INF3580/4580 – Semantic Technologies – Spring 2018

Lecture 5: Mathematical Foundations

Martin Giese

13th February 2018

Department of Informatics

University of Oslo

Mandatory exercises

Remember: Hand-in Oblig 3 by tomorrow.

Oblig 4 published after next lecture.

Today's Plan

1 Basic Set Algebra

2 Pairs and Relations

3 Propositional Logic

Outline

Motivation

- The great thing about Semantic Technologies is...
- ...Semantics!
- ~~“The study of meaning”~~
- RDF has a precisely defined semantics (=meaning)
- Mathematics is best at precise definitions
- RDF has a mathematically defined semantics

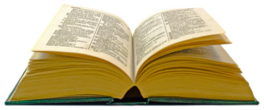


Image ©Colourbox.no

Sets: Cantor’s Definition

From the inventor of Set Theory, Georg Cantor (1845–1918):

> Unter einer „Menge“ verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die „Elemente“ von M genannt werden) zu einem Ganzen.

Translated:

> A ‘set’ is any collection M of definite, distinguishable objects m of our intuition or intellect (called the ‘elements’ of M) to be conceived as a whole.

There are some problems with this, but it’s good enough for us!

Sets

- A set is a mathematical object like a number, a function, etc.
- Knowing a set is
  - knowing what is in it
  - knowing what is not
- Need to know whether elements are equal or not!
- There is no order between elements
- Nothing can be in a set several times
- Two sets A and B are equal if they contain the same elements
  - everything that is in A is also in B
  - everything that is in B is also in A

Elements, Set Equality

Notation for finite sets:

$$\{\text{‘a’}, 1, \triangle\}$$

$$\{\cdots\}$$

Contains ‘a’, 1, and $\triangle$, and nothing else.

There is no order between elements

$$\{1, \triangle\} = \{\triangle, 1\}$$

Nothing can be in a set several times

$$\{1, \triangle, \triangle\} = \{1, \triangle\}$$

Sets with different elements are different:

$$\{1, 2\} \neq \{2, 3\}$$

Element of-relation

We use $\in$ to say that something is element of a set:

$$1 \in \{\text{‘a’}, 1, \triangle\}$$

$$\text{‘b’} \notin \{\text{‘a’}, 1, \triangle\}$$

$$\in$$

$\{3, 7, 12\}$: a set of numbers
- $3 \in \{3, 7, 12\}$, $0 \notin \{3, 7, 12\}$

$\{\text{‘a’}, \text{‘b’}, \ldots, \text{‘z’}\}$: a set of letters
- $\text{‘y’} \in \{\text{‘a’}, \text{‘b’}, \ldots, \text{‘z’}\}$, $\text{‘æ’} \notin \{\text{‘a’}, \text{‘b’}, \ldots, \text{‘z’}\}$,

$\mathbb{N} = \{1, 2, 3, \ldots\}$: the set of all natural numbers
- $3580 \in \mathbb{N}$, $\pi \notin \mathbb{N}$.

$\mathbb{P} = \{2, 3, 5, 7, 11, 13, 17, \ldots\}$: the set of all prime numbers
- $257 \in \mathbb{P}$, $91 \notin \mathbb{P}$.

The set P_{3580} of people in the lecture room right now
- Martin Giese $\in P_{3580}$, Georg Cantor $\notin P_{3580}$.

Sets as Properties

- Sets are used a lot in mathematical notation
- Often, just as a short way of writing things
- More specifically, that something has a property
- E.g. “n is a prime number.”
- In mathematics: $n \in \mathbb{P}$
- E.g. “Martin is a human being.”
- In mathematics, $m \in H$, where
  - H is the set of all human beings
  - m is Martin
- One *could* define $Prime(n)$, $Human(m)$, etc. but that is not usual
- Instead of writing “x has property XYZ” or “$XYZ(x)$”,
  - let P be the set of all objects with property XYZ
  - write $x \in P$.

The Empty Set

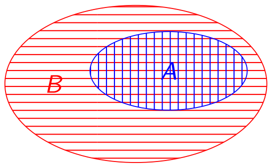
- Sometimes, you need a set that has no elements.
- This is called the *empty set*
- Notation: $\emptyset$ or $\{\}$
- $x \notin \emptyset$, whatever x is!

$$\emptyset$$

Subsets

Let A and B be sets
if every element of A is also in B
then A is called a *subset* of B
This is written

$$A \subseteq B$$



$$\subseteq$$

Examples

- $\{1\} \subseteq \{1, \text{‘a’}, \triangle\}$
- $\{1,3\} \not\subseteq \{1,2\}$
- $\mathbb{P} \subseteq \mathbb{N}$
- $\emptyset \subseteq A$ for any set A

$A = B$ if and only if $A \subseteq B$ and $B \subseteq A$

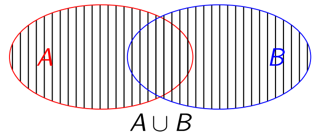
Set Union

The *union* of A and B contains

- all elements of A
- all elements of B
- also those in both A and B
- and nothing more.

It is written

$$A \cup B$$



$$\bigcup$$

(A cup which you pour everything into)

Examples

- $\{1,2\} \cup \{2,3\} = \{1,2,3\}$
- $\{1,3,5,7,9,\ldots\} \cup \{2,4,6,8,10,\ldots\} = \mathbb{N}$
- $\emptyset \cup \{1,2\} = \{1,2\}$

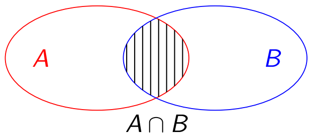
Set Intersection

The *intersection* of A and B contains

- those elements of A
- that are also in B
- and nothing more.

It is written

$$A \cap B$$



$$\bigcap$$

Examples

- $\{1,2\} \cap \{2,3\} = \{2\}$
- $\mathbb{P} \cap \{2,4,6,8,10,\ldots\} = \{2\}$
- $\emptyset \cap \{1,2\} = \emptyset$

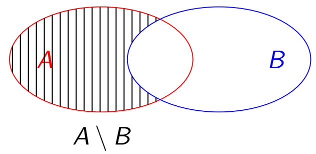
Set Difference

The *set difference* of A and B contains

- those elements of A
- that are *not* in B
- and nothing more.

It is written

$$A \setminus B$$



$$\setminus$$

Examples

- $\{1,2\} \setminus \{2,3\} = \{1\}$
- $\mathbb{N} \setminus \mathbb{P} = \{1,4,6,8,9,10,12,\ldots\}$
- $\emptyset \setminus \{1,2\} = \emptyset$
- $\{1,2\} \setminus \emptyset = \{1,2\}$

Set Comprehensions

- Sometimes enumerating all elements is not good enough
- E.g. there are infinitely many, and "..." is too vague
- Special notation:

$$\{x \in A \mid x \text{ has some property}\}$$

$$\Big\{ \cdots \Big| \cdots \Big\}$$

- The set of those elements of A which have the property.
- Examples:
  - $\{n \in \mathbb{N} \mid n = 2k \text{ for some } k \in \mathbb{N}\}$: the even numbers
  - $\{n \in \mathbb{N} \mid n < 5\} = \{1, 2, 3, 4\}$
  - $\{x \in A \mid x \notin B\} = A \setminus B$

Question

- The *symmetric difference* $A \triangle B$ of two sets contains
  - All elements that are in A or B...
  - ...but not in both.
- Can you write $A \triangle B$ using $\cap$, $\cup$, $\setminus$?

$$A \triangle B = (A \cup B) \setminus (A \cap B)$$

- Or:

$$A \triangle B = (A \setminus B) \cup (B \setminus A)$$

Outline

Motivation

- RDF is all about
  - Resources (objects)
  - Their properties (`rdf:type`)
  - Their relations amongst each other
- Sets are good to group objects with some properties!
- How do we talk about relations between objects?

Pairs

A pair is an *ordered* collection of two objects

Written

$$\langle x, y \rangle \qquad\qquad \Big\langle \cdot\ \cdot\ \cdot \Big\rangle$$

Equal if components are equal:

$$\langle a, b \rangle = \langle x, y \rangle \quad \text{if and only if} \quad a = x \quad \text{and} \quad b = y$$

Order matters:

$$\langle 1, \text{'a'} \rangle \neq \langle \text{'a'}, 1 \rangle$$

An object can be twice in a pair:

$$\langle 1, 1 \rangle$$

$\langle x, y \rangle$ is a pair, no matter if $x = y$ or not.

The Cross Product

Let A and B be sets.

Construct the set of all pairs $\langle a, b \rangle$ with $a \in A$ and $b \in B$.

This is called the *cross product* of A and B, written

$$A \times B \qquad\qquad \times$$

Example:

- $A = \{1, 2, 3\}$, $B = \{\text{'a'}, \text{'b'}\}$.
- $A \times B = \{ \ \langle 1, \text{'a'} \rangle, \ \langle 2, \text{'a'} \rangle, \ \langle 3, \text{'a'} \rangle, \ \langle 1, \text{'b'} \rangle, \ \langle 2, \text{'b'} \rangle, \ \langle 3, \text{'b'} \rangle \ \}$

Why bother?

Instead of "$\langle a, b \rangle$ is a pair of a natural number and a person in this room"...

...$\langle a, b \rangle \in \mathbb{N} \times P_{3580}$

But most of all, there are subsets of cross products...

Relations

A *relation* R between two sets A and B is...

...a set of pairs $\langle a, b \rangle \in A \times B$

$$R \subseteq A \times B$$

We often write aRb to say that $\langle a, b \rangle \in R$

Example:

- Let $L = \{\text{'a'}, \text{'b'}, \ldots, \text{'z'}\}$
- Let $\triangleright$ relate each number between 1 and 26 to the corresponding letter in the alphabet:

$$1 \triangleright \text{'a'} \qquad 2 \triangleright \text{'b'} \qquad \ldots \qquad 26 \triangleright \text{'z'}$$

- Then $\triangleright \subseteq \mathbb{N} \times L$:

$$\triangleright = \{\langle 1, \text{'a'} \rangle, \langle 2, \text{'b'} \rangle, \ldots, \langle 26, \text{'z'} \rangle\}$$

- And we can write:

$$\langle 1, \text{'a'} \rangle \in \triangleright \qquad \langle 2, \text{'b'} \rangle \in \triangleright \qquad \ldots \qquad \langle 26, \text{'z'} \rangle \in \triangleright$$

More Relations

A relation R *on* some set A is a relation between A and A:

$$R \subseteq A \times A = A^2$$

Example: $<$

- Consider the $<$ order on natural numbers:

$$1 < 2 \qquad 1 < 3 \qquad 1 < 4 \qquad \ldots \qquad 2 < 3 \qquad 2 < 4 \qquad \ldots$$

- $< \subseteq \mathbb{N} \times \mathbb{N}$:

$$< \ = \ \{ \ \langle 1, 2 \rangle, \ \langle 1, 3 \rangle, \ \langle 1, 4 \rangle, \ \ldots, \ \langle 2, 3 \rangle, \ \langle 2, 4 \rangle, \ \ldots, \ \langle 3, 4 \rangle, \ \ldots \ \ldots \ \}$$

- $< \ = \{\langle x, y \rangle \in \mathbb{N}^2 \mid x \text{ is less than } y\}$

Family Relations

Consider the set $S = \{\text{Homer}, \text{Marge}, \text{Bart}, \text{Lisa}, \text{Maggie}\}$.
Define a relation P on S such that

$$x\,P\,y \quad \text{iff} \quad x \text{ is parent of } y$$

For instance:

$$\text{Homer}\,P\,\text{Bart} \qquad \text{Marge}\,P\,\text{Maggie}$$

As a set of pairs:

$$P = \{ \langle \text{Homer}, \text{Bart} \rangle, \langle \text{Homer}, \text{Lisa} \rangle, \langle \text{Homer}, \text{Maggie} \rangle, \langle \text{Marge}, \text{Bart} \rangle, \langle \text{Marge}, \text{Lisa} \rangle, \langle \text{Marge}, \text{Maggie} \rangle \} \subseteq S^2$$

For instance:

$$\langle \text{Homer}, \text{Bart} \rangle \in P \qquad \langle \text{Marge}, \text{Maggie} \rangle \in P$$

Set operatrions on relations

Since relations are just sets of pairs, we can use set operations and relations on them.
We say that R is a subrelation P if $R \subseteq P$.
E.g.: if F is the father-of-relation,

$$F = \{ \langle \text{Homer}, \text{Bart} \rangle, \langle \text{Homer}, \text{Lisa} \rangle, \langle \text{Homer}, \text{Maggie} \rangle \}$$

then $F \subseteq P$.
If M is the mother-of-relation,

$$M = \{ \langle \text{Marge}, \text{Bart} \rangle, \langle \text{Marge}, \text{Lisa} \rangle, \langle \text{Marge}, \text{Maggie} \rangle \}$$

then $F \cup M = P$.

Special Kinds of Relations

Certain properties of relations occur in many applications

Therefore, they are given names

$R \subseteq A^2$ is *reflexive*

- $x\,R\,x$ for all $x \in A$.
- E.g. "$=$", "$\leq$" in mathematics, "has same color as", etc.

$R \subseteq A^2$ is *symmetric*

- If $x\,R\,y$ then $y\,R\,x$.
- E.g. "$=$" in mathematics, friendship in facebook, connected by rail, etc.

$R \subseteq A^2$ is *transitive*

- If $x\,R\,y$ and $y\,R\,z$, then $x\,R\,z$
- E.g. "$=$", "$\leq$", "$<$" in mathematics, "is ancestor of", etc.

Question

Let $A = \{1, 2\}$, a set of two elements.
How many different relations on A are there?

$$A \times A = \{ \langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 2, 1 \rangle, \langle 2, 2 \rangle \}$$

A relation on A is a subset of $A \times A$. So how many subsets are there?

$$\{\}, \quad \{\langle 1, 1 \rangle\}, \quad \{\langle 1, 2 \rangle\}, \quad \{\langle 1, 1 \rangle, \langle 1, 2 \rangle\}, \ldots$$

16 relations on A. Generally: $2^{(|A|^2)}$

Outline

Many Kinds of Logic

- In mathematical logic, many kinds of logic are considered
  - propositional logic (and, or, not)
  - description logic (a mother is a person who is female and has a child)
  - modal logic (Alice knows that Bob didn't know yesterday that...)
  - first-order logic (For all..., for some...)
- All of them formalizing different aspects of reasoning
- All of them defined mathematically
  - Syntax ($\approx$ grammar. What is a formula?)
  - Semantics (What is the meaning?)
    - proof theory: what is legal reasoning?
    - model semantics: declarative using set theory.
- For semantic technologies, description logic (DL) is most interesting
  - talks about sets and relations
- Basic concepts can be explained using predicate logic

Propositional Logic: Formulas

Formulas are defined "by induction" or "recursively":

1. Any letter p, q, r,...is a formula
2. *if* A and B are formulas, *then*
   - $(A \land B)$ is also a formula (read: "A and B")
   - $(A \lor B)$ is also a formula (read: "A or B")
   - $(A \to B)$ is also a formula (read "A implies B")
   - $\neg A$ is also a formula (read: "not A")

$$\land \lor \to \neg$$

Nothing else is. Only what rules [1] and [2] say is a formula.

Examples of formulae:

$$p \quad (p \land \neg r) \quad (q \land q) \quad (q \land \neg q) \quad ((p \lor \neg q) \land (\neg p \to q))$$

Examples of non-formulas:

$$pqr \quad p\neg q \quad \land (p$$

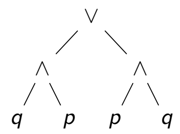
Propositional Formulas, Using Sets

The set of all formulas Φ is the least set such that

1. All letters $p, q, r, \ldots \in \Phi$
2. *if* $A, B \in \Phi$, then
   - $(A \land B) \in \Phi$
   - $(A \lor B) \in \Phi$
   - $(A \to B) \in \Phi$
   - $\neg A \in \Phi$

- Formulas are just a kind of strings until now:
  - no meaning
  - but every formula can be "parsed" uniquely.

$$((q \land p) \lor (p \land q))$$



Terminology

$\neg, \wedge, \vee, \rightarrow$ are called *connectives*.

A formula $(A \wedge B)$ is called a *conjunction*.

A formula $(A \vee B)$ is called a *disjunction*.

A formula $(A \rightarrow B)$ is calles an *implication*.

A formula $\neg A$ is called a *negation*.

Truth

Logic is about things being true or false, right?

Is $(p \wedge q)$ true?

That depends on whether p and q are true!

If p is true, and q is true, then $(p \wedge q)$ is true

Otherwise, $(p \wedge q)$ is false.

So truth of a formula depends on the truth of the letters

We also say the "interpretation" of the letters

In other words, in general, truth depends on the context

Let's formalize this context, a.k.a. interpretation, a.k.a. model

Interpretations

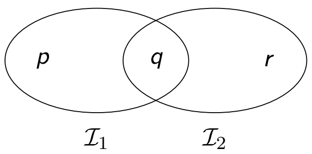
Idea: put all letters that are "true" into a set!

Define: An *interpretation* $\mathcal{I}$ is a set of letters.

Letter p is true in interpretation $\mathcal{I}$ if $p \in \mathcal{I}$.

E.g., in $\mathcal{I}_1 = \{p, q\}$, p is true, but r is false.

But in $\mathcal{I}_2 = \{q, r\}$, p is false, but r is true.



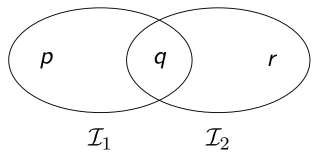
Semantic Validity $\models$

To say that p is true in $\mathcal{I}$, write

$$\mathcal{I} \models p$$

$\models$

For instance



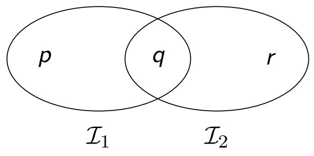
$$\mathcal{I}_1 \models p \qquad \mathcal{I}_2 \not\models p$$

In other words, for all letters p:

$$\mathcal{I} \models p \qquad \text{if and only if} \qquad p \in \mathcal{I}$$

Validity of Compound Formulas

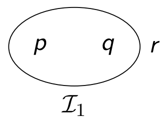
So, is $(p \wedge q)$ true?
That depends on whether p and q are true!
And that depends on the interpretation.
All right then, *given some* $\mathcal{I}$, is $(p \wedge q)$ true?
Yes, if $\mathcal{I} \models p$ *and* $\mathcal{I} \models q$
No, otherwise
For instance

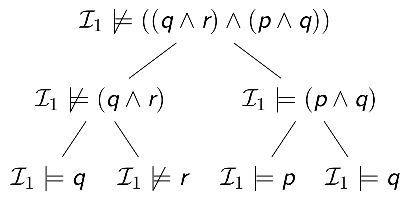


$\mathcal{I}_1 \models p \wedge q$ $\quad$ $\mathcal{I}_2 \not\models p \wedge q$

Validity of Compound Formulas, cont.

That was easy, p and q are only letters...
...so, is $((q \wedge r) \wedge (p \wedge q))$ true in $\mathcal{I}$?
Idea: apply our rule recursively
For any formulas A and B,...
...and any interpretation $\mathcal{I}$,...
...$\mathcal{I} \models A \wedge B$ if and only if $\mathcal{I} \models A$ and $\mathcal{I} \models B$
For instance, if $\mathcal{I}_1 = \{p, q\}$:





Semantics for $\neg$, $\rightarrow$ and $\vee$

The complete definition of $\models$ is as follows:
For any interpretation $\mathcal{I}$, letter p, formulas A, B:

- $\mathcal{I} \models p$ iff $p \in \mathcal{I}$
- $\mathcal{I} \models \neg A$ iff $\mathcal{I} \not\models A$
- $\mathcal{I} \models (A \wedge B)$ iff $\mathcal{I} \models A$ and $\mathcal{I} \models B$
- $\mathcal{I} \models (A \vee B)$ iff $\mathcal{I} \models A$ or $\mathcal{I} \models B$ (or both)
- $\mathcal{I} \models (A \rightarrow B)$ iff $\mathcal{I} \models A$ implies $\mathcal{I} \models B$

Semantics of $\neg$, $\wedge$, $\vee$, $\rightarrow$ often given as *truth table*:

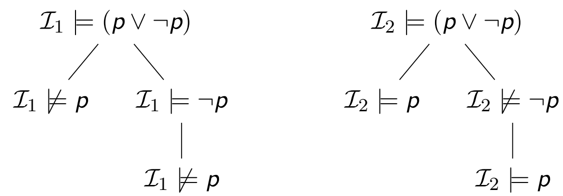
A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \rightarrow B$
f	*f*	*t*	*f*	*f*	*t*
f	*t*	*t*	*f*	*t*	*t*
t	*f*	*f*	*f*	*t*	*f*
t	*t*	*f*	*t*	*t*	*t*

Some Formulas Are Truer Than Others

Is $(p \vee \neg p)$ true?
Only two interesting interpretations:

$\mathcal{I}_1 = \emptyset$ $\qquad$ $\mathcal{I}_2 = \{p\}$

Recursive Evaluation:



$(p \vee \neg p)$ is true in *all* interpretations!

Tautologies

A formula A that is true in *all* interpretations is called a *tautology*

also *logically valid*

also a *theorem* (of propositional logic)

written:

$$\models A$$

$(p \vee \neg p)$ is a tautology

True whatever p means:

- The sky is blue or the sky is not blue.
- Marit B. will win the race or Marit B. will not win the race.
- The slithy toves gyre or the slithy toves do not gyre.

Possible to derive true statements mechanically...

...without understanding their meaning!

Checking Tautologies

Checking whether $\models A$ is the task of SAT-solving

(co-)NP-complete in general (i.e. in practice exponential time)

Small instances can be checked with a truth table:

$$\models (\neg p \vee (\neg q \vee (p \wedge q))) \quad ?$$

p	q	$\neg p$	$\neg q$	$(p \wedge q)$	$(\neg q \vee (p \wedge q))$	$(\neg p \vee (\neg q \vee (p \wedge q)))$
f	*f*	*t*	*t*	*f*	*t*	*t*
f	*t*	*t*	*f*	*f*	*f*	*t*
t	*f*	*f*	*t*	*f*	*t*	*t*
t	*t*	*f*	*f*	*t*	*t*	*t*

Therefore: $(\neg p \vee (\neg q \vee (p \wedge q)))$ is a tautology!

Entailment

Tautologies are true in all interpretations

Some Formulas are true only under certain assumptions

A entails B, written $A \models B$ if

- $\mathcal{I} \models B$
- for all interpretations $\mathcal{I}$ with $\mathcal{I} \models A$

Also: "B is a logical consequence of A"

Whenever A holds, also B holds

For instance:

$$p \wedge q \models p$$

Independent of meaning of p and q:

- If it rains and the sky is blue, then it rains
- If M.B. wins the race and the world ends, then M.B. wins the race
- If 'tis brillig and the slythy toves do gyre, then 'tis brillig

Checking Entailment

SAT solvers can be used to check entailment:

$$A \models B \qquad \text{if and only if} \qquad \models (A \to B)$$

We can check simple cases with a truth table:

$$(p \wedge \neg q) \models \neg(\neg p \vee q) \quad ?$$

p	q	$\neg p$	$\neg q$	$(p \wedge \neg q)$	$(\neg p \vee q)$	$\neg(\neg p \vee q)$
f	*f*	*t*	*t*	*f*	*t*	*f*
f	*t*	*t*	*f*	*f*	*t*	*f*
t	*f*	*f*	*t*	*t*	*f*	*t*
t	*t*	*f*	*f*	*f*	*t*	*f*

So $(p \wedge \neg q) \models \neg(\neg p \vee q)$

And $\neg(\neg p \vee q) \models (p \wedge \neg q)$

Equivalent formulas and redundant connectives

- In other words, $(p \wedge \neg q)$ and $\neg(\neg p \vee q)$ always have the same truth value, no matter the interpretation.
- We say that A and B are *equivalent* if A and B always have the same truth value.
- For this we often introduce another connective, $\leftrightarrow$.
- $\mathcal{I} \models (A \leftrightarrow B)$ iff $\mathcal{I} \models A$ if and only if $\mathcal{I} \models B$.
- To express that two formulas A, B are equivalent, we can write $\models (A \leftrightarrow B)$.

- We actually only need a subset of the connectives:
- E.g.:
  - $\models ((A \vee B) \leftrightarrow \neg(\neg A \wedge \neg B))$.
  - $\models ((A \rightarrow B) \leftrightarrow (\neg A \vee B))$.
  - $\models ((A \leftrightarrow B) \leftrightarrow ((A \rightarrow B) \wedge (B \rightarrow A)))$.
- So we actually only need $\neg$ and $\wedge$ to express any formula!
- Any formula is equivalent to a formula containing only the connectives $\neg$ and $\wedge$.

Recap

- Sets
  - are collections of objects without order or multiplicity
  - often used to gather objects which have some property
  - can be combined using $\cap, \cup, \setminus$
- Relations
  - are sets of pairs (subset of cross product $A \times B$)
  - $x\, R\, y$ is the same as $\langle x, y \rangle \in R$
  - can use set operations on relations, e.g. $F \subseteq P$.
- Predicate Logic
  - has formulas built from letters, $\wedge$, $\vee$, $\rightarrow$, $\neg$ *(syntax)*
  - which can be evaluated in an *interpretation (semantics)*
  - interpretations are sets of letters
  - recursive definition for semantics of $\wedge$, $\vee$, $\rightarrow$, $\neg$
  - $\models A$ if $\mathcal{I} \models A$ for all $\mathcal{I}$ *(tautology)*
  - $A \models B$ if $\mathcal{I} \models B$ for all $\mathcal{I}$ with $\mathcal{I} \models A$ *(entailment)*
  - truth tables can be used for checking validity and etailment.