Session 2

Languages & Language Operations

Definition of Languages

- A language over Σ is a subset of Σ^*
- L is a language over Σ if $L \subseteq \Sigma^*$
- How many languages are there for a given Σ?

How many languages are there?

- Let $m_0 \ldots m_n$ be the list of finite strings formed with symbols of Σ
- Let $S_0 \ldots S_n$ be the list of all subsets (languages) of Σ^*

	m_0	m_1	m_2	…	m_n
S_0	0	0	0	…	0
S_1	0	0	1	…	0
S_2	1	0	1	…	0
…	…	…	…	…	…
S_n	1	1	1	…	1

A diagonal set: $D(i) = S_i(i)$

	m_0	m_1	m_2	…	m_n
S_0	0	0	0	…	0
S_1	0	0	1	…	0
S_2	1	0	1	…	0
…	…	…	…	…	…
S_n	1	1	1	…	1

$$D = \{m_2, \ldots, m_n\}$$

The *anti*-diagonal set

	m_0	m_1	m_2	…	m_n
S_0	1	0	0	…	0
S_1	0	1	1	…	0
S_2	1	0	0	…	0
…	…	…	…	…	…
S_n	1	1	1	…	0

$$\overline{D} = \{m_0, m_1, \ldots\}$$

Is the *anti-diagonal* set in the list?

- $\overline{D}$ is not in the list $S_0 \ldots S_n$
  - Differs from S_0 in the first element
  - Differs from S_1 in the second element
  - …
  - Differs from S_n in the $n + 1$ element
- 2^{Σ^*} is not countable!
- There are more languages that we can think of!

Set operations on languages

- If $L_1, L_2 \subseteq \Sigma^*$
  - Union: $L_1 \cup L_2$ is a Language
  - Intersection: $L_1 \cap L_2$ is a Language
  - Difference: $L_1 - L_2$ is a Language
- If $L_1 \subseteq \Sigma_1^*$ and $L_2 \subseteq \Sigma_2^*$ then L_1, L_2 are both subsets of $(\Sigma_1 \cup \Sigma_2)^*$
  - $\overline{L_1} = \Sigma_1^* - L_1$
  - $\overline{L_1} = (\Sigma_1 \cup \Sigma_2)^* - L_1$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

- Concatenation of languages
  - If $L_1, L_2 \subseteq \Sigma^*$
  - The concatenation of L_1 and L_2, L_1L_2 is

$$L_1L_2 = \{xy \mid x \in L_1 \text{ and } y \in L_2\}$$

- Example:
  - L_1 = {hope, fear}
  - L_2 = {less, fully}
  - L_1L_2 = {hopeless, hopefully, fearless, fearfully}
- Concatenation with $\{\Lambda\}$:
  - $\{\Lambda\}L = L\{\Lambda\} = L$
  - $\{\Lambda\}$ is the unit of concatenation for languages

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Concatenation of Languages

fully	hopefully	fearfully
less	hopeless	fearless
L_2 \ L_1	hope	fear

$$L_1L_2 = \{xy \mid x \in L_1 \text{ and } y \in L_2\}$$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Closures of languages

- The Kleen-star operation L^*:

$$L^* = \bigcup_{i=0}^{\infty} L^i$$

- The Kleen-plus operation L^+:

$$L^+ = \bigcup_{i=1}^{\infty} L^i$$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Closures of languages

- The Kleen-star operation L^*:
  - $L = \{0, 1\}$ $(\Sigma = \{0, 1\}$ but $\Sigma \neq L)$
  - $L^* = L^0 \cup L^1 \cup L^2 \cup \ldots$
  - $L^* = \{0, 1\}^* = \{\Lambda, 0, 1, 00, 01, 10, 11, \ldots\}$
- The Kleen-plus operation L^+:
  - $L = \{0, 1\}$
  - $L^+ = L^1 \cup L^2 \cup \ldots$
  - $L^+ = \{0, 1\}^+ = \{0, 1, 00, 01, 10, 11, \ldots\}$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Closures of Languages

- The empty language: Φ
- The language containing only the empty string: $\{\Lambda\}$
- A finite language without Λ
- A finite language with Λ
- An infinite (denumerable) language with Λ
- An infinite (denumerable) language without Λ

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Notation of Powers

- If $a \in \Sigma$, $x \in \Sigma^*$ and $L \subseteq \Sigma^*$
  - $a^k = aa \ldots a$ k times
  - $x^k = xx \ldots x$ k times
  - $\Sigma^k = \Sigma\Sigma \ldots \Sigma$ k times
  - $L^k = LL \ldots L$ k times
- The case for $k = 0$
  - $a^0 = \Lambda$, $x^0 = \Lambda$ } Unit of concatenation for strings
  - $\Sigma^0 = \{\Lambda\}$, $L^0 = \{\Lambda\}$ } Unit of concatenation for languages

Unit of concatenation for Languages

L_2 \ L_1	hope	fear
Λ	hopeΛ	fearΛ

⇒

L_2 \ L_1	hope	fear
Λ	hope	fear

Closures of Φ

- Expansion of powers
  - $\Phi^0 = \{\Lambda\}$
  - $\Phi^1 = \Phi$ (it has no elements)
  - $\Phi^2 = \Phi\ \Phi = \Phi$
  - $\Phi^3 = \Phi\ \Phi^2 = \Phi\ \Phi = \Phi$
  - …
- Closures
  - $\Phi^* = \Phi^0 = \{\Lambda\}$
  - $\Phi^+ = \Phi^1 = \Phi$ (has no elements)

Closures of $\{\Lambda\}$

- Expansion of powers
  - $L^0 = \{\Lambda\}^0 = \{\Lambda\}$
  - $L^1 = L = \{\Lambda\}$
  - $L^2 = L^1L^1 = \{\Lambda\}\{\Lambda\} = \{\Lambda\Lambda\} = \{\Lambda\}$
  - $L^3 = L^1L^2 = \{\Lambda\}\{\Lambda\} = \{\Lambda\Lambda\} = \{\Lambda\}$
  - …
- Closures
  - $\{\Lambda\}^* = \{\Lambda\}$
  - $\{\Lambda\}^+ = \{\Lambda\}$
  - If $L = \{\Lambda\}$ then $L = L^* = L^+$

Closures of a finite language

- *Let* $L = \{0, 1\}$ $(\Sigma = \{0, 1\}$ but $\Sigma \neq L)$
  - $L^0 = \{\Lambda\}$
  - $L^1 = L = \{0, 1\}$
  - $L^2 = L^1L^1 = \{0,1\}\{0,1\} = \{00,01,10,11\}$
  - $L^3 = L^1L^2 = \{0,1\}\ \{00,01,10,11\} =$ $\{000, 001,010,011,100,101,110,111\}$
- Closures:
  - $L^* = \{\Lambda,0,1,00,01,10,11,000,001,010,011,\ldots\}$
  - $L^+ = \{0,1,00,01,10,11,000,001,010,011,\ldots\}$

Closures of a finite language

- *Let* $L = \{0, 11\}$ $(\Sigma = \{0, 1\}$ but $\Sigma \neq L)$
  - $L^0 = \{\Lambda\}$
  - $L^1 = L = \{0, 11\}$
  - $L^2 = L^1L^1 = \{0, 11\}\{0, 11\} = \{00, 011, 110, 1111\}$
  - $L^3 = L^1L^2 = \{0, 11\}\{00, 011, 110, 1111\}$ $=\{000,0011,0110,01111,1100,11011,11110,111111\}$
- L^n is the set of strings that result of concatenating n *strings* or *words* of L (not of symbols of Σ)
- Closures:
  - $L^* = \{\Lambda,0,11,00,011,110,1111,000,0011,0110,01111\ldots\}$
  - $L^+ = \{0,11,00,011,110,1111,000,0011,0110,01111\ldots\}$

Graphical computation L^{n+1}

L^2

1111	01111	111111
110	0110	11110
011	0011	11011
00	000	1100
	0	11

L

$L^3 = L\,L^2$

Concatenation of a language with its powers is commutative!

L

11	0011	01111	11011	111111
0	000	0110	1100	11110
	00	011	110	1111

L^2

$L^2 L = L\,L^2$

Any language commutes with its own closure

- $L^+ = L\,L^* = L^*L$
  - $L^* = L^0 + L^1 + L^2 + \ldots$
  - $L^+ = L^1 + L^2 + \ldots$
- $L\,L^* = L(L^0 + L^1 + L^2 + \ldots)$
  $= L(\{\Lambda\} + L^1 + L^2 + \ldots)$
  $= L\Lambda + LL^1 + LL^2 + \ldots$
  $= L + L^2 + L^3 + \ldots$
  $= L^+$
- $L^*L = (L^0 + L^1 + L^2 + \ldots)\,L$
  $= (\{\Lambda\} + L^1 + L^2 + \ldots)\,L$
  $= \Lambda L + L^1L + L^2L + \ldots$
  $= L + L^2 + L^3 + \ldots$
  $= L^+$
- $L\Lambda + LL^1 + LL^2 + \ldots = \Lambda L + L^1L + L^2L + \ldots$

Closures of a finite language with Λ

- *Let* $L = \{\Lambda, 0, 11\}$ $(\Sigma = \{0, 1\})$
  - $L^0 = \{\Lambda\}$
  - $L^1 = L = \{\Lambda, 0, 11\}$ (i.e. L^1 contains L^0)
  - $L^2 = L^1L^1 = \{\Lambda, 0, 11\}\{\Lambda, 0, 11\}$
    $= \{\Lambda\Lambda, \Lambda 0, \Lambda 11, 0\Lambda, 00, 011, 11\Lambda, 110, 1111\}$
    $= \{\Lambda, 0, 11, 0, 00, 011, 11, 110, 1111\}$
    $= \{\Lambda, 0, 11, 00, 011, 110, 1111\}$ (i.e. L^2 contains L!)
  - $L^2 = L \cup \{0, 11\}\{0, 11\} = L \cup (L^1 - L^0)(L^1 - L^0)$
  - $L^3 = L^2 \cup (L^1 - L^0)(L^2 - L^1) = (L^1 - L^0)L^2$
- Closures:
  - $L^* = \{\Lambda, 0, 11, 00, 011, 110, 1111, \ldots\}$
  - $L^+ = \{\Lambda, 0, 11, 00, 011, 110, 1111, \ldots\}$
  - $L^* = L^+$

Computing L^3

L^2

1111	Λ1111	01111	111111
110	Λ110	0110	11110
011	Λ011	0011	11011
00	Λ00	000	1100
11	Λ11	011	1111
0	Λ0	00	110
Λ	ΛΛ	0Λ	11Λ
	Λ	0	11

L

Computing L^3: applying id.

Λ is a language reproducer!

L^2 L^2

1111	1111	01111	111111
110	110	0110	11110
011	011	0011	11011
00	00	000	1100
11	11	011	1111
0	0	00	110
Λ	Λ	0	11
	Λ	0	11

L^1

L

The contribution of L^3: $(L^1 - L^0)(L^2 - L^0)$

L^2	L^2			
1111	1111	01111	111111	
110	110	0110	11110	
011	011	0011	11011	
00	00	000	1100	
11	11	011	1111	
0	0	00	110	
Λ	Λ	0	11	L^1
	Λ	0	11	L

$L^3 = L^0 \cup L^1 \cup L^2 \cup (L^1 - L^0)(L^2 - L^0)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The contribution of L^3: $c(L_1)c(L_2)$

L^2	L^2			
1111	1111	01111	111111	
110	110	0110	11110	
011	011	0011	11011	
00	00	000	1100	
11	11	011	1111	$L^2 - L^1$
0	0	00	110	
Λ	Λ	0	11	L^1
	Λ	0	11	L

$L^3 = L^0 \cup L^1 \cup L^2 \cup (L^1 - L^0)(L^2 - L^1)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

If $\Lambda \in L$

$$L^n = \bigcup_{i=0}^{n-1} L^i \bigcup (L^1 - L^0)(L^{n-1} - L^{n-2})$$

Is this true?

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Closures of an infinite language

- *Let* $L = \{0, 00, 000, \ldots\}$ $(\Sigma = \{0\})$
  - $L^0 = \{\Lambda\}$
  - $L^1 = L = \{0, 00, 000,\ldots\}$
  - $L^2 = L^1L^1 = \{0, 00, 000,\ldots\}\{0, 00, 000,\ldots\}$
    $= \{00, 000, 0000,\ldots, 000, 0000, 00000,\ldots\}$
    $= \{00, 000, 0000, 00000,\ldots\} = L^1 - \{0\}$
- Closures:
  - $L^* = \{\Lambda, 0, 00, 000, \ldots\}$
  - $L^+ = \{0, 00, 000, \ldots\}$
  - $L^+ = L$

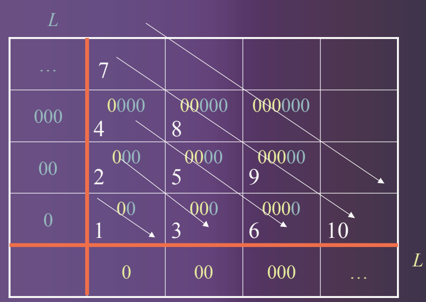
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Enumerating the concatenating of infinite languages

L					
…					
000	0000	00000	000000		
00	000	0000	00000		
0	00	000	0000		
	0	00	000	…	L

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

$L^2 = \{00,000,000,0000,0000,0000,\ldots\}$



Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Closures of an infinite language with Λ

- *Let* $L = \{\Lambda, 0, 00, 000, \ldots\}$ $(\Sigma = \{0\})$
  - $L^0 = \{\Lambda\}$
  - $L^1 = L = \{\Lambda, 0, 00, 000, \ldots\}$
  - $L^2 = L^1L^1 = \{\Lambda, 0, 00, 000, \ldots\}\{\Lambda, 0, 00, 000, \ldots\}$
    $= \{\Lambda\Lambda, \Lambda 0, \Lambda 00, \Lambda 000 \ldots, 0\Lambda, 00, 000, 0000, \ldots,$
    $000, 0000, 00000, \ldots\}$
    $= \{\Lambda, 0, 00, 000, 0000, 00000 \ldots\}$
  - $L^2 = L^1$
- Closures:
  - $L^* = \{\Lambda, 0, 00, 000, \ldots\}$
  - $L^+ = \{\Lambda, 0, 00, 000, \ldots\}$
  - $L^* = L^+ = L$ and, indeed, $L^{i+1} = L^i$ for all $i \geq 0$

Closure again!

- Any string formed by 0 or more repetitions of the elements of L is in L^*
  - The first element of the string belongs to L
  - The first *two* elements of the string are in L^2, even if they are the same!
  - The first *three* elements of the strings are in L^3, because they are produced by the concatenation of L^2 con L, even if they are the same!
  - …

Closures

- The closure of $L = \{0, 1\}$ is the set of all strings formed by 0 and 1, and Λ
- 100111 belongs to the closure:
  - 100111 belongs to L
  - 100111 belongs to L^2
  - 100111 belongs to L^3
  - 100111 belongs to L^4
  - 100111 belongs to L^5
  - 100111 belongs to L^6

Closures

- The closure of $L = \{10, 111\}$ is the set of all strings formed by the pairs 10 and 111, including repetitions, and Λ:
- 10111111 belongs to the closure:
  - 10111111 belongs to L
  - 10111111 belongs to L^2
  - 10111111 belongs to L^3
- But 110111does not!
  - 101111 belongs to L^1
  - 101111 belongs to L^2
  - 101111 but belongs to no power of L!

Closures

- The closure of $L = \{11, 111, 11111, 1111111\ldots\}$ is the set of strings formed by a sequence of prime numbers in monadic notation, and Λ!
- 1111111111 belongs to the closure:
  - 1111111111 belongs to L
  - 1111111111 belongs to L^2
  - 1111111111 belongs to L^3
- Is there a number $n > 1$ which is not in the closure of L?

Closure

- It is perhaps hard to think of the closure of a language!
- But it is much simpler to see whether a given string is in the closure of a language:
  - Read the string from left to right, and scan members of the language one at a time
  - If the string is completely scanned, and we have successfully extracted members of the language at every stage of the process, the string is in the closure of the language!
- It is simple to generate string in the closure: just concatenate symbols of L, with repetitions allowed!

Example 1

- Give an example of INFINITE Languages L_1, L_2 over $\{a, b\}$ where L_1 ⌦ L_1L_2 and L_2 ⌦ L_1L_2. Justify your answer by giving strings in L_1 and L_2 which are not in L_1L_2
- Assume:
  - $L_1 = \{a, aa, aaa, \dots\}$
  - $L_2 = \{b, bb, bbb, \dots\}$
  - $L_1L_2 = \{ab, abb, aab, abbb, aabb, aaabb \dots\}$
- Consequently:
  - L_1 ⌦ L_1L_2 and L_2 ⌦ L_1L_2
  - In particular, $a^k \in L_1$ but $a^k \notin L_1L_2$
  - Similarly, $b^k \in L_2$ but $b^k \notin L_1L_2$

Example 2

- Give an example of languages L_1, L_2 and L_3 where
  $(L_1 \cap L_2)L_3 \neq (L_1L_3) \cap (L_2L_3)$
- Let $L_1 = \{\Lambda\}$, $L_2 = \{a, b\}^+ = \Sigma^+$ and $L_3 = \{a, b\}^* = \Sigma^*$
- Right side:

$$(L_1 \cap L_2)L_3 = (\{\Lambda\} \cap \Sigma^+)\Sigma^* = \Phi\Sigma^* = \Phi$$

- Left side:

$$(L_1L_3) \cap (L_2L_3) = (\{\Lambda\}\Sigma^*) \cap (\Sigma^+\Sigma^*) = \Sigma^* \cap \Sigma^+ = \Sigma^+$$

- So, $\Phi \neq \Sigma^+$ and $(L_1 \cap L_2)L_3 \neq (L_1L_3) \cap (L_2L_3)$

Summary

- Given an alphabet Σ
- There is universe of strings Σ^*
- A language L is a subset of Σ^*
- There are many subsets of Σ^*: 2^{Σ^*}
- There are quite a few languages (more than we can think of) for any Σ
- Some of these are finite… other denumerable (countable infinite)
- Given an initial language or set of languages, we can define new languages through language operations (e.g. union, concatenation and closures)

Summary

- A Problem: given a language and a string, tell whether the strings belongs to the language:
  - Declarative specification: $s \in L$
  - Procedural specification:

(L, s) → [Decision procedure: An algorithm] → Yes or no!

- Two sides of the coin:
  - Generating all the strings of a language
  - Telling whether a string belongs to a language

Outline of the course