, \ldots, w_n \in L.\ w = w_1 \ldots w_n$$

$\{a\}^* = \{\varepsilon, a, aa, aaa, \ldots\}$

$\{a, bb\}^* = \{\varepsilon, a, bb, abb, bba, aa, bbbb, aabb, \ldots\}$ (describe this language)

- words whose all contiguous blocks of b-s have even length

Can L^* be finite for some L? If so, describe all such L

- $\{\varepsilon\}^* = \{\varepsilon\}$, $\varnothing^* = \{\varepsilon\}$; all others have at least one word of length ≥ 1, so L^* is infinite

Star and the Empty Word

Concatenating with an empty word has no effect, so we have the following:

$$L^* = (L \setminus \{\varepsilon\})^* = \{\varepsilon\} \cup \bigcup_{n \geq 1} (L \setminus \{\varepsilon\})^n$$

Moreover, $w \in L^*$ if and only if either $w = \varepsilon$ or, for some n where $1 \leq n \leq |w|$,

$$w = w_1 \dots w_n$$

where $w_i \in L$ and $|w_i| \geq 1$ for all i where $1 \leq i \leq n$.

Star and the Empty Word

Concatenating with an empty word has no effect, so we have the following:

$$L^* = (L \setminus \{\varepsilon\})^* = \{\varepsilon\} \cup \bigcup_{n \geq 1} (L \setminus \{\varepsilon\})^n$$

Moreover, $w \in L^*$ if and only if either $w = \varepsilon$ or, for some n where $1 \leq n \leq |w|$,

$$w = w_1 \dots w_n$$

where $w_i \in L$ and $|w_i| \geq 1$ for all i where $1 \leq i \leq n$.

- we omit ε because it leaves concatenation the same
- we can assume $n \leq |w|$ because all blocks have length at least one

Star and the Empty Word

Concatenating with an empty word has no effect, so we have the following:

$$L^* = (L \setminus \{\varepsilon\})^* = \{\varepsilon\} \cup \bigcup_{n \geq 1} (L \setminus \{\varepsilon\})^n$$

Moreover, $w \in L^*$ if and only if either $w = \varepsilon$ or, for some n where $1 \leq n \leq |w|$,

$$w = w_1 \dots w_n$$

where $w_i \in L$ and $|w_i| \geq 1$ for all i where $1 \leq i \leq n$.

- we omit ε because it leaves concatenation the same
- we can assume $n \leq |w|$ because all blocks have length at least one

If L is computable (has a computable characterstic function), is L^* also computable?

Star and the Empty Word

Concatenating with an empty word has no effect, so we have the following:

$$L^* = (L \setminus \{\varepsilon\})^* = \{\varepsilon\} \cup \bigcup_{n \geq 1} (L \setminus \{\varepsilon\})^n$$

Moreover, $w \in L^*$ if and only if either $w = \varepsilon$ or, for some n where $1 \leq n \leq |w|$,

$$w = w_1 \dots w_n$$

where $w_i \in L$ and $|w_i| \geq 1$ for all i where $1 \leq i \leq n$.

- we omit ε because it leaves concatenation the same
- we can assume $n \leq |w|$ because all blocks have length at least one

If L is computable (has a computable characterstic function), is L^* also computable?

- try all possible ways of splitting w
- if $k = |w|$, for each point between the letters of w you can decide to split there or not, so there are 2^{k-1} ways to split: $w = \Box \underbrace{|\Box| \dots |\Box|}_{k-1} \Box$
- there is a better way - see exercises

Starring: $\{a, ab\}$

Let $A = \{a, b\}$ and $L = \{a, ab\}$.
Come up with a property "…" that describes the language L^*:

$$L^* = \{w \in A^* \mid \ldots\}$$

Prove that the property and L^* denote the same language.

Starring: $\{a, ab\}$

Let $A = \{a, b\}$ and $L = \{a, ab\}$.
Come up with a property "..." that describes the language L^*:

$$L^* = \{w \in A^* \mid \ldots\}$$

Prove that the property and L^* denote the same language.

Properties:

- does not being with b
- does not contain bb

in one phrase: there is an "a" before every "b"