Chapter 1 Sets, Relations, and Language

Elements Of The Theory Of Computation
Zhejiang University/CS/Course/Xiaogang Jin
E-Mail: xiaogangj@cise.zju.edu.cn

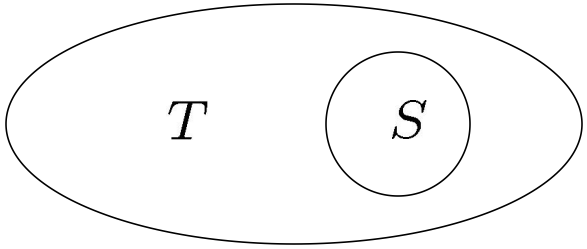
1.1 Sets

Set

– an unordered collection of elements

empty set $\emptyset$

Subsets and proper subsets



– Subset notation: $\subseteq$

$S \subseteq T \Leftrightarrow (\forall x \in S \Rightarrow x \in T)$

– Proper Subset: $\subset$

– Two sets are **equal** iff they contain the same elements

$S = T \Leftrightarrow (S \subseteq T) \wedge (T \subseteq S)$

Set Operations and Its Identities

– Union, Intersection, Difference, Symmetric difference, complement

– Commutative Law, Associative Law, Distributive law, Absorption, DeMorgan's Law, Idempotent law

Power Set

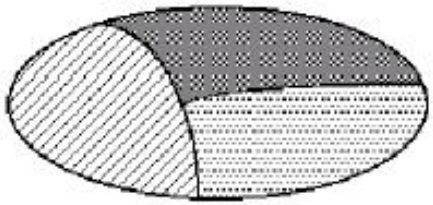
– 2^S = set of all subset of S

$$2^S = \{T \mid T \subseteq S\}$$

Partition

A **partition** of a nonempty set A is a subset Π of 2^A such that

1) $\emptyset \notin \Pi$;
2) $\forall S, T \in \Pi,$ and $S \neq T, S \cap T = \emptyset$
3) $\bigcup \Pi = A$.



1.2 Relations and Functions

Ordered Pair and Binary Relation

– Ordered Pair: (a, b)

$$(a, b) = (c, d) \Leftrightarrow (a = c) \wedge (b = d)$$

– Cartesian Product: $A \times B$

$$A \times B = \{(a, b) | a \in A \wedge b \in B\}$$

– Binary Relation on A and B:

$$R \subseteq A \times B$$

Ordered Tuple and n-ary Relation (Omitted)

Operations of Relations

- Inverse

$$R \subseteq A \times B \Rightarrow R^{-1} \subseteq B \times A$$

- Composition

When ideas fail, words come in very handy.
- Goethe (1749-1832)

Function

Definition: A **function** f: $A \rightarrow B$ must satisfy:

- $f \subseteq A \times B$
- $\forall a \in A, \exists$ exactly one $b \in B$ with $(a, b) \in f$

Note:

We write $(a, b) \in f$ as $f(a) = b$.

Domain, range

one-to-one function:

$\forall a, b \in A \wedge a \neq b \Rightarrow f(a) \neq f(b)$

onto function:

$\forall b \in B \ \exists a \in A$ such that $f(a) = b$

bijection function:

(one-to-one correspondence)

one-to-one + onto

1.3 Special Types of Binary Relations

Representation of Relations

– Directed graph: node, edge, path
– Matrix: Adjacency matrix

Properties of Relations $(R \subseteq A \times A)$

– reflexive: $\forall a \in A \Rightarrow (a, a) \in R$
– symmetric: $(a, b) \in R \wedge a \neq b \Rightarrow (b, a) \in R$
  antisymmetric: $(a, b) \in R \Rightarrow (b, a) \notin R$
– transitive: $(a, b) \in R, (b, c) \in R \Rightarrow (a, c) \in R$

Equivalence Relation

– reflexive, symmetric, transitive

– equivalence classes

$$[a] = \{b|(a,b) \in R\}$$

Theorem Let R be an equivalence relation on a nonempty set A. Then the equivalence classes of R constitute a partition of A.

Partial Order

- reflexive, antisymmetric, transitive
- total order
- minimal element and maximal element

"Sometimes when reading Goethe, I have the paralyzing suspicion that he is trying to be funny.
- Guy Davenport

1.4 Finite and Infinite sets

Equinumerous

– Sets A and B equinumerous $\Leftrightarrow \exists$ bijection $f : A \rightarrow B$

– Cardinality and generalized Cardinality

– Finite and Infinite Sets

Countable and Uncountable Infinite

– A set is said to be **countably infinite** $\Leftrightarrow$ it is equinumerous with $\mathbb{N}$.

– S is an **uncountable** set $\Leftrightarrow |S| > |\mathbb{N}|$.

– The union of a countably infinite collection of countably infinite sets is countably infinite.

Example: Show that $\mathbb{N} \times \mathbb{N}$ is countably infinite.

Theorem: $|\mathbb{R}| > |\mathbb{N}|$.

Proof: See next section by diagonalization.

Question: Is $|\mathbb{R}| > |(0, 1)|$?

$$f(x) = \frac{1}{\pi} arctan(x) + \frac{1}{2}$$

Continuum Hypothesis

$|\mathbb{N}| = \aleph_0 \qquad |\mathbb{R}| = \aleph_1$

$\aleph_0 < \aleph_1$

$\exists?\ w$ such that $\aleph_0 < w < \aleph_1$

Independent of the axioms ! [Cohen, 1966]

For an informal account on infinities, see e.g.: Rucker, *Infinity and the Mind* , Harvester Press, 1982.

1.5 Three Fundamental Proof Techniques

The Principle of Mathematical Induction

Let A be a set of natural numbers such that
1) $0 \in A$, and
2) for each natural number n, if $\{0, 1, 2, \cdots, n\} \in A$, then $n + 1 \in A$.

The Pigeonhole Principle

If A and B are finite sets and $|A| > |B|$, then there is no one-to-one function from A to B.

Proof. *Basis step.*

$|B| = 0$ $(B = \emptyset)$ $\Rightarrow$ no function from A to B

$\Rightarrow$ no one-to-one function.

Induction Hypothesis. Suppose $f : A \rightarrow B$, $|A| > |B|$, and $|B| \leq n$, where $n \geq 0 \Rightarrow f$ is not one-to-one.

Induction step. Suppose $f : A \rightarrow B$, and $|A| > |B| = n + 1$.

Choose some $a \in A$.

Case 1: If $\exists a' \in A$, such that $f(a) = f(a')$.

$\Rightarrow f$ is not one-to-one.

Case 2: a is the only element mapped by f to $f(a)$

Consider then the sets $A-\{a\}$ to $B-\{f(a)\}$, and the function $g : A - \{a\} \rightarrow B - \{f(a)\}$ that agree with f on all elements of $A - \{a\}$.

By induction hypothesis

$$|B - \{f(a)\}| = n$$

$$|A - \{a\}| = |A| - 1 > |B| - 1 = |B - \{f(a)\}|$$

$\Rightarrow \exists a, b \in A-\{a\}$, $a \neq b$, such that $g(a) = g(b)(f(a) = f(b))$.

$\Rightarrow$ f is not one-to-one

The Diagonalization Principle

Let R be a binary relation on a set A, and let D, the diagonal set for R, be $\{a : a \in A \wedge (a, a) \notin R\}$. For each $a \in A$, let

$$R_a = \{b : b \in A \wedge (a, b) \in R\}.$$

Then D is distinct from each R_a.

Example: Let us consider the relation $R = \{(a,b),(a,d),(b,b),(b,c),(c,c),(d,b),(d,c),(d,e),(d,f),(e,e),(e,f),(f,a),(f,c),(f,d),(f,e)\}$.

Notice that $R_a = \{b,d\}$, $R_b = \{b,c\}$,
$R_c = \{c\}$, $R_d = \{b,c,e,f\}$,
$R_e = \{e,f\}$, $R_f = \{a,c,d,e\}$.

The corresponds to the diagonal set is

$$D = \{a,d,f\}.$$

Theorem: (G. Cantor1845-1918) The set $2^{\mathbb{N}}$ is uncountable.

Proof: Suppose that $2^{\mathbb{N}}$ is countably infinite.

We assume that there is a way of enumerating all members of $2^{\mathbb{N}}$ as

$$2^{\mathbb{N}} = \{R_0, R_1, R_2, \cdots\}$$

Consider the relation

$$R = \{(i, j) : j \in R_i\}.$$

R_i — in the statement of the diagonalization principle.

Now consider the set $D = \{n \in \mathbb{N} : n \notin R_n\}$

—$D \subseteq \mathbb{N}$, should be appear somewhere in the enumeration $\{R_0, R_1, R_2, \cdots\}$

— D can not be the $R_i (i = 0, 1, 2, \cdots)$

Suppose that $D = R_k$ for some $k \geq 0$. $k \in R_k$?

- $k \in R_k$: Since $D = \{n \in \mathbb{N} : n \notin R_n\} \Rightarrow k \notin D$; but $D = R_k$.
- $k \notin R_k \Rightarrow k \in D$. But $D = R_k$. **Contradiction!**

Theorem: $(0, 1)$ is uncountable.

1.6 Closures

The Transitive Closure

the "smallest" relation that includes R and is transitive (usually called R^+)

e.g. If R is Parent-of, then the transitive closure of R is Ancestor-Of

More formally:

R^+ is a relation such that

* $R \subseteq R^+$
* R^+ is transitive
* $\forall R', R \subseteq R'$ and R' is transitive, $\Rightarrow R^+ \subseteq R'$

Closures of Relations

Given any binary relation R, one can form closures with respect to any combinations of the properties:

- reflexive
- symmetric
- transitive

e.g. Symmetric, transitive closure of "Parent-Of" is:

...?...

Note:

Reflexive, transitive closure of R is usually denoted R^*.

1.7 Alphabet and Language

□ **Alphabet:** finite set of symbols

e.g. $\sum_1 = \{0, 1\}$, $\sum_2 = \{a, b, \cdots, x, y, z\}$

- String : finite symbol sequence
- Length: # of symbols
- Empty string : e

□ **Operations of Strings:**

- **Concatenation:** $x \circ y$ or xy

  Substring, suffix, prefix

Example: $\forall w, we = ew = w$

- **String exponentiation**

$w^0 = e$, the empty string
$w^{i+1} = w^i \circ w$, for each $i \geq 0$

—definition by induction

- **Reversal**

If w is a string of length 0, then $w^R = w = e$.
IF w is a string of length $n + 1 > 0$, then $w = ua$ for some $a \in \sum$, and $w^R = au^R$.

- **Language:** set of strings
  - $\sum$–alphabet, $\sum^*$–the set of all strings ($e \in \sum^*$)
  - Language $L \subseteq \sum^*$
  - $\emptyset$, $\sum$ and $\sum^*$ are languages.
  - Finite Language: by listing all the strings

    Infinite Language: specify by the following scheme

$$L = \{w \in \sum^* : w \text{ has property } P\}$$

Example: $L = \{ab, aabb, aaabbb, \cdots\} = \{a^n b^n \mid n \geq 1\}$

Theorem: If $\sum$ is a finite alphabet, then $\sum^*$ is countably infinite set.

Proof: Construct a bijection $f : \mathbb{N} \to \sum^*$.

Fix some ordering of the alphabet, say $\sum = \{a_1, a_2, \cdots, a_n\}$.

The member of $\sum^*$ can be enumerated in the following way:

1) For each $k \geq 0$, all string of length k are enumerated before all strings of length $k + 1$.
2) The n^k strings of length exactly k are enumerated lexicographically.

Operations of Languages:

- **Union, Intersection, Difference, Complement**

$(\overline{A} = \sum^* - A)$

- **Concatenation:**

$$L^0 = \{e\}$$
$$L^{i+1} = LL^i, \text{ for each } i \geq 0$$

$$L_1L_2 = \{w_1w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$$

Example:

$L_1 = \{w \in \{0,1\}^* : w$ has an even number of 0's$\}$

$L_2 = \{w \in \{0,1\}^* : w$ starts with a 0, the rest symbol are 1's$\}$

$$L_1L_2 = \{w \in \{0,1\}^* : w \text{ has an odd number of 0's}\}.$$

- **Kleene Star**

$$L^* = \{w \in \sum^* : w = w_1 \cdots w_k, k \geq 0, w_1, \cdots, w_k \in L\}$$

$$= L^0 \cup L^1 \cup L^2 \cup \cdots$$

$$L^+ = L^1 \cup L^2 \cup L^3 \cup \cdots$$

Example: $L = \{w \in \{0,1\}^* : w$ has an unequal number of 0's and 1's$\}$. Then $L^* = \{0,1\}^*$.

Hint: $L_1 \subseteq L_2 \Rightarrow L_1^* \subseteq L_2^*$ $\{0,1\} \subseteq L$

Remark:

1) The use of $\sum^*$ to denote the set of all strings over $\sum$ is consistent with the notation for the Kleene star of $\sum$.

2) $\emptyset^* = \{e\}$

3) $L^+ = LL^*$

4) For any language L, $(L^*)^* = L^*; L\emptyset = \emptyset L = \emptyset$

1.8 Finite Representations of Languages

- **Finite Representations:**
  - must be a string
  - different languages to have different representations

Representations ($\sum^*$, countable)	Languages ($2^{\sum^*}$, uncountable)

$\Rightarrow \exists$ undescribable languages!

Question: Prove that $2^{\sum^*}$ is uncountable.

Regular Expressions

Example Let $L = \{w \in \{0,1\}^* : w$ has two or three occurrences of 1, the first and second of which are not consecutive$\}$.

- The language can be described using only singleton sets and the symbols $\cup$, $\circ$, and $*$ as

$$\{0\}^* \circ \{1\} \circ \{0\}^* \circ \{0\} \circ \{1\} \circ \{0\}^*((\{1\} \circ \{0\}^*) \cup \emptyset^*)$$

- The language can be written simply as

$$0^*10^*010^*(10^* \cup \emptyset^*).$$

– **Regular Expressions**

Definition: The regular expressions are all strings over the alphabet $\sum \cup \{(,), \cup, \star\}$ that can be obtained as follows.

1) Θ and $\{x\}(\forall x \in \sum)$ is a regular expression.

2) If α and β are regular expressions, then so are $(\alpha\beta)$, $(\alpha \cup \beta)$, $\alpha^\star$.

3) Nothing is regular expression unless it follows from 1) through 2).

Example: a^*b^* $\quad a^* \cup b^*$ $\quad a(a^* \cup b^*)$

$(a^* \cup b^*)a(a^* \cup b^*)$ $\quad aaaaa^*$

□ Regular expressions & languages.

The function $\mathscr{L}$ is defined as follows.

1) $\mathscr{L}(\Theta) = \emptyset$, and $\mathscr{L}(a) = \{a\}$ for each $a \in \sum$.
2) If α and β are regular expressions, then

$$\mathscr{L}(\alpha\beta) = \mathscr{L}(\alpha)\mathscr{L}(\beta)$$
$$\mathscr{L}(\alpha \cup \beta) = \mathscr{L}(\alpha) \cup \mathscr{L}(\beta)$$
$$\mathscr{L}(\alpha^\star) = \mathscr{L}(\alpha)^*$$

Example: What is $\mathscr{L}(((a \cup b)^\star a))$? ?

$$\begin{aligned}
\mathscr{L}(((a \cup b)^\star a)) &= \mathscr{L}((a \cup b)^\star)\mathscr{L}(a) \\
&= \mathscr{L}((a \cup b)^\star)\{a\} \\
&= \mathscr{L}((a \cup b))^*\{a\} \\
&= (\mathscr{L}(a) \cup \mathscr{L}(b))^*\{a\} \\
&= (\{a\} \cup \{b\})^*\{a\} \\
&= (\{a, b\})^*\{a\} \\
&= \{w \in \{a, b\}^* : w \text{ ends with an } a\}
\end{aligned}$$

Example: What language is represented by $(c^\star(a \cup (bc^\star))^\star)$?

$L = \{w \in \{a, b, c\}^* :$ not have the substring $ac\}$.

Regular Expression Identities

- $SR \neq RS$
- $S \cup R = R \cup S$
- $R(ST) = (RS)T$
- $R(S \cup T) = RS \cup RT,\ (R \cup S)T = RT \cup ST$
- $\emptyset^* = \{e\}$
- $(R^*)^* = R^*$
- $(R^*S^*)^* = (R \cup S)^*$
- $(\{e\} \cup R)^* = R^*$

Remark:

1) Every language that can be represented by a regular expression can be represented by infinitely many of them.

2) The class of regular languages over an alphabet $\sum$ is defined to consist of all languages L such that $L = L(a)$ for some regular expression a over $\sum$. i.e. the class of regular languages over an alphabet $\sum$ is precisely the closure of the set of languages

$$\{\{\sigma\} : \sigma \in \sum\} \cup \{\emptyset\}$$

3) The regular expressions are an inadequate specification method in general.
For example, $\{0^n1^n : n \geq 0\}$ cannot be described by regular expressions.

4) Two important and useful means of representing languages:
– language recognition device

to answer questions of the form "Is string w a member of L?
– language generators

Homework 1:	
P46	1.7.4 (c)(d) 1.7.6
P51	1.8.3(c) 1.8.5