1

ภาษา

LANGUAGE

LANGUAGE

ภาษาอังกฤษ

หน่วยของภาษา

- อักขระ letters
- คำ words
- ประโยค sentences
- ย่อหน้า paragraphs
- เรื่องราว coherent stories

COLLECTION
&
SEQUENTIAL

How do they do that ?

LANGUAGE

ภาษาคอมพิวเตอร์

หน่วยของภาษาคอมพิวเตอร์

- อักขระ letters
- คำสำคัญ keywords
- คำสั่ง commands
- โปรแกรม programs
- ระบบ systems

It is very hard to state all the rules for the language "spoken English".

Commands can be recognized by certain sequences of words.
Language structure is based on explicitly rules.

LANGUAGE

ภาษา

LANGUAGE

Definition

Language means simply a set of strings involving symbols from alphabet.

LANGUAGE

ทฤษฎีภาษารูปนัย

THEORY OF FORMAL LANGUAGES

Formal refers

- explicitly rules
  - What sequences of symbols can occur ?
  - No liberties are tolerated.
  - No reference to any "deeper understanding" is required.
- the form of the sequences of symbols
- not the meaning

LANGUAGE

นิยามและสัญลักษณ์

- One finite set of fundamental units , called "alphabet", denoted Σ.
- An element of alphabet is called "character".
- A certain specified set of strings of characters will be [illegible]ge" denoted L.
- Those s[illegible] are permissible in the language we call "words".
- The string without letter is called "empty string" or "null string", denoted by Λ.
- The language that has no word is denoted by $\varnothing$.

specified

LANGUAGE

SYMBOLS

Union operation	+
Different operation	–
Alphabet	Σ
Empty string	Λ
	ε
Language	L
	Γ
Empty language	$\varnothing$

LANGUAGE

ภาษา

การนิยามหรือการบรรยายภาษา

IMPLICITLY DEFINING

Given an alphabet Σ = { a b c … z ʻ - }.
We can now specify a language L as
{ all words in a standard dictionary },
named "ENGLISH-WORDS".

LANGUAGE

ภาษา

การนิยามหรือการบรรยายภาษา

INFINITE LANGUAGE DEFINING

The trick of defining the language L,

By listing all rules of grammar.

This allows us to give a finite description of an infinite language.

Consider this sentence "I eat three Sundays".

This is grammatically correct.

RIDICULOUS LANGUAGE

LANGUAGE

ภาษา

การนิยามหรือการบรรยายภาษา

Method of exhaustion

Let $\Sigma = \{x\}$ be an alphabet.

Language L can be defined by

$L = \{ x \ \ xx \ \ xxx \ \ xxxx \ \ \ldots \}$

$L = \{ x^n \text{ for } n = 1 \ \ 2 \ \ 3 \ \ \ldots \}$.

Language

$L_2 = \{ x \ \ xxx \ \ xxxxx \ \ xxxxxxx \ \ \ldots \}$

$L_2 = \{ x^{odd} \}$

$L_2 = \{ x^{2n-1} \text{ for } n = 1 \ \ 2 \ \ 3 \ \ \ldots \}$.

LANGUAGE

ภาษา นิยาม

We define the function length of a string to be the number of letters in the string.

For example, if a word *a* = xxxx in L, then length(*a*)=4.

In any language that includes Λ, we have length(Λ)=0.

The function reverse is defined by if *a* is a word in L, then reverse(*a*) is the same string of letters spelled backward, called the reverse of *a*.

For example, reverse(123)=321.

Remark: The reverse(*a*) is not necessary in the language of *a*.

LANGUAGE

ภาษา นิยาม

We define the function $n_a(w)$ of a *w* to be the number of letter *a* in the string *w*.

For example, if a word *w* = aabbac in L, then $n_a(w)=3$.

Concatenation of two strings means that two strings are written down side by side.

For example, x^n concatenated with x^m is x^{n+m}

LANGUAGE

ภาษา นิยาม

Language is called PALINDROME over the alphabet if

Language = { Λ and all strings x such that reserve(x)=x }.

For example, let Σ={ a, b }, and

PALINDROME={ Λ a b aa bb aaa aba bab bbb …}.

Remark: Sometimes, we obtain another word in PALINDROME when we concatenate two words in PALINDROME. We shall see the interesting properties of this language later.

LANGUAGE

ภาษา นิยาม

Consider the language

PALINDROME={ Λ a b aa bb aaa aba bab bbb …}.

We usually put words in size order and then listed all the words of the same length alphabetically. This order is called lexicographic order.

LANGUAGE

ภาษา นิยาม

KLEENE CLOSURE

Given an alphabet ∑, the language that any string of characters in ∑ are in this language is called the closure of the alphabet. It is denoted by

$$\Sigma^*.$$

This notation is sometimes known as the Kleene star.

Kleene star can be considered as an operation that makes an infinite language. When we say "infinite language", we mean infinitely many words, each of finite length.

LANGUAGE

ภาษา นิยาม

KLEENE CLOSURE

More general,

if S is a set of words, then by S* we mean the set of all finite strings formed by concatenating words from S and from S*.

Example:

If S = { a ab }then

S* = { Λ and any word composed of factors of a and ab }.

{ Λ and all strings of a and b except strings with double b }.

{ Λ a aa ab aaa aab aba aaaa aaab aaba … }.

LANGUAGE

ภาษา นิยาม

KLEENE CLOSURE

Example:

If S = { a ab }then

S* = { Λ and any word composed of factors of a and ab }.

{ Λ and all strings of a and b except strings with double b }.

{ Λ a aa ab aaa aab aba aaaa aaab aaba … }.

To prove that a certain word is in the closure language S* , we must show how it can be written as a concatenation of words from the base set S.

Example: abaaba can be factored as (ab)(a)(ab)(a) and it is unique.

LANGUAGE

ภาษา นิยาม

KLEENE CLOSURE

Example:

If S = { xx xxxxx }then

S* ={ Λ xx xxxx xxxxx xxxxxx xxxxxxx xxxxxxxx … }.

{ Λ and xx and x^n for n = 4 5 6 7 … }.

How can we prove this statement ?

Hence: proof by constructive algorithm (showing how to create it).

LANGUAGE

ภาษา นิยาม

KLEENE CLOSURE

Example:

If S = { a b ab } and T = { a b ba }, then S* = T* = { a b }*.

Proof: It is clear that { a b }*⊂ S* and { a b }*⊂T*.

We have to show that S* and T* ⊂ { a b }*.

For x ∈ S*, in the case that x is composed of ab.

Replace ab in x by a, b which are in { a b }*.

Then S*⊂ { a b }*.

The proof of T*⊂ { a b }* is similarity. QED

LANGUAGE

ภาษา นิยาม

POSITIVE CLOSURE

Given an alphabet ∑, the language that any string (not zero) of characters in ∑ are in this language is called the positive closure of the alphabet. It is denoted by

$$\Sigma^+.$$

Example: Let Γ={ ab }.

Then Γ^+ = { ab abab ababab … }.

LANGUAGE

ภาษา นิยาม

TRIVIAL REMARK

Given an alphabet Σ={ aa bbb }. Then Σ* is the set of all strings where a's occur in even clumps and b's in groups of 3, 6, 9…. Some words in Σ* are

bbb aabbbaaaa bbbaa

If we concatenate these three elements of Σ*, we get one big word in Σ**, which is again in Σ*.

bbbaabbbaaaabbbaa = (bbb)(aa)(bbb)(aa)(aa)(bbb)(aa)

Note : Σ** means (Σ*)*.

LANGUAGE

ภาษา นิยาม

THEOREM

Theorem

For any set S of strings, we have S*= S**.

Proof: Every words in S** is made up of factors from S*.

Every words in S* is made up of factors from S.

Therefore every words in S** is made up of factors from S.

We can write this $S^{**} \subset S^*$.

In general, it is true that $S \subset S^*$. So $S^* \subset S^{**}$.

Then S*= S**. QED

โจทย์ น่าคิด

ให้ L เป็นภาษาที่นิยามบน $\sum=\{0,1\}$ จงอธิบายความสัมพันธ์ของ

- L^{*+}
- L^{+*}

?