Sets and Languages

Debdeep Mukhopadhyay
IIT Madras

Introduction to Set Theory

- A *set* is a new type of structure, representing an *unordered* collection (group, plurality) of zero or more *distinct* (different) objects.
- Set theory deals with operations between, relations among, and statements about sets.
- Sets are ubiquitous in computer software systems.
- *All* of mathematics can be defined in terms of some form of set theory (using predicate logic).

Naïve set theory

- **Basic premise:** Any collection or class of objects (*elements*) that we can *describe* (by any means whatsoever) constitutes a set.
- But, the resulting theory turns out to be *logically inconsistent*!
  - This means, there exist naïve set theory propositions *p* such that you can prove that both *p* and $\neg p$ follow logically from the axioms of the theory!
  - $\therefore$ The conjunction of the axioms is a contradiction!
  - This theory is fundamentally uninteresting, because any possible statement in it can be (very trivially) "proved" by contradiction!
- More sophisticated set theories fix this problem.

Basic notations for sets

- For sets, we'll use variables *S*, *T*, *U*, …
- We can denote a set *S* in writing by listing all of its elements in curly braces:
  - {a, b, c} is the set of whatever 3 objects are denoted by a, b, c.
- *Set builder notation*: For any proposition $P(x)$ over any universe of discourse, $\{x|P(x)\}$ is *the set of all x such that P(x).*

Basic properties of sets

- Sets are inherently *unordered*:
  - No matter what objects a, b, and c denote, {a, b, c} = {a, c, b} = {b, a, c} = {b, c, a} = {c, a, b} = {c, b, a}.
- All elements are *distinct* (unequal); multiple listings make no difference!
  - If a=b, then {a, b, c} = {a, c} = {b, c} = {a, a, b, a, b, c, c, c, c}.
  - This set contains (at most) 2 elements!

Definition of Set Equality

- Two sets are declared to be equal *if and only if* they contain exactly the same elements.
- In particular, it does not matter *how the set is defined or denoted.*
- **For example:** The set {1, 2, 3, 4} =

  {x | x is an integer where x>0 and x<5 } =

  {x | x is a positive integer whose square is >0 and <25}

Infinite Sets

- Conceptually, sets may be *infinite* (*i.e.,* not *finite*, without end, unending).
- Symbols for some special infinite sets:
  N = {0, 1, 2, …}   The **N**atural numbers.
  Z = {…, -2, -1, 0, 1, 2, …} The **Z**ntegers.
  R = The "**R**eal" numbers, such as 374.1828471929498181917281943125…
- "Blackboard Bold" or double-struck font ($\mathbb{N}$,$\mathbb{Z}$,$\mathbb{R}$) is also often used for these special number sets.
- Infinite sets come in different sizes!

Venn Diagrams

John Venn
1834-1923

0 2 4 6 8
-1 1 3 5 7 9

Even integers from 2 to 9

Odd integers from 1 to 9

Primes <10

Positive integers less than 10

Integers from -1 to 9

Basic Set Relations: Member of

- $x \in S$ ("x is in S") is the proposition that object x is an $\in$*lement* or *member* of set S.
  - *e.g.* $3 \in \mathbf{N}$, "a"$\in \{x \mid x$ is a letter of the alphabet$\}$
  - Can define set equality in terms of $\in$ relation:
    $\forall S,T\text{: } S=T \leftrightarrow (\forall x\text{: } x \in S \leftrightarrow x \in T)$
    "Two sets are equal iff they have all the same members."
- $x \notin S :\equiv \neg(x \in S)$ "x is not in S"

The Empty Set

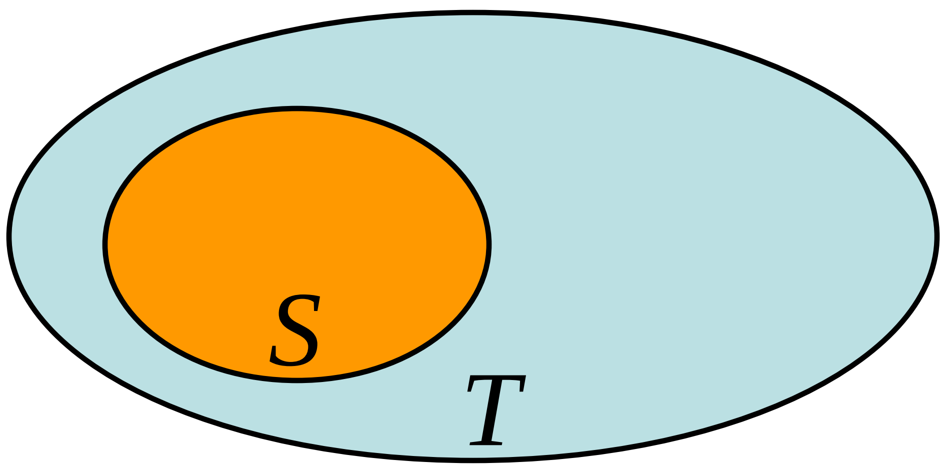
- $\varnothing$ ("null", "the empty set") is the unique set that contains no elements whatsoever.
- $\varnothing = \{\} = \{x|\mathbf{False}\}$
- No matter the domain of discourse, we have the axiom $\neg\exists x$: $x \in \varnothing$.

Subset and Superset Relations

- $S \subseteq T$ ("S is a subset of T") means that every element of S is also an element of T.
- $S \subseteq T \Leftrightarrow \forall x\, (x \in S \rightarrow x \in T)$
- $\emptyset \subseteq S$, $S \subseteq S$.
- $S \supseteq T$ ("S is a superset of T") means $T \subseteq S$.
- Note $S = T \Leftrightarrow S \subseteq T \wedge S \supseteq T$.
- $S \not\subseteq T$ means $\neg(S \subseteq T)$, *i.e.* $\exists x(x \in S \wedge x \notin T)$

Proper (Strict) Subsets & Supersets

- $S \subset T$ ("S is a proper subset of T") means that $S \subseteq T$ but $T \not\subseteq S$. Similar for $S \supset T$.



Example:
$\{1,2\} \subset \{1,2,3\}$

Venn Diagram equivalent of $S \subset T$

Sets Are Objects, Too!

- The objects that are elements of a set may *themselves* be sets.
- *E.g.* let $S=\{x \mid x \subseteq \{1,2,3\}\}$
  then $S=\{\varnothing,$
  $\{1\}, \{2\}, \{3\},$
  $\{1,2\}, \{1,3\}, \{2,3\},$
  $\{1,2,3\}\}$
- Note that $1 \neq \{1\} \neq \{\{1\}\}$!!!!

Cardinality and Finiteness

- |*S*| (read "the *cardinality* of *S*") is a measure of how many different elements *S* has.
- *E.g.*, $|\emptyset|=0$, $|\{1,2,3\}| = 3$, $|\{a,b\}| = 2$,
  $|\{\{1,2,3\},\{4,5\}\}| =$ __2__
- If $|S| \in \mathbf{N}$, then we say *S* is *finite*. Otherwise, we say *S* is *infinite*.
- What are some infinite sets we've seen?
  N Z R

The *Power Set* Operation

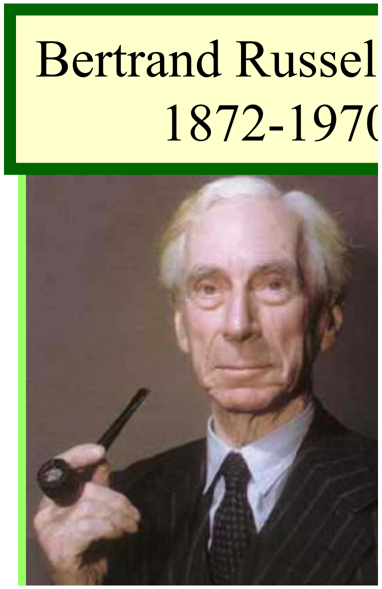
- The *power set* P(*S*) of a set *S* is the set of all subsets of *S*. $P(S) :\equiv \{x \mid x \subseteq S\}$.
- *E.g.* P({a,b}) = {∅, {a}, {b}, {a,b}}.
- Sometimes P(*S*) is written $\mathbf{2}^S$. Note that for finite *S*, $|P(S)| = 2^{|S|}$.
- It turns out $\forall S{:}|P(S)|>|S|$, *e.g.* $|P(\mathbf{N})| > |\mathbf{N}|$. *There are different sizes of infinite sets*!

Review: Set Notations So Far

- Variable objects x, y, z; sets S, T, U.
- Literal set {a, b, c} and set-builder $\{x|P(x)\}$.
- $\in$ relational operator, and the empty set $\varnothing$.
- Set relations $=$, $\subseteq$, $\supseteq$, $\subset$, $\supset$, $\not\subset$, etc.
- Venn diagrams.
- Cardinality $|S|$ and infinite sets **N**, **Z**, **R**.
- Power sets P(S).

Naïve Set Theory is Inconsistent

- There are some naïve set *descriptions* that lead to pathological structures that are not *well-defined*.
  - (That do not have self-consistent properties.)
- These "sets" mathematically *cannot* exist.
- *E.g.* let $S = \{x \mid x \notin x\}$. Is $S \in S$?
- Therefore, consistent set theories must restrict the language that can be used to describe sets.
- For purposes of this class, don't worry about it!



Ordered *n*-tuples

- These are like sets, except that duplicates matter, and the order makes a difference.
- For $n \in \mathbf{N}$, an *ordered n-tuple* or a *sequence* or *list of length n* is written $(a_1, a_2, \ldots, a_n)$. Its *first* element is a_1, *etc.*
- Note that $(1, 2) \neq (2, 1) \neq (2, 1, 1)$.
- Empty sequence, singlets, pairs, triples, quadruples, quintuples, …, *n*-tuples.

Contrast with sets' {}

Cartesian Products of Sets

- For sets A, B, their *Cartesian product* $A \times B :\equiv \{(a, b) \mid a \in A \wedge b \in B\}$.
- *E.g.* $\{a,b\} \times \{1,2\} = \{(a,1),(a,2),(b,1),(b,2)\}$
- Note that for finite A, B, $|A \times B| = |A||B|$.
- Note that the Cartesian product is *no* commutative: *i.e.*, $\neg \forall AB: A \times B = B \times A$.
- Extends to $A_1 \times A_2 \times \ldots \times A_n$...

René Descartes
(1596-1650)

Review

- Sets S, T, U… Special sets **N**, **Z**, **R**.
- Set notations {a,b,...}, $\{x|P(x)\}$…
- Set relation operators $x \in S$, $S \subseteq T$, $S \supseteq T$, $S=T$, $S \subset T$, $S \supset T$. (These form propositions.)
- Finite vs. infinite sets.
- Set operations $|S|$, P(S), $S \times T$.
- More set ops: $\cup$, $\cap$, $-$.

The Union Operator

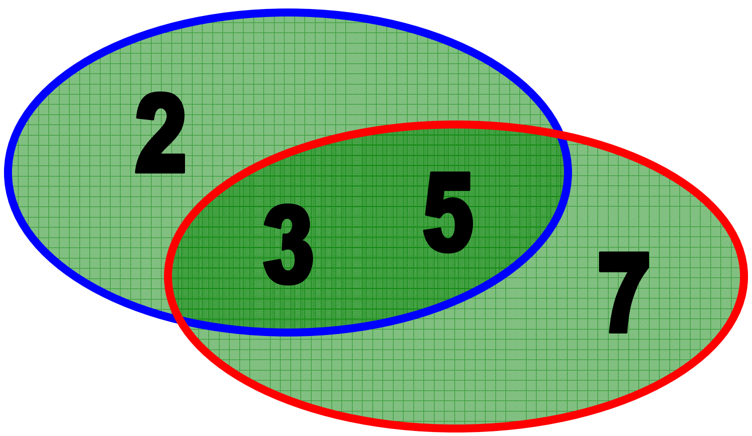
- For sets A, B, their *union* $A \cup B$ is the set containing all elements that are either in A, **or** ("$\vee$") in B (or, of course, in both).
- Formally, $\forall A,B$: $A \cup B = \{x \mid x \in A \vee x \in B\}$.
- Note that $A \cup B$ is a **superset** of both A and B (in fact, it is the smallest such superset):

$$\forall A, B: (A \cup B \supseteq A) \wedge (A \cup B \supseteq B)$$

Union Examples

- {a,b,c}∪{2,3} = {a,b,c,2,3}
- {2,3,5}∪{3,5,7} = {2,3,5,3,5,7} = {2,3,5,7}

Required Form



Think "The **Uni**ted States of America includes every person who worked in <u>any</u> U.S. state last year." (This is how the IRS sees it...)

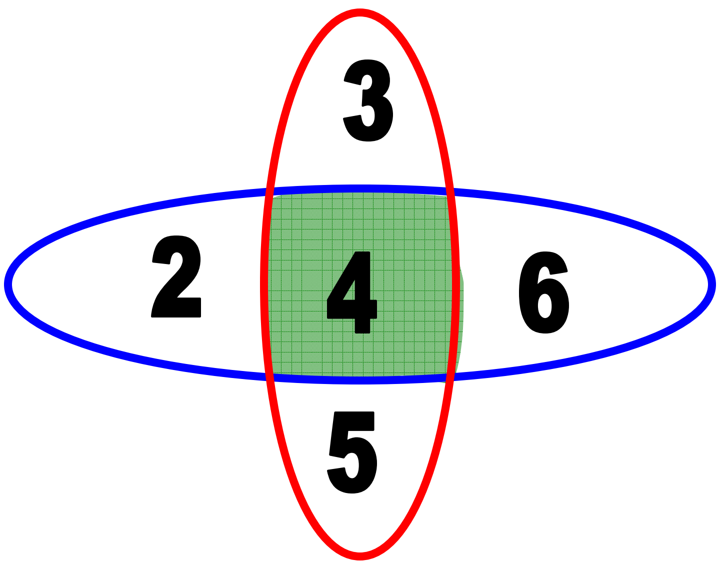
The Intersection Operator

- For sets *A*, *B*, their *intersection* $A \cap B$ is the set containing all elements that are simultaneously in *A* **and** ("$\land$") in *B*.
- Formally, $\forall A,B: A \cap B = \{x \mid x \in A \land x \in B\}$.
- Note that $A \cap B$ is a **subset** of both A and B (in fact it is the largest such subset):

$$\forall A, B: (A \cap B \subseteq A) \land (A \cap B \subseteq B)$$

Intersection Examples

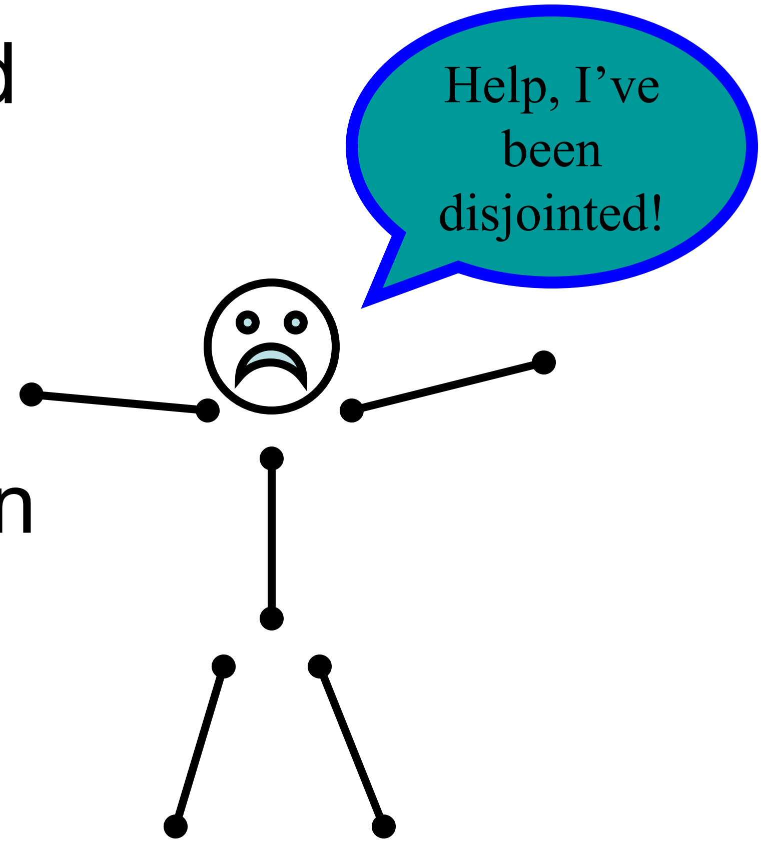
- $\{a,b,c\} \cap \{2,3\}$ = $\varnothing$
- $\{2,4,6\} \cap \{3,4,5\}$ = $\{4\}$



Think "The **intersection** of University Ave. and W 13th St. is just that part of the road surface that lies on *both* streets."

Disjointedness

- Two sets *A*, *B* are called *disjoint* (*i.e.*, unjoined) iff their intersection is empty. ($A \cap B = \varnothing$)
- Example: the set of even integers is disjoint with the set of odd integers.



Inclusion-Exclusion Principle

- How many elements are in $A \cup B$?
  $|A \cup B| = |A| + |B| - |A \cap B|$
- Example: How many students are on our class email list? Consider set $E = I \cup M$,
  $I = \{s \mid s$ turned in an information sheet$\}$
  $M = \{s \mid s$ sent the TAs their email address$\}$
- Some students did both!
  $|E| = |I \cup M| = |I| + |M| - |I \cap M|$

Set Difference

- For sets *A*, *B*, the *difference of A and B*, written *A*–*B*, is the set of all elements that are in *A* but not *B*. Formally:

$$A - B :\equiv \{x \mid x \in A \land x \notin B\}$$
$$= \{x \mid \neg(x \in A \rightarrow x \in B)\}$$

- Also called:
  The *complement of B with respect to A*.

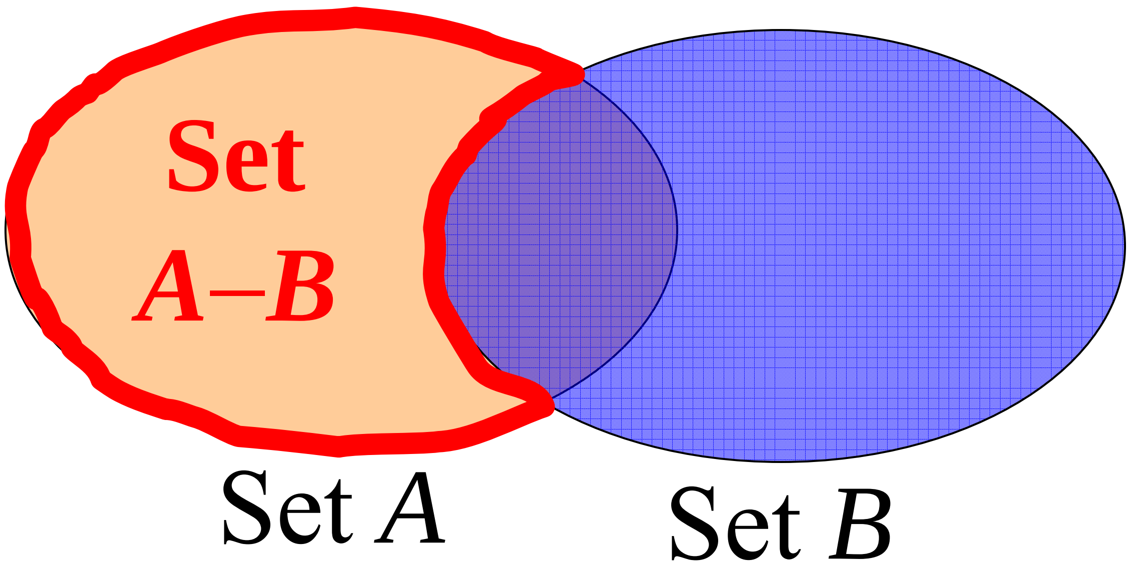
Set Difference Examples

- {1,2,3,4,5,6} − {2,3,5,7,9,11} =
  {1,4,6}
  ____

- **Z** − **N** = {… , −1, 0, 1, 2, … } − {0, 1, … }
  = {*x* | *x* is an integer but not a nat. #}
  = {*x* | *x* is a negative integer}
  = {… , −3, −2, −1}

Set Difference - Venn Diagram

- $A-B$ is what's left after B "takes a bite out of A"



Set Complements

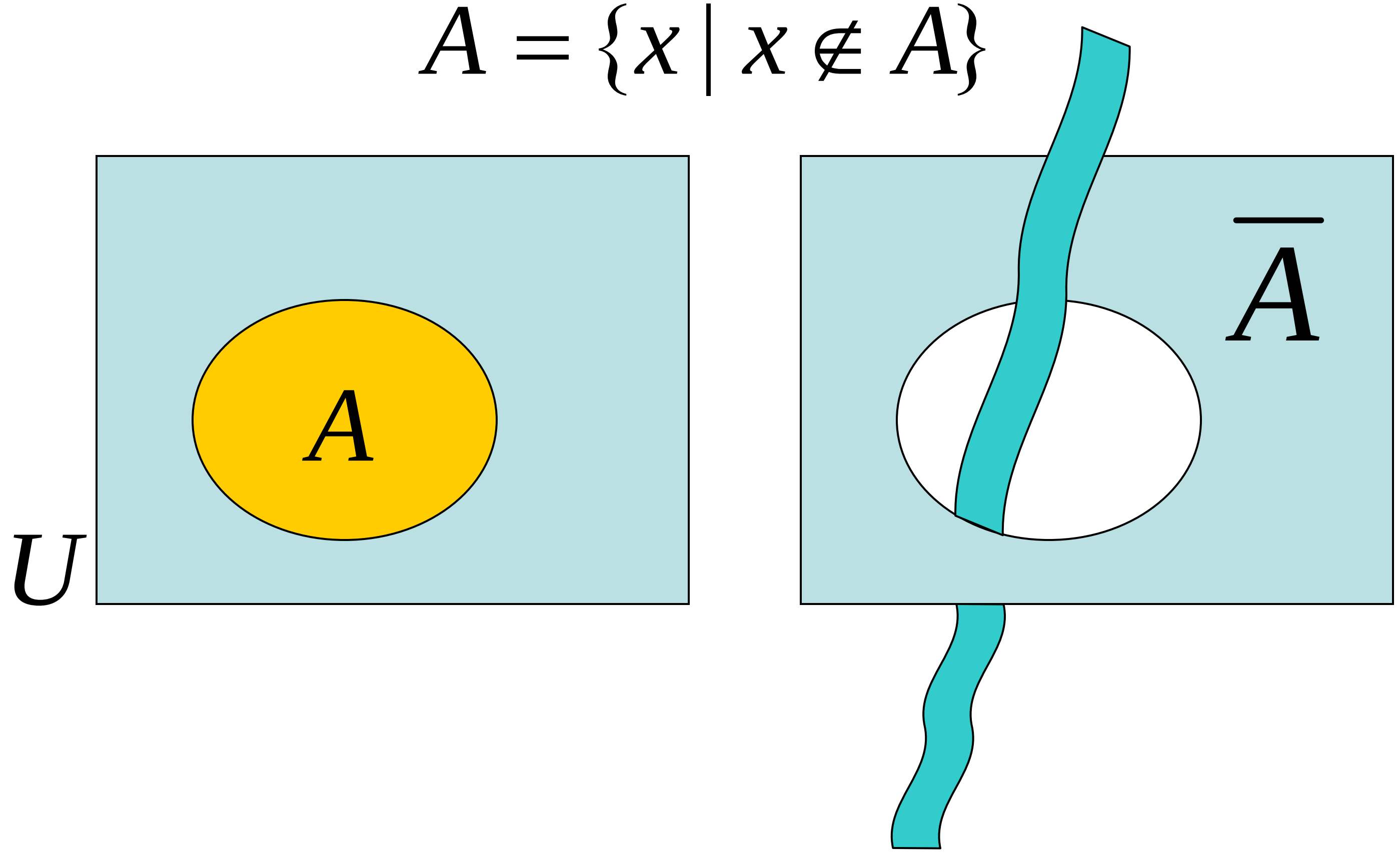
- The *universe of discourse* can itself be considered a set, call it *U*.
- When the context clearly defines *U*, we say that for any set $A \subseteq U$, the *complement* of *A*, written $\overline{A}$, is the complement of *A* w.r.t. *U*, *i.e.*, it is *U*–*A.*
- *E.g.*, If *U*=**N**,

$$\overline{\{3,5\}} = \{0,1,2,4,6,7,...\}$$

More on Set Complements

- An equivalent definition, when *U* is clear:

$$\overline{A} = \{x \mid x \notin A\}$$



Set Identities

- Identity: $A \cup \varnothing = A = A \cap U$
- Domination: $A \cup U = U$, $A \cap \varnothing = \varnothing$
- Idempotent: $A \cup A = A = A \cap A$
- Double complement: $\overline{(\overline{A})} = A$
- Commutative: $A \cup B = B \cup A$, $A \cap B = B \cap A$
- Associative: $A \cup (B \cup C) = (A \cup B) \cup C$,
  $A \cap (B \cap C) = (A \cap B) \cap C$

DeMorgan's Law for Sets

- Exactly analogous to (and provable from) DeMorgan's Law for propositions.

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Proving Set Identities

To prove statements about sets, of the form $E_1 = E_2$ (where the Es are set expressions), here are three useful techniques:

1. Prove $E_1 \subseteq E_2$ and $E_2 \subseteq E_1$ separately.
2. Use set builder notation & logical equivalences.
3. Use a *membership table*.

Method 1: Mutual subsets

Example: Show $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

- Part 1: Show $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.
  - Assume $x \in A \cap (B \cup C)$, & show $x \in (A \cap B) \cup (A \cap C)$.
  - We know that $x \in A$, and either $x \in B$ or $x \in C$.
    - Case 1: $x \in B$. Then $x \in A \cap B$, so $x \in (A \cap B) \cup (A \cap C)$.
    - Case 2: $x \in C$. Then $x \in A \cap C$, so $x \in (A \cap B) \cup (A \cap C)$.
  - Therefore, $x \in (A \cap B) \cup (A \cap C)$.
  - Therefore, $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.
- Part 2: Show $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. …

Method 3: Membership Tables

- Just like truth tables for propositional logic.
- Columns for different set expressions.
- Rows for all combinations of memberships in constituent sets.
- Use “1” to indicate membership in the derived set, “0” for non-membership.
- Prove equivalence with identical columns.

Membership Table Example

Prove $(A \cup B) - B = A - B$.

A	B	$A \cup B$	$(A \cup B) - B$	$A - B$
0	0	0	0	0
0	1	1	0	0
1	0	1	1	1
1	1	1	0	0

Membership Table Exercise

Prove $(A \cup B) - C = (A - C) \cup (B - C)$.

A	B	C	$A \cup B$	$(A \cup B) - C$	$A - C$	$B - C$	$(A - C) \cup (B - C)$
0	0	0					
0	0	1					
0	1	0					
0	1	1					
1	0	0					
1	0	1					
1	1	0					
1	1	1					

Review

- Sets S, T, U… Special sets **N**, **Z**, **R**.
- Set notations {a,b,...}, $\{x|P(x)\}$…
- Relations $x \in S$, $S \subseteq T$, $S \supseteq T$, $S = T$, $S \subset T$, $S \supset T$.
- Operations $|S|$, P(S), $\times$, $\cup$, $\cap$, $-$, $\overline{S}$,
- Set equality proof techniques:
  - Mutual subsets.
  - Derivation using logical equivalences.

Generalized Unions & Intersections

- Since union & intersection are commutative and associative, we can extend them from operating on *ordered pairs* of sets (A,B) to operating on sequences of sets $(A_1,\ldots,A_n)$, or even on unordered *sets* of sets, $X=\{A \mid P(A)\}$.

Generalized Union

- Binary union operator: $A \cup B$
- n-ary union:
  $A \cup A_2 \cup \ldots \cup A_n :\equiv ((\ldots((A_1 \cup A_2) \cup \ldots) \cup A_n)$
  (grouping & order is irrelevant)
- "Big U" notation:
  $$\bigcup_{i=1}^{n} A_i$$
- Or for infinite sets of sets:
  $$\bigcup_{A \in X} A$$

Generalized Intersection

- Binary intersection operator: $A \cap B$
- *n*-ary intersection:
  $A_1 \cap A_2 \cap \ldots \cap A_n \equiv ((\ldots((A_1 \cap A_2) \cap \ldots) \cap A_n)$
  (grouping & order is irrelevant)
- "Big Arch" notation: $\bigcap_{i=1}^{n} A_i$
- Or for infinite sets of sets: $\bigcap_{A \in X} A$

Representations

- A frequent theme of this course will be methods of *representing* one discrete structure using another discrete structure of a different type.
- *E.g.*, one can represent natural numbers as
  – Sets: **0**:≡∅, **1**:≡{**0**}, **2**:≡{**0**,**1**}, **3**:≡{**0**,**1**,**2**}, …
  – Bit strings:
  0:≡`0`, **1**:≡`1`, **2**:≡`10`, **3**:≡`11`, **4**:≡`100`, …

Representing Sets with Bit Strings

For an enumerable u.d. U with ordering $x_1, x_2, \ldots$, represent a finite set $S \subseteq U$ as the finite bit string B=$b_1 b_2 \ldots b_n$ where $\forall i: x_i \in S \leftrightarrow (i<n \wedge b_i=1)$.

E.g. $U=\mathbf{N}$, $S=\{2,3,5,7,11\}$, B=001101010001.

In this representation, the set operators "$\cup$", "$\cap$", "$\overline{}$" are implemented directly by bitwise OR, AND, NOT!

Set Operations on Σ^*

- What is Σ and Σ^* ?
- Σ : Also, called alphabet. Finite and nonempty set of symbols or characters.
- Each element of this set can form what is called string of length say *l*.
- The length denotes the number of characters in the string.
- Λ : The empty string, $l=0$

What is Σ^*

- Σ^* : set of all finite strings of symbols from the alphabet. This includes Λ (empty string).
- Definition:
  - Let Σ be an alphabet. Then Σ^* is defined as follows:
    1. (Basis) $\Lambda \in \Sigma^*$
    2. (Induction) If $x \in \Sigma^*$ and $a \in \Sigma$, then $ax \in \Sigma^*$
    3. (Extremal) Nothing is an element of the set Σ^* unless it can be constructed with a finite number of applications of clause 1 and 2.

Examples:

$If\ ,\Sigma = \{a,b\},then\ \Sigma^* = \{\Lambda, a, b, aa, ab, ba, bb, aaa, aab, ...\}$

$If\ ,\Sigma = \{0,1\},then\ \Sigma^*$ is the set of all finite binary sequences, including the empty sequence.

Concatenation of strings

- Let, Σ be an alphabet and x and y are elements of Σ^* . If $x=a_1a_2\ldots a_m$ and $y=b_1b_2\ldots b_n$ where a_i,b_j belongs to the alphabet and m,n are integers from 0,1,2,…. Then the concatenation of x with y, denoted by x.y, x||y or simply as xy is the string: $xy=a_1a_2\ldots a_mb_1b_2\ldots b_n$. If $x=\Lambda$, then xy=y for every y; similarly if $y=\Lambda$, then xy=x for every x.

Concatenation of x with itself

- Defn: Let x be an element of Σ^* . For each n belongs to N, x^n is defined as:

$$1.\ x^0 = \Lambda$$

$$2.\ x^{n+1} = x^n x$$

$If, \Sigma = \{a,b\}$ and $x = ab, then\ x^0 = \Lambda, x^1 = ab, x^2 = abab$

$S = \{a^n b^n \mid n \geq 0\}$ denotes the set, $S = \{\Lambda, ab, aabb, aaabbb, ...\}$

Language

- Language over Σ is a subset of Σ^*.

(a) The set $\{a, ab, abb\}$ is a language over $\Sigma = \{a, b\}$

(b) The set of strings consisting of sequences of a's followed by sequences of b's: $\{a^n b^m \mid n, m \in N\}$

Let A and B be languages over Σ. The set product of A and B is denoted by A.B, or simply AB is the language:

$AB = \{xy \mid x \in A \wedge x \in B\}$

Note: In general AB is not the same as BA. The set product is not commutative.

Properties of the set product

Let A, B, C and D be arbitrary languages over Σ. The following relations hold:

(a) $A\Phi=\Phi A=\Phi$

(b) $A\{\Lambda\}=\{\Lambda\}A=A$

(c) $(AB)C=A(BC)$

(d) If $A\subset B$ and $C\subset D$, then $AC\subset BD$

(e) $A(B\cup C)=AB\cup AC$

(f) $(B\cup C)A=BA\cup CA$

(g) $A(B\cap C)\subset AB\cap AC$

(h) $(B\cap C)A\subset BA\cap CA$

Product of language

Let A be a language over $\sum$. The language A^n is defined as follows:

1. $A^0 = \{\Lambda\}$
2. $A^{n+1} = A^n A$, for $n \in \mathbb{N}$

The language A^n is the set product of A with itself n times.

So, if $z \in A^n$, then $z = w_1 w_2 \ldots w_n$, where each $w_i \in A$, for each i from 1 to n.

Theorem

Let A and B be subsets of Σ^*, and let m and n be arbitrary elements of $\mathbb{N}$. Then,

(a) $A^m A^n = A^{m+n}$

$(b)(A^m)^n = A^{mn}$

$(c) A \subset B \Rightarrow A^n \subset B^n$

Proof: part (a) and (b) is left as an exercise.
Part (c) follows from mathematical induction.

Kleene Closure or Star closure of A

Let A be a subset of Σ^*. Then the set A^* is defined as:

$$A^* = \bigcup_{n \in N} A^n$$

That is $A^* = \{\Lambda\} \cup A \cup A^2 \cup \ldots$

Positive Closure of A

The set A^* is defined as:

$$A^+ = \bigcup_{n=1}^{\infty} A^n$$

Examples

(a) If A={a}, then

$$A^{+} = \{a\} \cup \{aa\} \cup \{aaa\} \cup \ldots$$

$$= \{a^{n} \mid n \geq 1\}$$

$$A^{*} = \{a^{n} \mid n \geq 0\}$$

$$(b)\Phi^{*} = \{\Lambda\}, \Phi^{+} = \Phi$$

Properties of the language closure

$(a)\ A^* = \{\Lambda\} \cup A^+$

$(b)\ A^n \subset A^*,\ \text{for n} \geq 0$

$(c)\ A^n \subset A^+,\ \text{for n} \geq 1$

$(d)\ A \subset AB^*$

$(e)\ A \subset B^*A$

$(f)\ (A \subset B) \Rightarrow (A^* \subset B^*)$

$(g)\ (A \subset B) \Rightarrow (A^+ \subset B^+)$

$(h)\ AA^* = A^*A = A^+$

$(i)\ \{\Lambda\} \in A \Leftrightarrow A^+ = A^*$

$(j)\ (A^*)^* = A^*A^* = A^*$

$(k)\ (A^*)^+ = (A^+)^+ = A^*$

$(l)\ A^*A^+ = A^+A^* = A^+$

$(m)\ (A^*B^*)^* = (A \cup B)^* = (A^* \cup B^*)^*$

We shall see the proves of some of the results in the class. Rest are left to you as an exercise

Dean Arden's Theorem

- Let A and B be arbitrary subsets of Σ^* such that $\Lambda \notin A$.
  Then the equation $X = AX \cup B$ has the unique solution $X = A^*B$

$$X \supseteq A^*B:$$

$$X = AX \cup B \Rightarrow (X \supseteq B) \wedge (X \supseteq AX)$$

$$\Rightarrow (X \supseteq AB)$$

$$(X \supseteq AB) \wedge (X \supseteq AX) \Rightarrow (X \supseteq AAB)$$

Thus, in general $(X \supseteq A^nB) \Rightarrow (X \supseteq A^*B)$

Dean Arden's Theorem

$X \subseteq A^* B$:

Consider, ($x \in X$).

Thus, $x \in B$ or, $x \in AX$.

Since, $\Lambda \notin$ A x must belong to B or must have a non-empty prefix which belongs to A and the rest of the string is a shorter string in X.

By the same reason the shorter string also belongs to B or we can remove another prefix string belonging to A and obtain another string in X.

Since, the original string is a finite string after a finite number of steps we have a string in B.

Thus, in a nutshell the original string must consist of a (possibly) empty sequence of prefixes, each belonging to A, followed by a suffix which is in B. Thus, $x \in A^* B$.

But does the solution always exist?

Examples

- If A={a}, B=Φ, then the equation X=AXUB, has the unique solution X=A*B= Φ
- If A={a,ab}, B={cc}, then the equation X=AXUB, has the unique solution

  X=A*B= {a,ab}*{cc}