Ontology Engineering for the Semantic Web

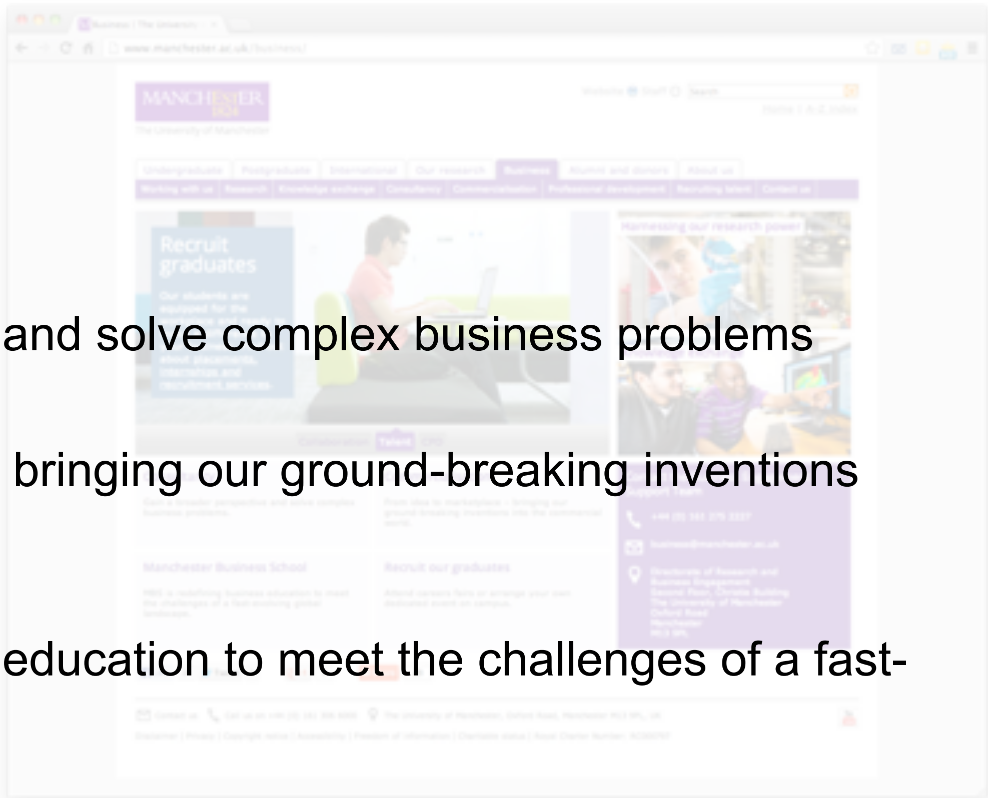
COMP62342

Sean Bechhofer and Uli Sattler
University of Manchester
sean.bechhofer@manchester.ac.uk
ulrike.sattler@manchester.ac.uk

What's the Problem?

- Typical web page markup consists of:
  - Rendering information (e.g., font size and colour)
  - Hyper-links to related content
- Semantic content is accessible to humans but not (easily) to computers…

Information we can see

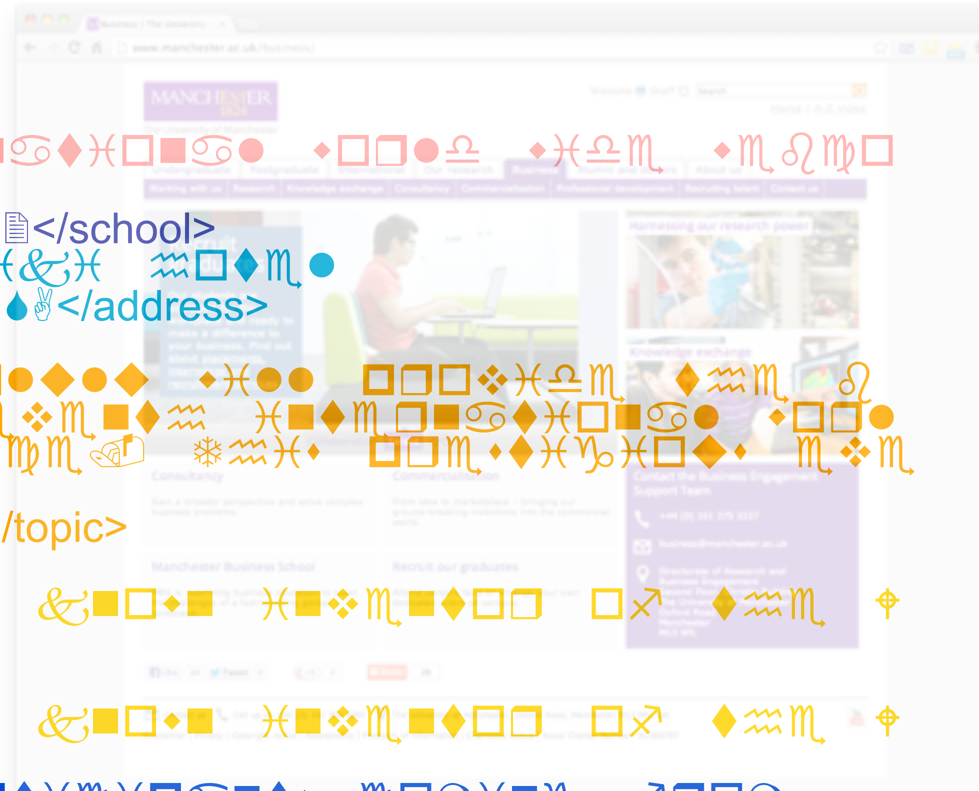


- University of Manchester
  - The Business School
- Consultancy
  - Gain a broader perspective and solve complex business problems
- Commercialisation
  - From idea to marketplace -- bringing our ground-breaking inventions into the commercial world
- Manchester Business School
  - MBS is redefining business education to meet the challenges of a fast-evolving global landscape
- Recruit our graduates
  - Attend careers fairs or arrange your own dedicated event on campus
- Contact the Business Engagement Support Team
  - +44 161 275 2227
  - business@manchester.ac.uk
- ....

Information a machine can see…

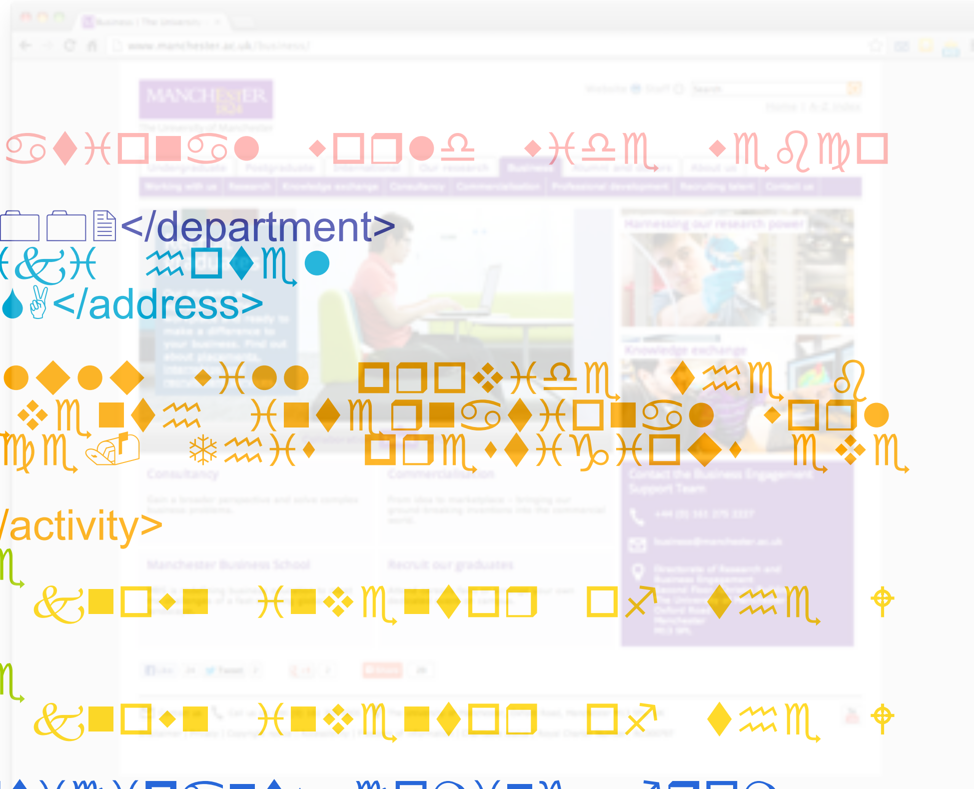
⌖⌖⌖📄📁📁📄
❄♒♏ ♏●♏❖♏■⧫♒ ♓■⧫♏❒■♋⧫♓□■♋● ⬥□❒●♎ ⬥♓♎♏ ⬥♏♌
♍□■
💧♒♏❒♋⧫□■ ⬥♋♓🙰♓🙰♓ ♒□⧫♏●
🏱□■□●◆●◆📪 ♒♋⬥♋♓♓📪 ✞💧✌
🖮📫📂📂 ❍♋⍓ 📄📁📁📄
📂 ●□♍♋⧫♓□■ 🗄 ♎♋⍓⬧ ●♏♋❒■ ♓■⧫♏❒♋♍⧫
☼♏♑♓⬧⧫♏❒♏♎ ◻♋❒⧫♓♍♓◻♋■⧫⬧ ♍□❍♓■♑ ♐❒□❍
♋◆⬧⧫❒♋●♓♋📪 ♍♋■♋♎♋📪 ♍♒♓●♏ ♎♏■❍♋❒🙰📪 ♐❒♋■
♍♏📪 ♑♏❒❍♋■⍓📪 ♑♒♋■♋📪 ♒□■♑ 🙰□■♑📪 ♓■♎♓♋
📪 ♓❒♏●♋■♎📪 ♓⧫♋●⍓📪 ⛎♋◻♋■📪 ❍♋●⧫♋📪 ■♏⬥ ⌘
♏♋●♋■♎📪 ⧫♒♏ ■♏⧫♒♏❒●♋■♎⬧📪 ■□❒⬥♋⍓📪 ⬧♓■♑♋◻□❒
♏📪 ⬧⬥♓⧫⌘♏❒●♋■♎📪 ⧫♒♏ ◆■♓⧫♏♎ 🙰♓■♑♎□❍📪 ⧫♒♏
◆■♓⧫♏♎ ⬧⧫♋⧫♏⬧📪 ❖♓♏⧫■♋❍📪 ⌘♋♓❒♏
☼♏♑♓⬧⧫♏❒ ■□⬥
⚐■ ⧫♒♏ 🖮⧫♒ 💣♋⍓ 🏱□■□●◆●◆ ⬥♓●● ◻❒□❖♓♎♏ ⧫♒♏
♌♋♍🙰♎❒□◻ □♐ ⧫♒♏ ♏●♏❖♏■⧫♒ ♓■⧫♏❒■♋⧫♓□■♋● ⬥
□❒●♎ ⬥♓♎♏ ⬥♏♌ ♍□■♐♏❒♏■♍♏📬 ❄♒♓⬧ ◻❒♏⬧⧫♓♑♓□
◆⬧ ♏❖♏■⧫ ⑤
💧◻♏♋🙰♏❒⬧ ♍□■♐♓❒❍♏♎
❄♓❍ ♌♏❒■♏❒⬧📫●♏♏
❄♓❍ ♓⬧ ⧫♒♏ ⬥♏●● 🙰■□⬥■ ♓■❖♏■⧫□❒ □♐ ⧫♒♏ 🕈♏♌
📪…

Solution: XML markup with "meaningful" tags?



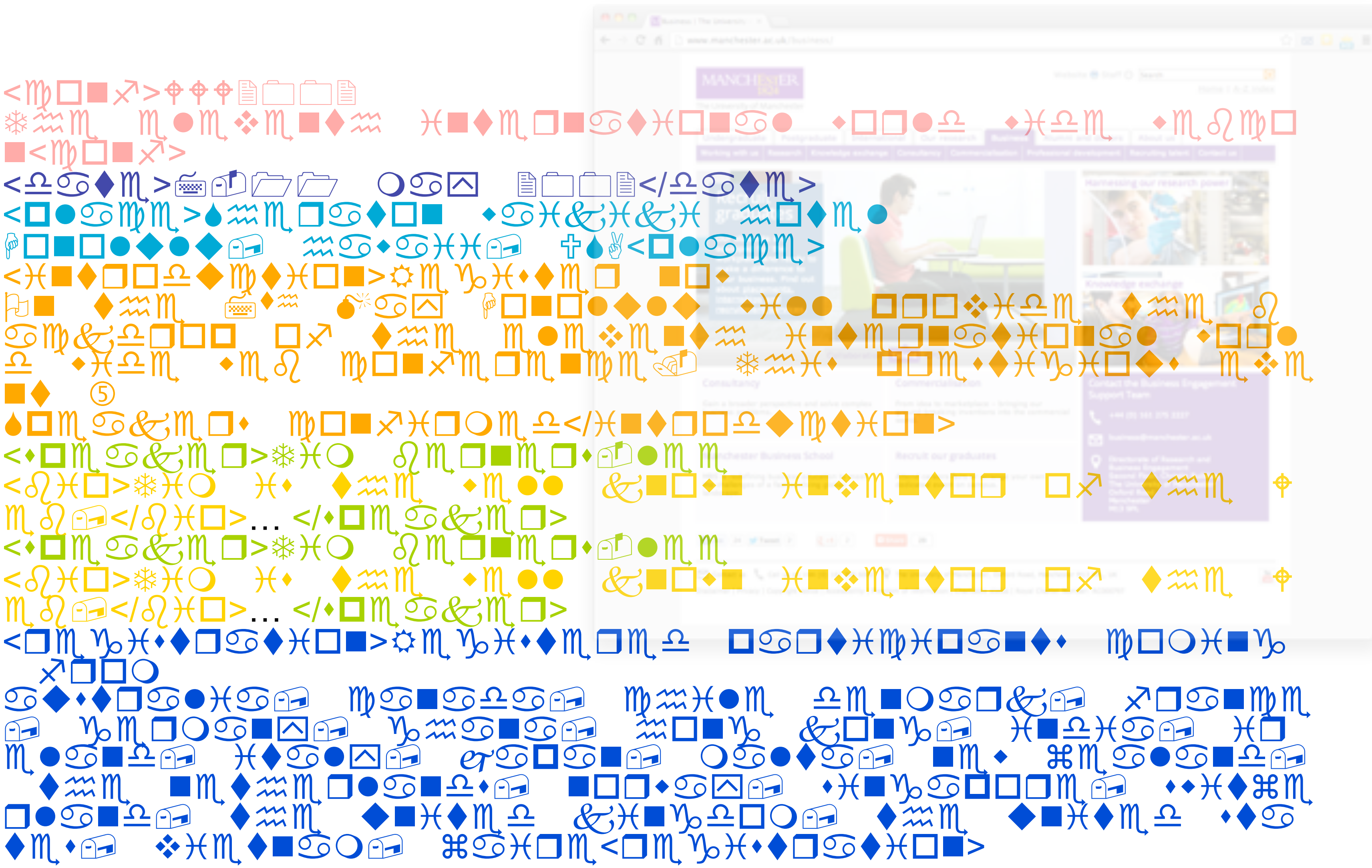
<university> </university>
<school> </school>
<address> </address>
<topic> </topic>
<topic> <details> </details>… </topic>
<topic> <details> </details>… </topic>
<contact> <contact>

But what about....?



<university>⌖⌖⌖🗏🗀🗀🗏
❄♒♏ ♏●♏❖♏■⧫♒ ♓■⧫♏❒■♋⧫♓□■♋● ⬧□❒●♎ ⬧♓♎♏ ⬧♏♌♍□
■</university>
<department>🖮🖂🗁🗁 ❍♋⌧ 🗏🗀🗀🗏</department>
<address>💧♒♏❑♋⧫□■ ⬧♋♓🙰♓🙰♓ ♒□⧫♏●
☞□■□●◆●◆📪 ♒♋⬧♋♓♓📪 ✞💧✌</address>
<activity>☼♏♑♓⬧⧫♏❑ ■□⬧
🏱■ ⧫♒♏ 🖮⧫♒ 💣♋⌧ ☞□■□●◆●◆ ⬧♓●● ❑❑❑❖♓♎♏ ⧫♒♏ ♌
♋♍🙰♎❑□□ □⌧ ⧫♒♏ ♏●♏❖♏■⧫♒ ♓■⧫♏❑■♋⧫♓□■♋● ⬧□❑●
♎ ⬧♓♎♏ ⬧♏♌ ♍□■⌧♏❑♏■♍♏📬 ❄♒♓⬧ ❑❑♏⬧⧫♓♑♓□◆⬧ ♏❖♏
■⧫ ⑤
💧□♏♋🙰♏❑⬧ ♍□■⌧♓❑❍♏♎</activity>
<activity>❄♓❍ ♌♏❑■♏❑⬧📂●♏♏
<details>❄♓❍ ♓⬧ ⧫♒♏ ⬧♏●● 🙵■□⬥■ ♓■❖♏■⧫□❑ □⌧ ⧫♒♏ ⌖
♏♌📪</details>… </activity>
<activity>❄♓❍ ♌♏❑■♏❑⬧📂●♏♏
<details>❄♓❍ ♓⬧ ⧫♒♏ ⬧♏●● 🙵■□⬥■ ♓■❖♏■⧫□❑ □⌧ ⧫♒♏ ⌖
♏♌📪</details>… </activity>
<contact>☼♏♑♓⬧⧫♏❑ ❑♋❑⧫♓♍♓□♋■⧫⬧ ♍□❍♓■♑ ⌧❑□❍
♋◆⬧⧫❑♋●♓♋📪 ♍♋■♋♎♋📪 ♍♒♓●♏ ♎♏■❍♋❑🙰📪 ⌧❑♋■♍♏
📪 ♑♏❑❍♋■⌧📪 ♑♒♋■♋📪 ♒□■♑ 🙰□■♑📪 ♓■♎♓♋📪 ♓□
♏●♋■♎📪 ♓⧫♋●⌧📪 ♐♋□♋■📪 ❍♋●⧫♋📪 ■♏⬧ ⌘♏♋●♋■♎📪
⧫♒♏ ■♏⧫♒♏❑●♋■♎⬧📪 ■□❑⬧♋⌧📪 ⬧♓■♑♋❑□❑♏📪 ⬧⬧♓⧫⌘♏
❑●♋■♎📪 ⧫♒♏ ◆■♓⧫♏♎ 🙰♓■♑♎□❍📪 ⧫♒♏ ◆■♓⧫♏♎ ⬧⧫♋
⧫♏⬧📪 ❖♓♏⧫■♋❍📪 ⌘♋♓❑♏<contact>

Still the Machine only sees…



MANCHESTER 1824

The University of Manchester

Need to Add "Semantics"

- External agreement on meaning of annotations
  - E.g., *Dublin Core* for annotation of library/bibliographic information
    - Agree on the meaning of a set of annotation tags
  - Problems with this approach
    - Infle
    - Limit
- Use Vocabu nnotations
  - Ontologi
  - New terr es
    - "Conceptual Lego"
  - Meaning (semantics) of such terms is formally specified

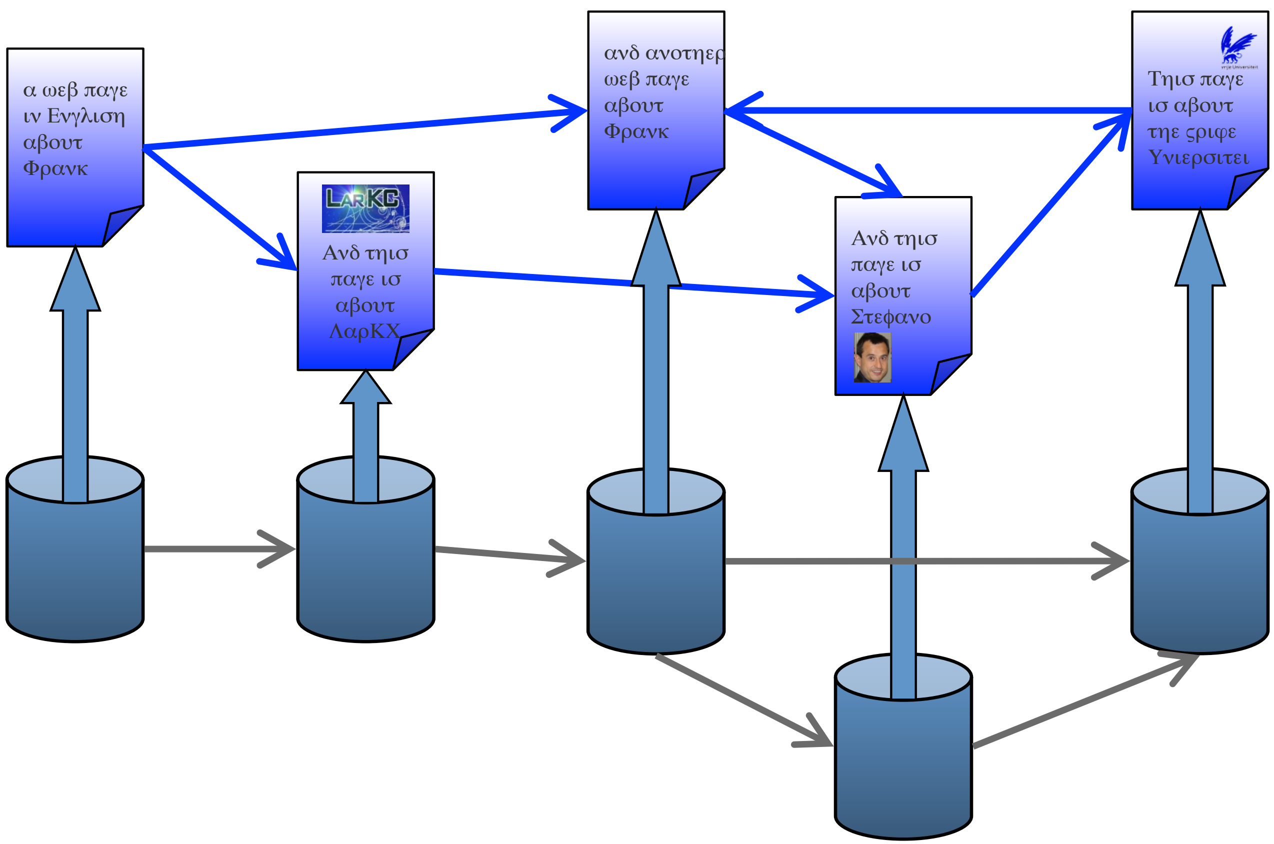
Machine **Processable**

not

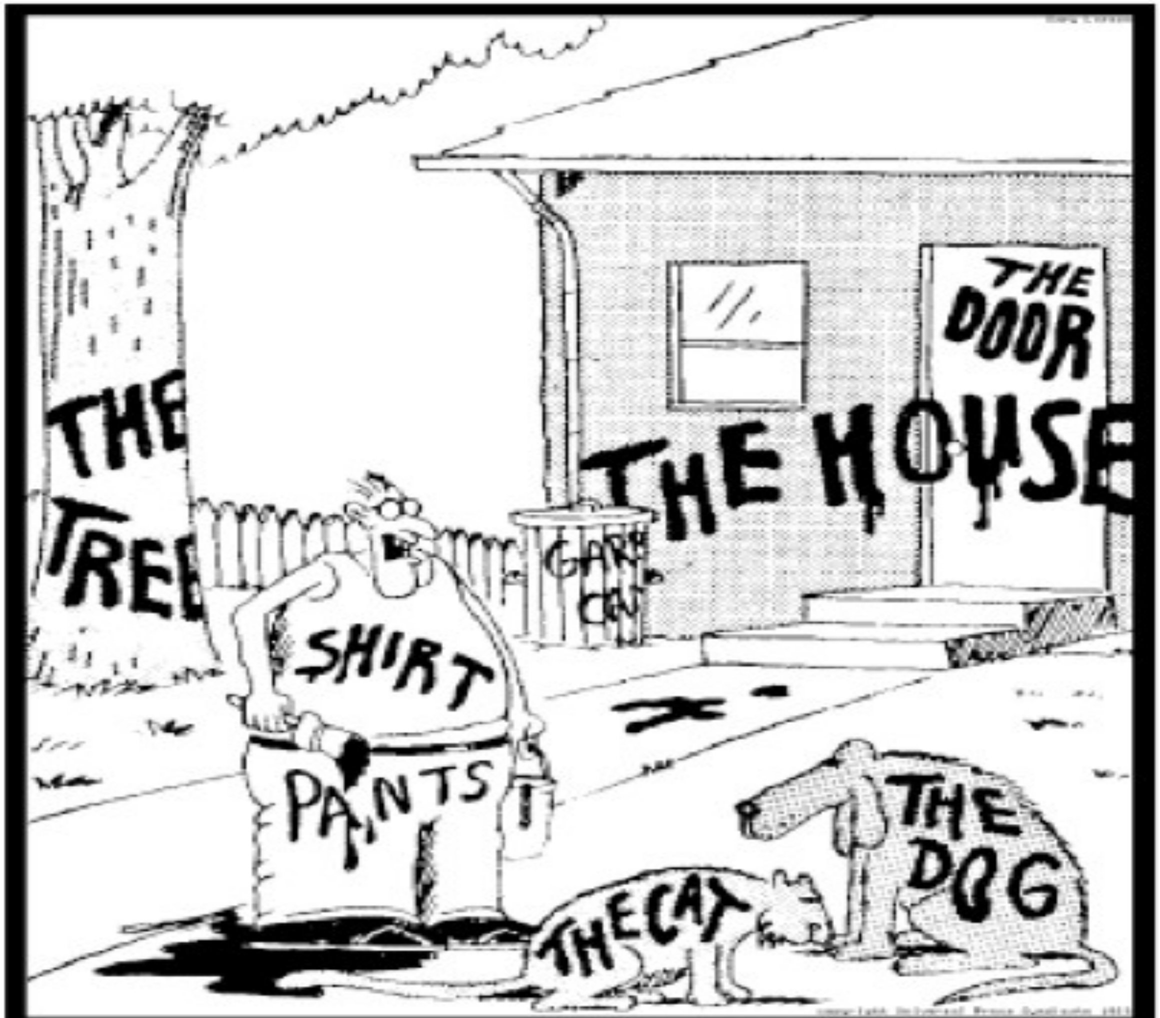
Machine **Understandable**

Four principles towards a Semantic Web of Data*

* With thanks to Frank van Harmelen

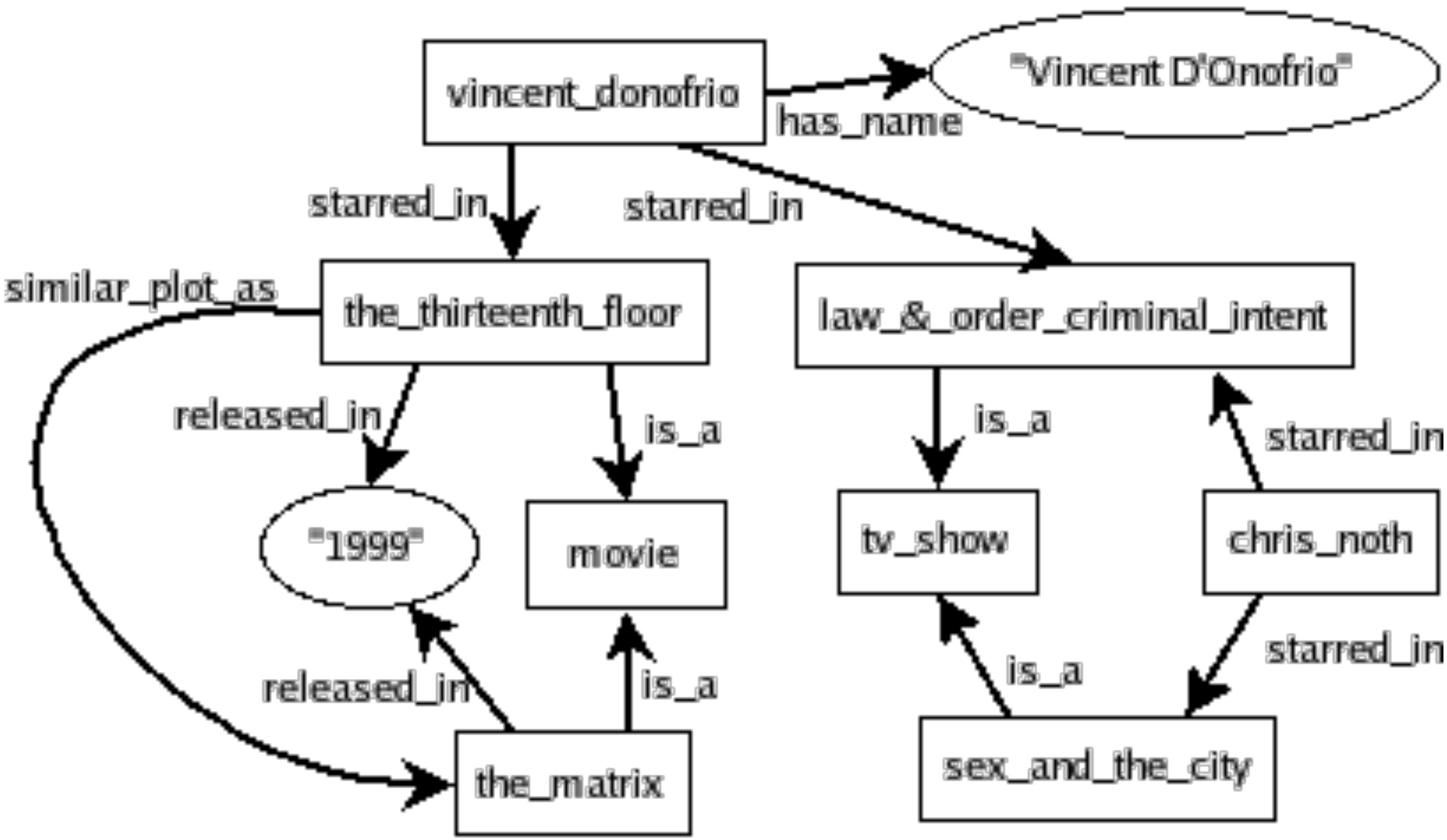


P1: Give all things a name

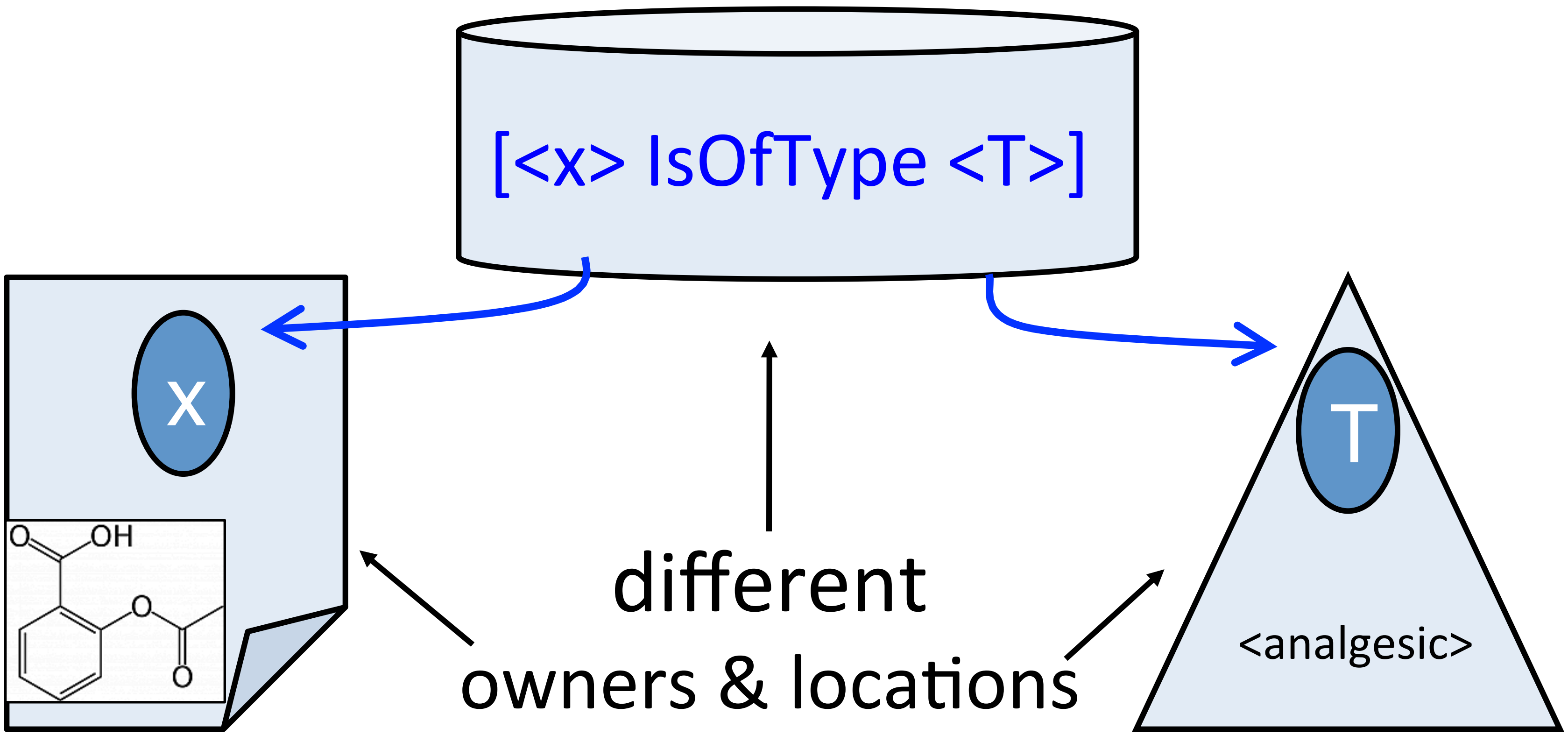


"Now! *That* should clear up a few things around here!"

P2: Relationships form a graph between things



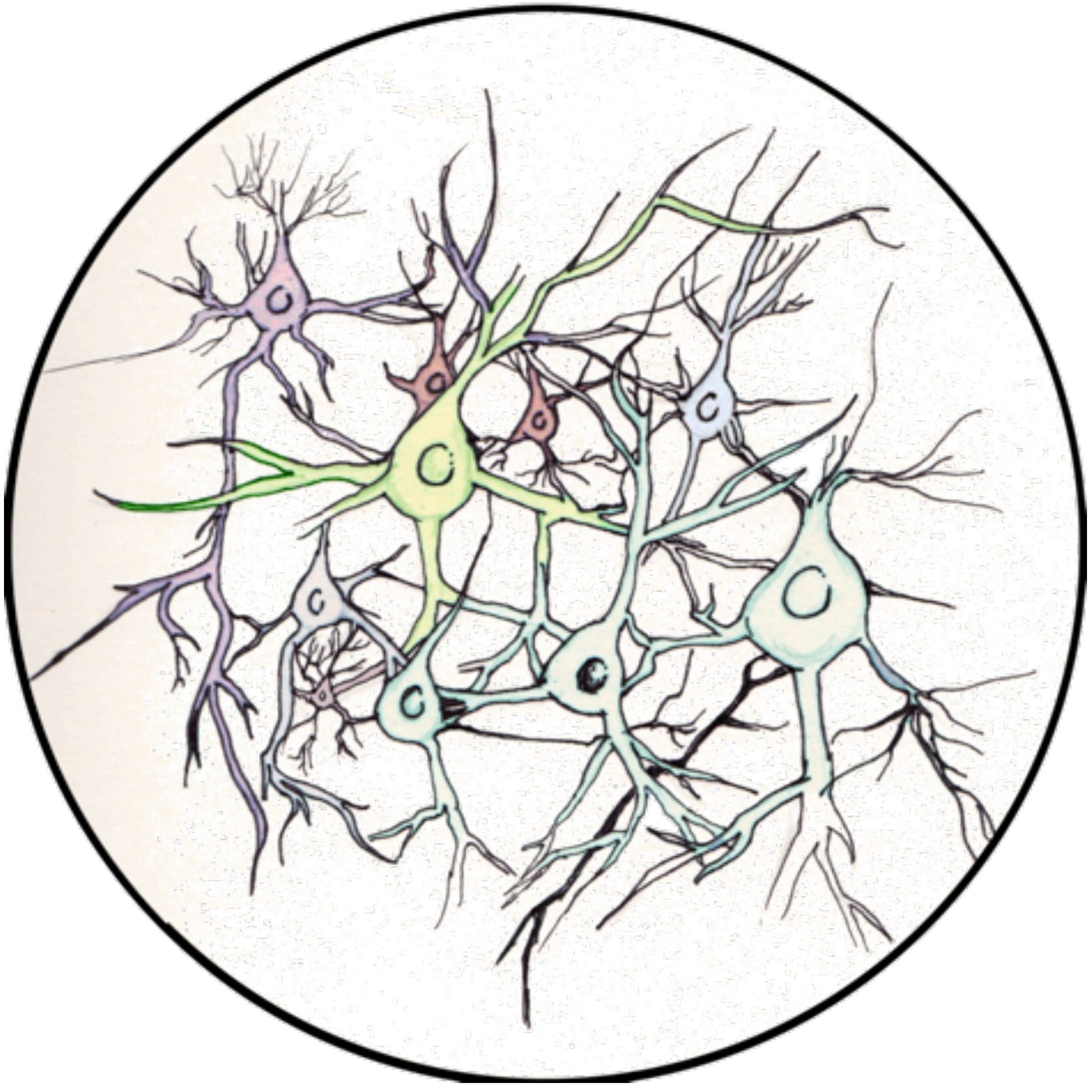
P3: The names are addresses on the Web



MANCHESTER
1824

The University of Manchester

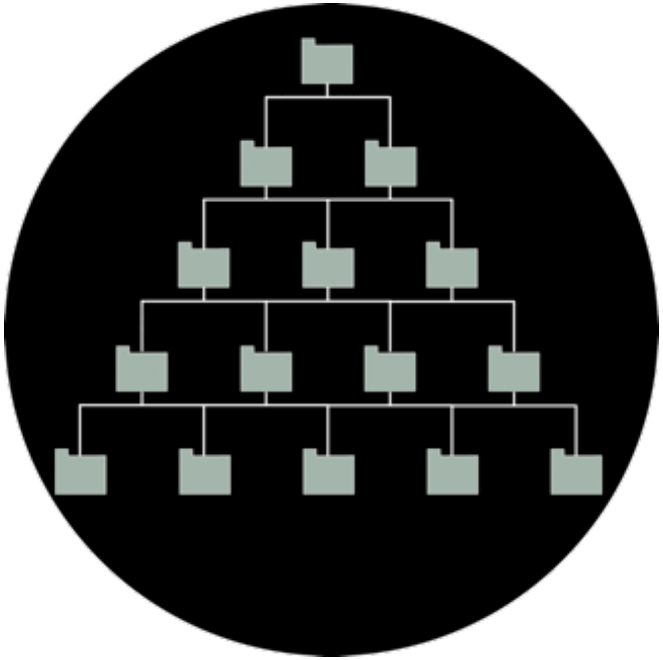
P1 + P2 + P3 = Giant Global Graph



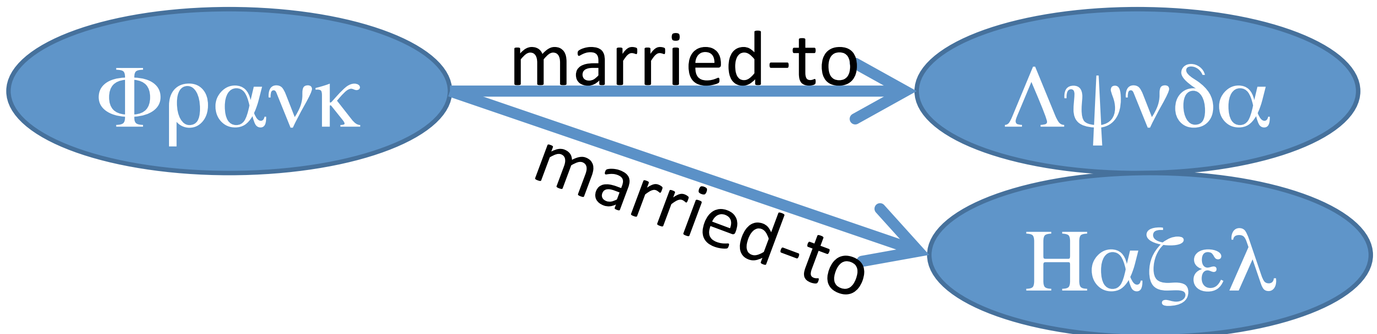
P4: Explicit, Formal Semantics

- Assign Types to Things
- Assign Types to Relations
- Organise Types in a Hierarchy
- Impose Constraints on Possible Interpretations

This is where we will spend most of our time on this course unit -- looking at the ontologies that provide this semantics



Semantics



- Φρανκ is male
- married-to relates males to females

lowerbound

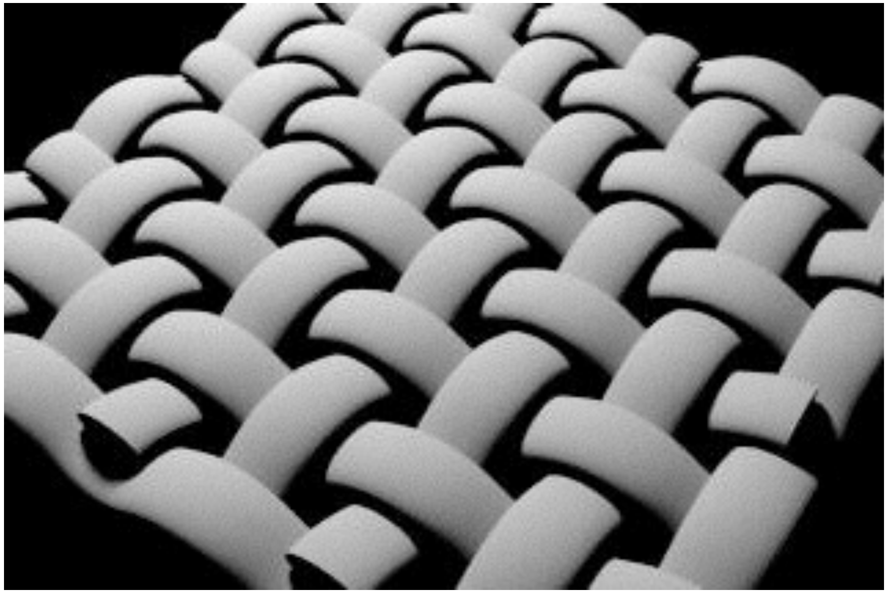
- married-to relates 1 male to 1 female
- Λψνδα = Ηαζελ

upperbound

Semantics = predictable inference

KR: Cloth Weaves
[Maier & Warren, Computing with Logic, 1988]

- An example showing how we can represent the qualities and characteristics of cloth types using a simple propositional logic knowledge base.



Cloth

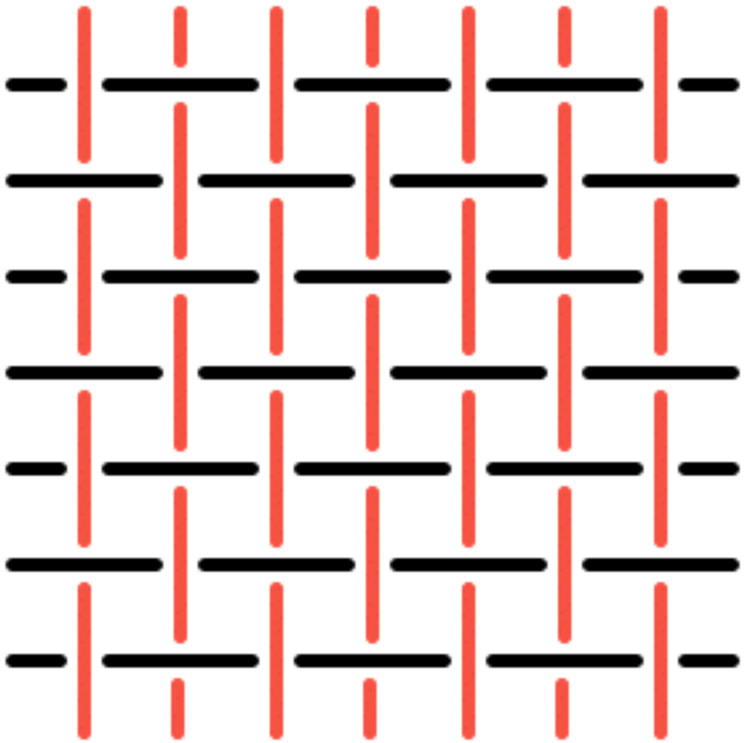
- *Woven fabrics* consist of two sets of *threads* interlaced at right angles.
- The *warp threads* run the length of the fabric
- The *weft* (fill, pick or woof) *threads* are passed back and forth between the warp threads.
- When weaving, the warp threads are raised or lowered in patterns, leading to different weaves.
- Factors include:
  - The pattern in which warps and wefts cross
  - Relative sizes of threads
  - Relative spacing of threads
  - Colours of threads

MANCHESTER 1824

The University of Manchester

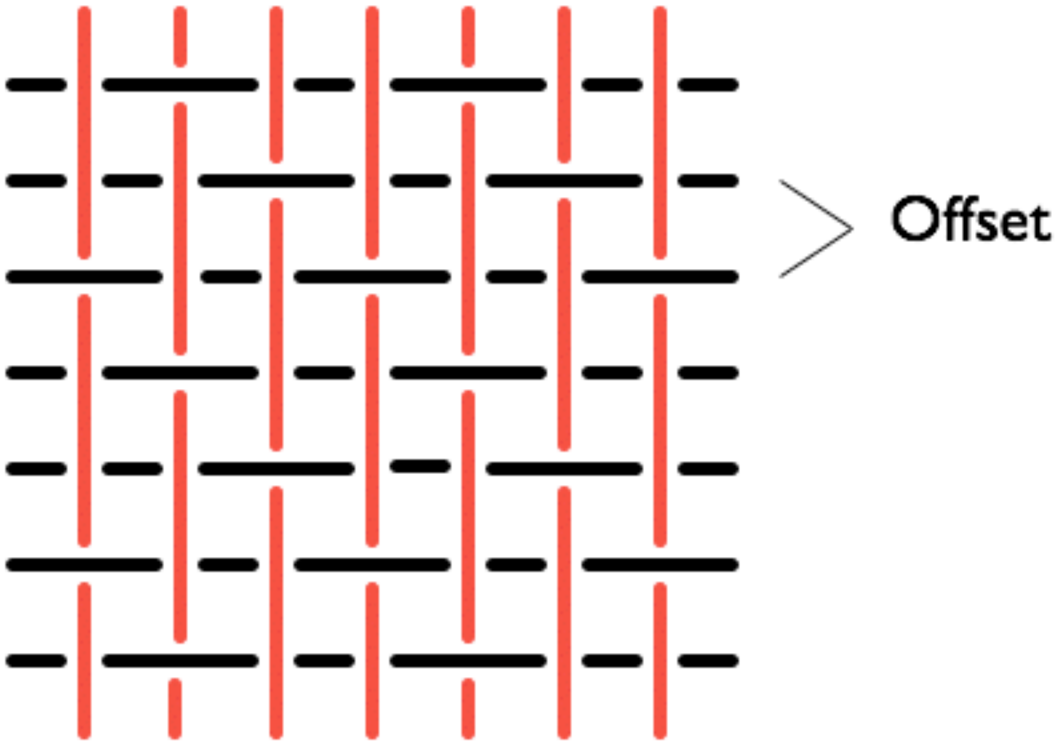
Plain Weave

- Over and under in a regular fashion



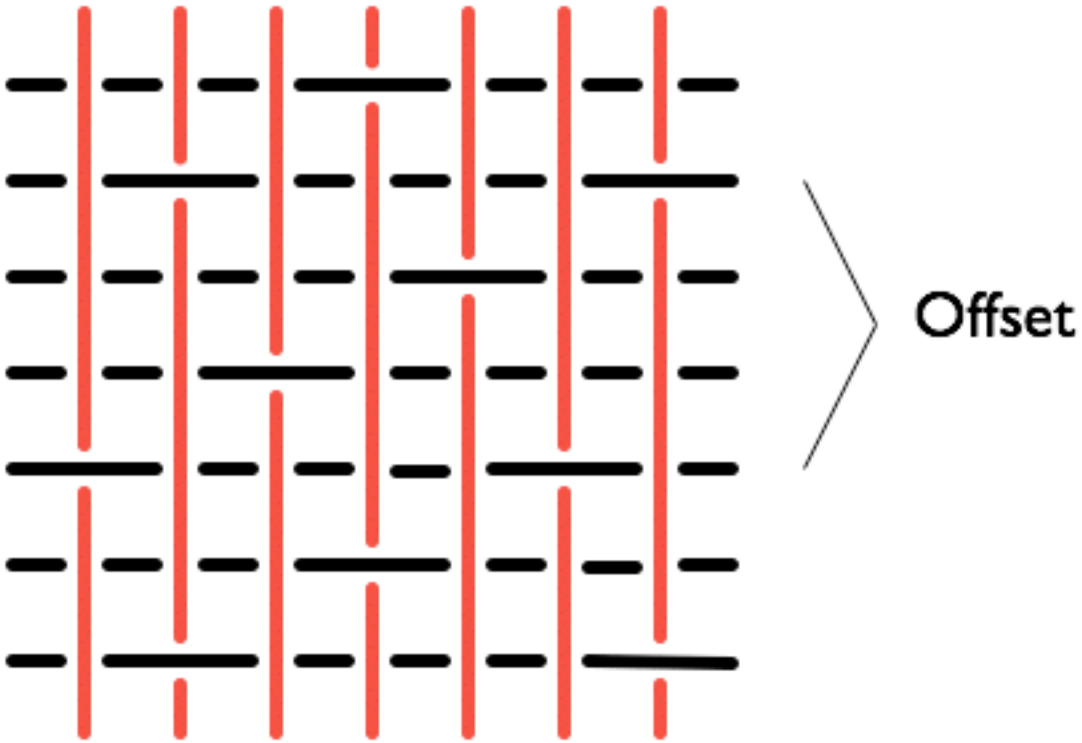
Twill Weave

- Warp end passes over more than one weft
  - Known as “floats”
- Successive threads offset by 1



Satin Weave

- Longer “floats”
- Offsets larger than 1



Classifying Cloth

- The example provides a number of rules that describe how particular kinds of cloth are described.
- alternatingWarp → plainWeave
  - *If a piece of cloth has alternating warp, then it's a plain weave.*
- hasFloats, warpOffsetEq1 → twillWeave
  - *If a piece of cloth has floats and a warp offset of 1, then it's a twill weave.*
- There are many other properties concerning the colour of threads, spacings etc.

Using the Rules

- We could use these rules to build a system that would be able to recognise different kinds of cloth through recognising the individual characteristics.
- The example given shows that once we have recognised the following characteristics
  - diagonalTexture
  - floatGTSink
  - colouredWarp
  - whiteFill

  then we can determine that this cloth is denim.

MANCHESTER 1824

The University of Manchester

Knowledge Representation

- Although this is relatively simple (in terms of both the expressivity of the language used and the number of facts), this really is an example of Knowledge Representation.
  - The rules represent some knowledge about cloth -- objects in the real world
  - Together they form a knowledge base
  - The knowledge base along with some deductive framework allow us to make inferences (which we hope reflect the characteristics/behaviour of the real world objects)

What is a Knowledge Representation?

Davis, Shrobe & Szolovits

http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html

- Surrogate

That is, a representation

- Expression of ontological commitment

of the world

- Theory of intelligent reasoning

and our knowledge of it

- Medium of efficient computation

that is accessible to programs

- Medium of human expression

and usable

KR as Surrogate

- Reasoning is an internal process, while the things that we wish to reason about are (usually) external
- A representation acts as a surrogate, standing in for things that exist in the world.
  - Reasoning operates on the surrogate rather than the things
- Surrogates can serve for tangible and intangible objects
  - Bicycles, cats, dogs, proteins
  - Actions, processes, beliefs

KR as Surrogate

- What is the correspondence between the representation and the things it is intended to represent?
  - Semantics
- How close is the representation?
  - What's there?
  - What's missing?
- Representations are *not* completely accurate
  - Necessarily abstractions
  - Simplifying assumptions will be present
- Imperfect representation means that incorrect conclusions are inevitable.
- We can ensure that our reasoning processes are sound
  - Only guarantees that the reasoning is not the source of the error.

KR as Set of Ontological Commitments

- A representation encapsulates a collection of decisions about what to see in the world and how to see it.
- Determine the parts in focus and out of focus
  - Necessarily so because of the imperfection of representation

- Choice of representation
- Commitments as layers

- KR != Data Structure
  - Representational languages carry meaning
  - Data structures may be used to implement representations
  - Semantic Nets vs. graphs

KR as Fragmentary Theory of Intelligent Reasoning

- Incorporates only part of the insight or belief
- Insight or belief is only part of the phenomenon of intelligent reasoning

- Intelligent inference
  - Deduction
- Sanctioned inferences
  - What can be inferred
- Recommended inferences
  - What should be inferred

KR as Medium for Efficient Computation

- To use a representation, we must compute with it.
- Programs have to work with representations
  - The representation management system is a component in a larger system
  - If the representation management system is inefficient, programmers will compensate
- Representations get complex quickly
  - People need prosthetics to work well with them

KR as Medium of Human Expression

- Representations as the means by which we
  - express things about the world;
  - tell the machine about the world;
  - tell one another about the world
- Representations as a medium for communication and expression by us.
  - How general is it?
  - How precise is it?
  - Is the expressiveness adequate?
- How easy is it for us to talk or think in the representation language?
  - How easy is it? vs. can we?

KR - ontologies - OWL

- Since the conception of the *Semantic Web,* (many) people use
  - knowledge base
  - ontology

  synonymously…we do here

- OWL is one language to for writing ontologies
  - just like Java is one language for writing programmes

Ontologies

- **Metadata**
  - Resources marked-up with descriptions of their content. No good unless everyone **speaks the same language**;
- **Terminologies**
  - Provide shared and common vocabularies of a domain, so search engines, agents, authors and users can communicate. No good unless everyone **means the same thing**;
- **Ontologies**
  - Provide a **shared and common understanding** of a domain that can be communicated across people and applications, and will play a major role in supporting information exchange and discovery.

Ontology

- A representation of the shared **background knowledge** for a community
- Providing the intended meaning of a formal vocabulary used to describe a certain **conceptualisation** of objects in a domain of interest
- In CS, ontology taken to mean an **engineering** artefact
- A vocabulary of terms plus **explicit characterisations** of the **assumptions** made in interpreting those terms
- Nearly always includes some notion of **hierarchical classification** (is-a)
- Richer languages allow the definition of classes through **description** of their **characteristics**
  - Introduce the possibility of using **inference** to help in management and deployment of the knowledge.

Ontologies and Ontology Representations

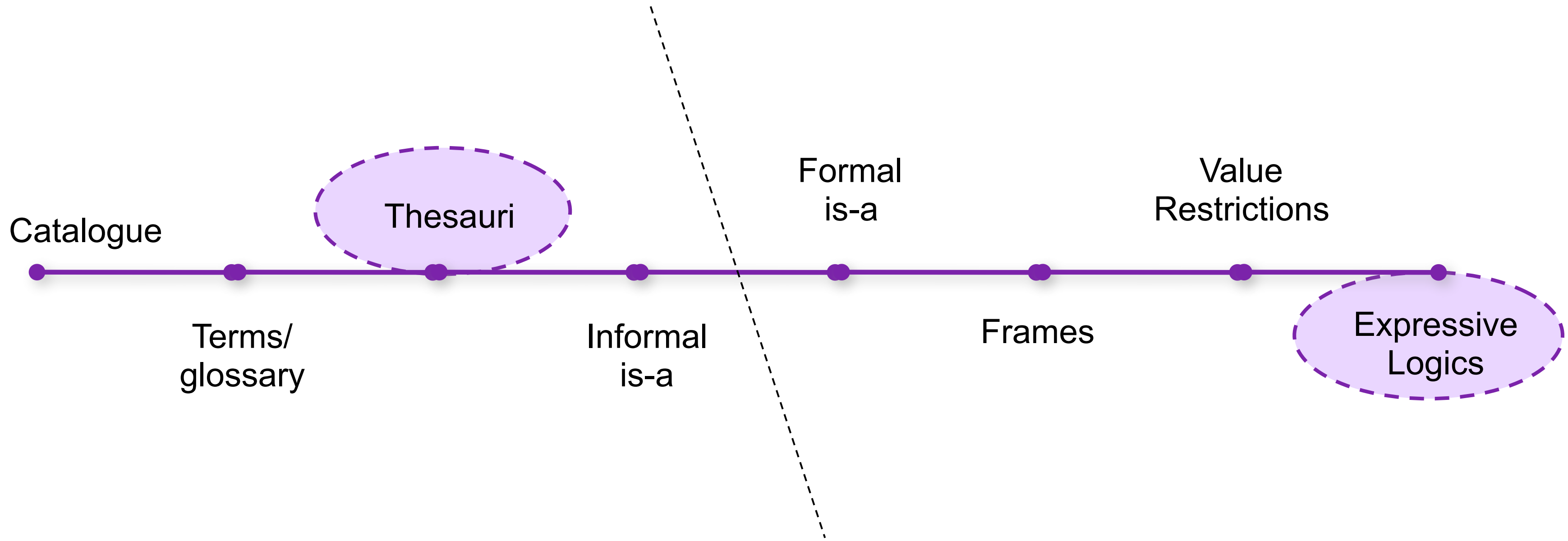
- “Ontology” – a word borrowed from philosophy
  - But we are necessarily building logical systems
- “Concepts” and “Ontologies”/ “conceptualisations” in their original sense are psychosocial phenomena
  - We don’t really understand them
- “Concept representations” and “Ontology representations” are engineering artefacts
  - At best approximations of our real concepts and conceptualisations (ontologies)
    - And we don’t even quite understand what we are approximating

Ontologies and Ontology Representations (cont)

- Most of the time we will just say “concept” and “ontology” but whenever anybody starts getting religious, remember…
  - *It is only a representation!*
    - We are doing engineering, not philosophy – although philosophy is an important guide

- There is no *one* way!
  - But there are consequences to different ways
    - and there are wrong ways
      - and better or worse ways for a given purposes
  - The test of an engineering artefact is whether it is fit for purpose
    - Ontology representations are engineering artefacts

A Spectrum of Representation



So why is it hard?

- Ontologies are tricky
  - People do it too easily;
    People are not logicians
    - Intuitions hard to formalise
- Ontology languages are tricky
  - "All tractable languages are useless;
    all useful languages are intractable"
- The evidence
  - The problem has been about for 3000 years
    - *But now it matters!*
    - *The semantic web means knowledge representation matters*

Ontology Engineering

- How do we build ontologies that are
  - Fit for purpose? (and what does that mean?)
  - Extensible?
  - Flexible?
  - Maintainable?
- Methodologies and guidelines
  - Knowledge acquisition
  - Ontology patterns
  - Normalisation
  - Upper level ontologies

Beware

- OWL is not all of Knowledge Representation
- Knowledge Representation is not all of the Semantic Web
- The Semantic Web is not all of Knowledge Management
- The field is still full of controversies

- This course unit is to teach you about implementation in OWL