Sets, tuples, maps, and languages

ian.mcloughlin@gmit.ie

Set: collection of objects

- Denoted by capital letters: A, B, X

- Objects in a set are called elements.

- Elements are denoted by lower case letters: a, b, x

- Curly braces around elements: $A = \{a_0, a_1, a_2\}$

Examples

$A = \{1, 2, 3\}$

$B = \{p \mid p \text{ is a prime number}\}$

No order and no count

A set doesn't maintain an order of its elements:

$$\{1, 2, 3\} \; = \; \{1, 3, 2\} \; = \; \{2, 1, 3\} \; = \; \{2, 3, 1\}$$
$$= \; \{3, 1, 2\} \; = \; \{3, 2, 1\}$$

An object is either in the set or not:

$$\{1, 2, 2, 3\} \; = \; \{1, 2, 3\}$$

Question: is 1.0 an element of $\mathbb{Z}$?

To a programmer, the answer is likely no, since 1.0 is a floating point number and not an integer. Compilers will sometimes give errors if you give 1.0 where an integer is expected.

```
IPython 6.1.0 -- An enhanced Interactive Python.
Type '?' for help.
In [1]: i = 1.0
In [2]: a = [1,4,9,16]
In [3]: a[i]
TypeError: list indices must be integers or slices,
not float
In [4]: a[int(i)]
Out[4]: 4
```
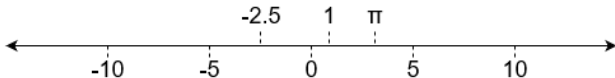
Question: is 1.0 an element of $\mathbb{Z}$?

Most mathematicians, on the other hand, will likely say 1.0 is an integer. To them, 1 and 1.0 are just different representations of the same point on a number line. They think of $\mathbb{Z}$ as a subset of the real numbers $\mathbb{R}$. Every number on the usual number line is a real number, including π, 1.5 and 10.

There's not a lot we can do, other than to provide some context when we consider such sets. The discussion can get philosophical, especially in discussions about symbols and semantics. Again, this is an important concept in the theory of computation.

Sets containing sets

Subsets

A is a subset of B if all of the elements of A are in B.

$$A = 1, 2, 3, 4 \qquad B = 2, 3 \qquad B \subset A$$

Powersets

Some sets contain other sets as elements. The powerset of a set is the set containing all subsets of it:

$$A = 1, 2, 3$$
$$\mathcal{P}(A) = \{\{\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

Note A contains 3 elements and $\mathcal{P}(A)$ contains $2^3 = 8$.

Famous sets

$\mathbb{N}$ – the natural numbers $\{1, 2, 3, \ldots\}$.

$\mathbb{N}_0$ – the natural numbers with zero $\{0, 1, 2, 3, \ldots\}$.

$\mathbb{Z}$ – the integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

$\mathbb{Q}$ – the rational numbers $\{\frac{m}{n} \mid m, n \in \mathbb{Z}\}$.

$\mathbb{R}$ – the real numbers.

$\mathbb{C}$ – the complex numbers $\{a + bi \mid a, b \in \mathbb{R}, i^2 = -1\}$.

Sizes of sets

$$A = \{a, b, c\} \;\Rightarrow\; |A| = 3$$

- Number of elements is denoted with vertical lines.

- The set of all prime numbers is an *infinite* set.

- Infinite sets can still have a notion of size.

- $\mathbb{R}$ is bigger than $\mathbb{N}$ even though they're both infinite – important consequences for computation.

Operations on sets

$$A = \{1, 2, 3\} \qquad B = \{2, 3, 4\}$$

Union: $A \cup B = \{1, 2, 3, 4\}$, in A **or** B.

Intersection: $A \cap B = \{2, 3\}$, in A **and** B.

Difference: $A \setminus B = \{1\}$, in A **not** B.

Tuples: finite list of elements taken from sets

$$t = (2, 1, 1) \qquad t \in \mathbb{N} \times \mathbb{N} \times \mathbb{N} \qquad |t| = 3$$

- Round brackets denote tuples, and t is a 3-tuple or a triple.

- Tuples have order, and can repeat elements.

- Sometimes we omit the brackets and commas: $t = 211$.

- $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ is sometimes shortened to $\mathbb{N}^3$.

- The first $\mathbb{N}$ means the first element comes from $\mathbb{N}$.

- The second $\mathbb{N}$ means the second element comes from $\mathbb{N}$, etc.

- Note that there is a single empty tuple: $()$.

Cartesian products of sets

$$A = \{1, 2, 3\} \qquad B = \{x, y\}$$

$$A \times B = \{(1, x), (2, x), (3, x), (2, y), (2, y), (3, y)\}$$

– $A \times B$ is called the cartesian product of A and B – the set of tuples with first element from A and second from B.

– $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ is the usual 2D plane where we draw plots.

– Can extend to any length of tuple: $\mathbb{R}^3$ is the 3D plane.
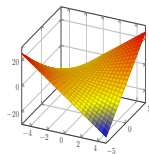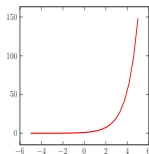
Multisets

$$M = \{(a, 2), (b, 10), (x, 5)\}$$

– We can use sets and tuples to define other data structures.

– A multiset over a set A is a subset of $A \times \mathbb{N}$ such that every element of A appears exactly once as the first element in a tuple.

Maps

Definition of map

A map from a set A to a set B is a subset M of $A \times B$ where each element of A appears as the first element of a tuple in M exactly once.

$$A = \{a, b, c\} \qquad B = \{x, y, z\}$$

Maps

– $\{(a, x), (b, x), (c, x)\}$
– $\{(a, x), (b, y), (c, z)\}$

Not maps

– $\{(a, x), (a, y), (b, x), (c, x)\}$
– $\{(a, x), (b, y)\}$

One-to-one map

$$A = \{1, 2, 3\} \quad B = \{a, b, c, d\} \quad M = \{(1, a), (2, b), (3, d)\}$$

- In a map, $M \subseteq A \times B$, two or more distinct elements of A can be mapped to the same element of B.

- A map where this does not happen is described as **one-to-one**.

- So, a map in which distinct elements of A go to distinct elements of B is one-to-one.

Onto map

$$A = \{1, 2, 3\} \quad B = \{a, b, c, d\} \quad M = \{(1, a), (2, a), (3, b)\}$$

- In a map, not all of the elements of B have to be involved in the map.

- A map in which they are all involved is described as **onto**.

- So, a map in which each element of B is paired with an element of A is onto.

Bijections

$$A = \{1, 2, 3\} \quad B = \{a, b, c\} \quad M = \{(1, a), (2, c), (3, b)\}$$

- Maps can be neither one-to-one nor onto, one or the other, or both.

- A **bijection** is map that is both one-to-one and onto.

- Both A and B must have the same size in a bijection.

Partial maps

$$A = \{1, 2, 3\} \quad B = \{a, b, c\} \quad P = \{(1, a), (3, c)\}$$

- A map M from a set A to a set B must involve every element of A.

- A partial map is like a map, but with that condition relaxed.

- A partial map from a set A to a set B is a subset of $A \times B$ where any element of A that appears as the first element in a tuple does so exactly once.

- The term *partial map* is a bit of a misnomer, as a partial map is not necessarily a map, but a map is a partial map.

Alphabets and strings

- In some contexts, sets are called alphabets and tuples over them are called strings or words.
- We omit the brackets and commas, and the empty tuple is called the empty string and denoted ϵ.

Example

Let A be the alphabet $\{0, 1\}$. The following are strings over that alphabet:

$$\epsilon, 0, 1, 00, 01, 10, 11, 001, 010, 011, 100, \ldots$$

Concatenation

We can concatenate strings. For example, concatenating 00 and 101 gives 00101. Technically: $(0, 0) \circ (1, 0, 1) = (0, 0, 1, 0, 1)$.

Kleene star

$$A = \{0, 1\}$$

$A^1 = \{0, 1\}$, the strings of length one over A.

$A^2 = \{00, 01, 10, 11\}$, the strings of length two over A.

$A^i =$ the strings of length i over A.

$A^0 = \{\epsilon\}$, the strings of length zero over A.

Definition

The Kleene star (denoted A^*) of A is the union of all the A^i:

$$\bigcup_i A^i = \{\epsilon, 0, 1, 00, 01, 10, \ldots\}$$

Note the alphabet can be different.

Languages

$$L \subseteq A^*$$

Any subset of the star of an alphabet is a **language**.

Operations
Create three new languages from two languages L_1 and L_2.

Union: $L_1 \cup L_2$, all strings in L_1 or L_2.

Concatenate: $L_1 \circ L_2$, join each string in L_1 with each in L_2.

Star: L_1^*, the Kleene star of L_1 using concatenation.

We could consider two languages over different alphabets, if we took the union of the alphabets.

File types as languages

- The set of all valid PDF files, according to the PDF specification, is a language over the alphabet $A = \{0, 1\}$.

- So, the set of valid PDFs is a subset of A^*.

- So is the set of valid DOCX file.

- A computer program that converts PDF files to DOCX files just maps one subset of A^* to another subset of A^*.

- Remember too, that executable files themselves are stored as files in 0's and 1's.