Maths for Regular Expressions

A **set** is an unordered collection of values where each value occurs only once. Values can be numbers, symbols of letters. The contents of a set are represented using curly brackets. For instance, the set

$$A = \{1,\ 2,\ 3,\ 4,\ 5\}$$

defines all the integers between 1 and 5 inclusive, where the name of the set is A.

Notation of special sets

- N is the infinites set of natural numbers from 0 to infinity
- $N = \{0,1,2,3,4,\ \ldots.\}$ (infinite set has ellipses)
- $x \in N$ means x is a member of the set N
- R is the set of real numbers
- Empty sets are sets that contain no elements and are represented using {} or Ø

Set comprehension is a short hand for writing out sets. For instance they take the form of:

$$A = \{n \mid n \in N \wedge n < 7\}$$

This means the set is a set of natural numbers and the values are less than 7. The ^ character means Boolean AND and the | character means such that.
Therefore:

$$A = \{0,1,2,3,4,5,6\}$$

Examples of set comprehension

$A = \{\, x^2 \mid x \in N \wedge x < 4\}$	$A = \{0,1,4,9\}$
$A = \{2x \mid x \in N \wedge x < 4\}$	$A = \{0,2,4,6\}$
$A = \{\, 3x \mid x \in N \wedge x > 4 \wedge x < 10\}$	$A = \{15,18,21,24,27\}$

Compact representation of Sets

$$A = \{\ 1^n\, 0^n \mid n\ \geq\ 1\}$$

would produce the set:

$$A = \{10,1100,111000,11110000, \ldots\ldots\}$$

A **finite** set can be counted up by natural numbers. It has a certain number of elements.

An **infinite** set has an infinite number of elements

The **cardinality** of a finite set is the number of members in a set

A **countable** set can be counted off by a finite subset of the natural numbers. A countable set has the same number of elements as a subset of the natural numbers

Index	0	1	2	3	4
Value	2	4	5	-2	-7

A **countably infinite** set can be counted off by the natural numbers.

Real numbers are not countable and you do not know which is the next value because they can be infinitesimal.

The **Cartesian product** of two sets A and B (A x B) is the set of all combinations of pairs of elements in A and B.

e.g. $A = \{2,4,6\}$ and $B = \{3,5,7\}$ C=A X B,

$$C = \{(2,3), (2,5), (2,7), (4,3), (4,5), (4,7), (6,3), (6,5), (6,7)\}$$

Membership

A **Proper subset** A contains everything in the set B, but there is at least one element in set B is not contained in subset A. A is a proper subset of B (or B is a superset of A)

$$A \subset B, B \supset A$$

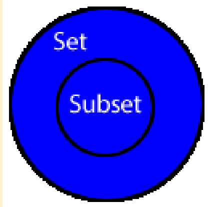
$$\text{e.g } A = \{0, 1, 2\}, B = \{0,\ 1,\ 2,\ 3\}$$

That is subset A will always have fewer elements than set B even if it only has one fewer element. In other words, All members of a subset will also be in a set. If there are n elements in a subset, a **proper** subset consists of a most n-1 elements

The difference between a subset and a proper subset is that subset A contains everything in the set B, and that all element in set B can be contained in subset A. A is a subset of B

$$A \subseteq B, B \supseteq A$$

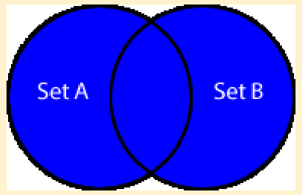
$$\text{e.g } A = \{0, 1, 2, 3\}, B = \{0,\ 1,\ 2,\ 3\}$$



Union

$A\ \cup B$ – A union B (Add together two sets)

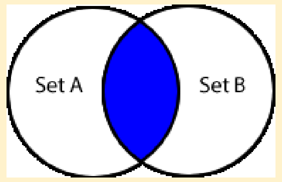
The Union of two sets consists of all elements from both sets. Where values are duplicated a value appears only once in the new set.



$$A = \{1,2,3,4,5,6\}$$
$$B = \{5,6,7,8,9,10\}$$
$$C\ =\ A\ U\ B$$
$$C\ =\ \{1,2,3,4,5,6,7,8,9,10\}$$

Intersection

$A \cap B$ – A intersects B – contains the members that both sets have in common

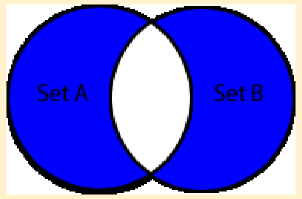


An intersection between 2 sets consists of elements that occur in both sets.

If A={1,2,3,4,5,6} and B={5,6,7,8,9,10}
The intersection between A and B is
{5,6}

Difference

A/B - Difference of set A and B



The difference between the 2 sets consists of elements that occur in one or other of the sets.

If A={1,2,3,4,5,6} and B={5,6,7,8,9,10}
The difference between A and B will be
{1,2,3,4,7,8,9,10}

Regular Expressions

A regular expression is a shorthand way of representing a set. The following characters are applied to the preceding value:

*	Zero or more repetitions
+	One or more repetitions
?	0 or 1
|	alternative
()	Group expression

Examples of regular expressions (The ellipses refer to infinite series)

ab	ab	*ab* only
a*b	b,ab,aab,aaab, ..	Any number of *a* followed by a single *b*
(ab)*	ab, abab, ababab, ..	Zero or more repetitions of *ab*
ab+	ab,abb,abbb,..	A single *a* followed by one or more *b*
a*b*	a,b,ab,aab,abb,..	Zero or more *a* followed by zero or more *b*
a?b	ab, b	Zero or one *a* followed by one *b*
a|b	a,b	*a* or *b*

Finite state machines can be used represent regular expressions. A regular language can be represented by a regular expression or FSM. A FSM recognises whether strings are valid for a language.

e.g. The finite state machine for *a+b* is given as: