The Representation of Reality in Computational Form

Nick Rossiter

Visiting Fellow

with Michael Heather and Dimitrios Sisiaridis

Computing, Engineering and Information Sciences

Northumbria University, Newcastle NE2 1XE, UK

Presentation at ANPA 33

Cambridge University 9-12 August 2012

Aristotle

- Aristotle coined the word 'categories' to describe a structure of classes
    - Used as the title of one of the books of his treatise on logic, the *Organon*.
- This begun a 2,000 year history for the category to describe the structured level at the foundation of logic.

Symbolic Logic Diversion

- The study of logic diverted to the symbolic logic of set theory around 1900
  - So the concept of a category with classes at various levels was no longer easy to represent.
  - For the elements for the set are independent of one another and a set cannot be a member of itself
  - No inherent possibility to represent recursion nor relations other than by external functions.

Indispensable Categories

- Nevertheless the category has become an indispensable component for many disciplines and the concept is still developing today.
- It is the primary classification system for Wikipedia which itself has about 500 types of categories defined.
- With advances in information systems the concept of typing is an aspect of categories that has increased in importance.

Accepted Approaches to Category Theory

In many areas work has centred around cartesian closed categories.

Such areas include:

Compiler design and evaluation

Language analysis

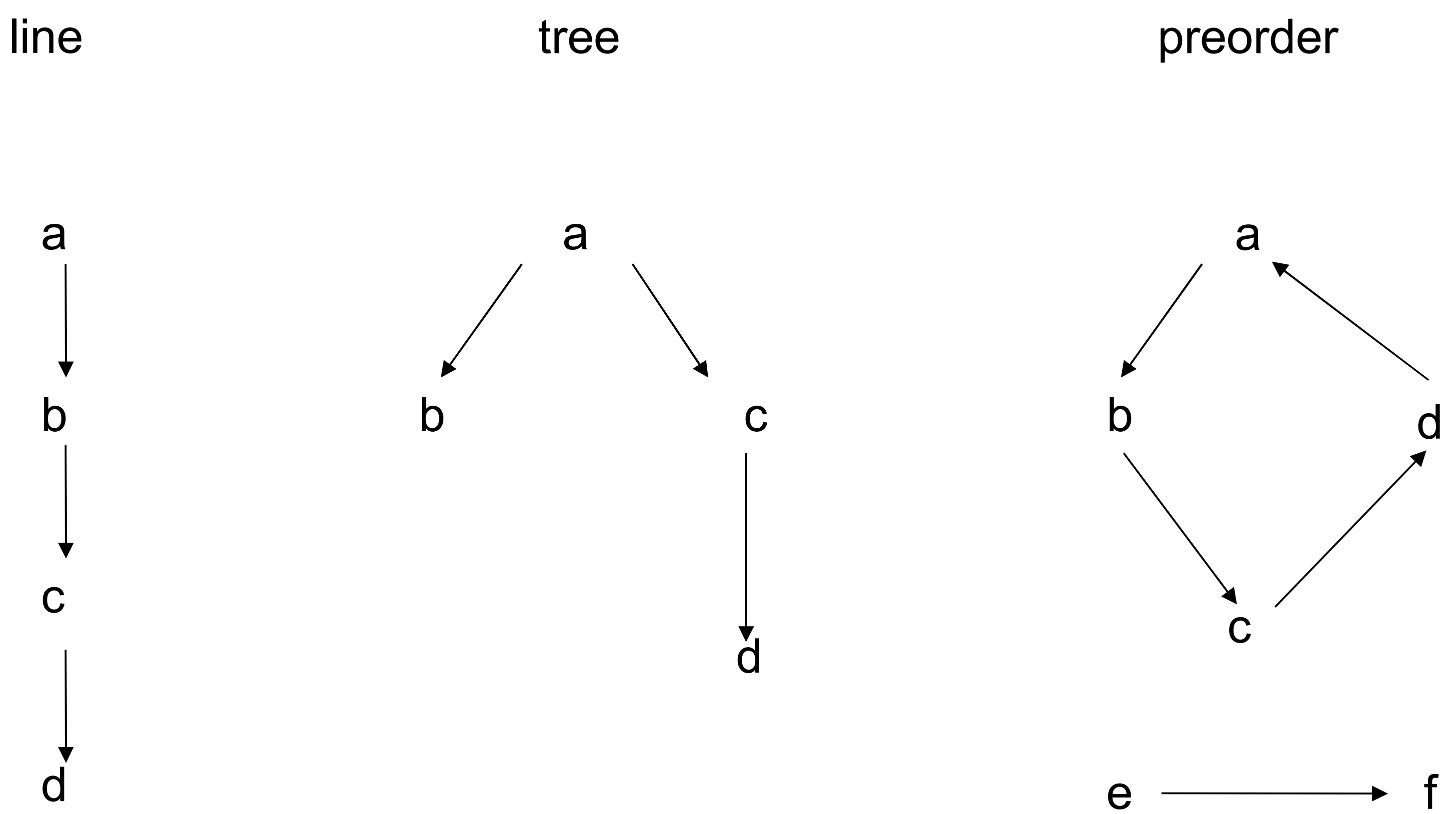
Database design

Metaphysics

Starting Point

- Reality concerns some area of interest.
    - The Universe of Discourse.
- Reality is a preorder
    - with cycles, trees and lines

Orders in Set



Tree is a special case of partial order where each element has only one parent.
Preorder can include unrelated fragments.
Sets are equipped with an order (as an add-on)

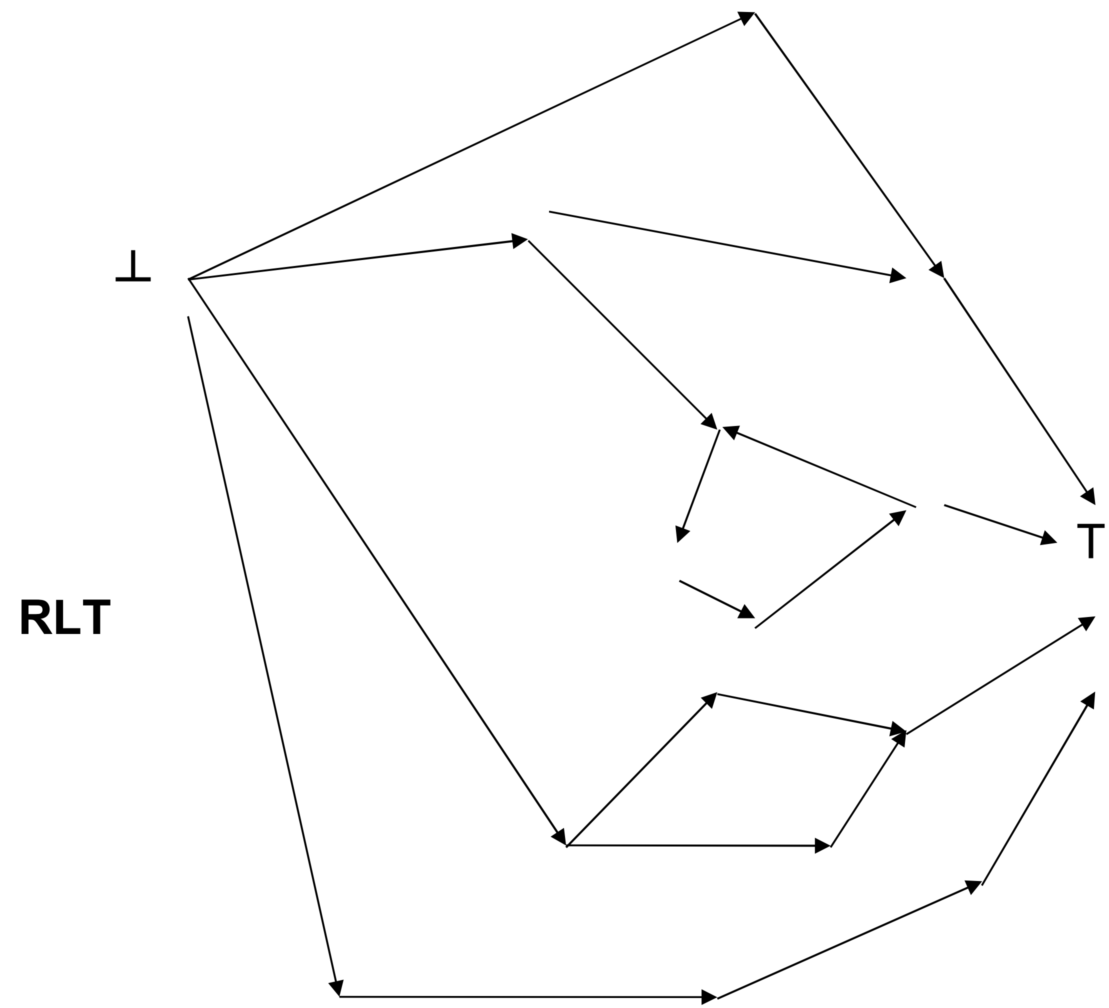
Orders in Categories

- Ordering is fundamental in category construction.
- Arrow is overloaded for many tasks
  - Calculation as by function e.g. square
  - Relationship e.g. marriage
  - Ordering e.g.
    - Class list on merit (line)
    - Relationships as in family tree (partial order)
    - Hours on an analogue clock/watch (preorder)
      - Directed graph, not acyclic
      - Connection of points in a topological space

Reality

- A general intensional category of type preorder
- A subcategory (not a partition) of the Universe
- Sometimes termed the Universe of Discourse
- With just one universe of discourse, will have an upper bound
  - The terminal object $\top$
- And a lower bound
  - The initial object $\bot$
- Everything must be connected
  - Fragments not permitted as outside bounds

Example Reality Category RLT



Terminal Object T
Unique upper bound on category

Initial Object ⊥
Unique lower bound on category

Properties of RLT

- Rich enough to represent reality with all types of order inter-mingled
- Provides an upper bound T, lower bound $\perp$ and connectivity
- An entry point $\perp$ exists, so that it could potentially be addressed by an identity functor

But

- Not cartesian closed (no products)

Handling RLT computationally (best industry practice)

In relational database, model cannot handle the great variety of orders directly

Need to map to a collection of relations, maintaining the arrow dependencies

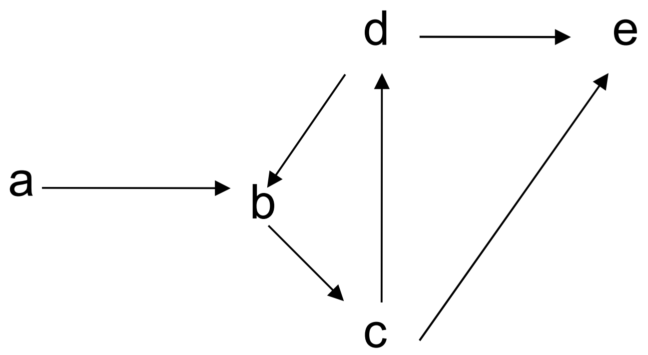
Object-oriented database model is more versatile, with less adaption required

Principle is that a single large cumbersome structure is unwieldy for updating, because not all the data may be available at the same time

So the structure is fragmented

Fragmentation

Take a simpler example:



RLT

Initial object a
Terminal object e
Preorder on b,c,d

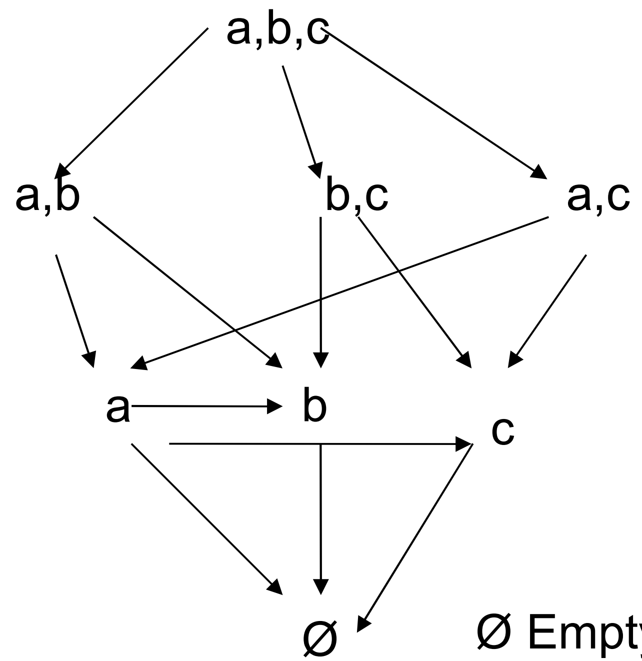
Universal Relation U

Construct the universal relation U including all the objects a,b,c,d,e in RLT

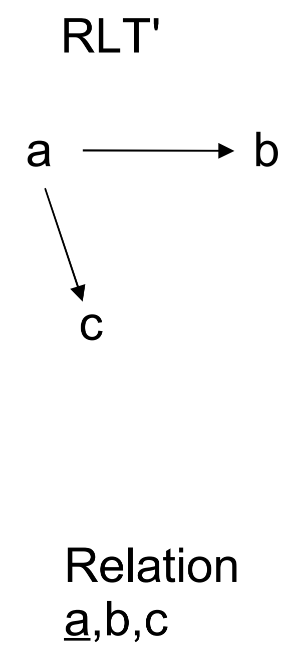
U contains all possible subsets of the objects in RLT

Plus the dependency arrows from RLT

Example of Dependency Diagram – all possible projected subsets



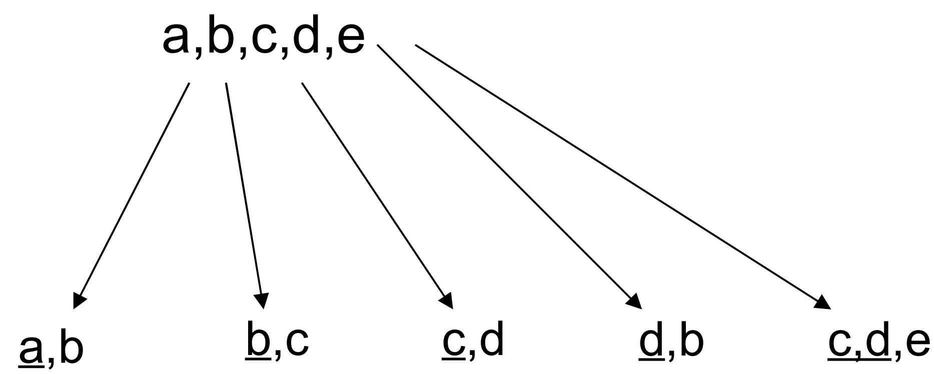
Ø Empty set



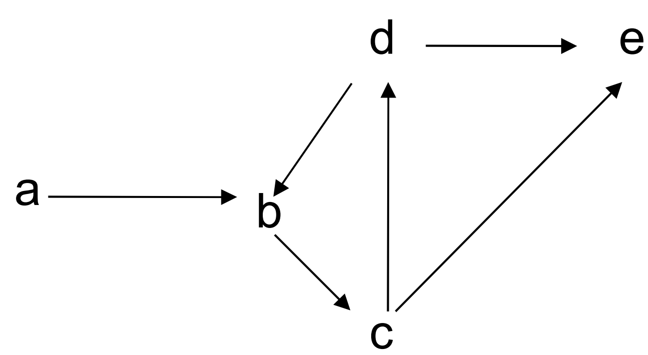
This is example for 3 elements, requires a lot more space for 5 (32 to be precise!). For 10 elements would require 1024 members.

For earlier example RLT, project out a dependency diagram D in U, including all the arrows in RLT

Arrows are projections
Underlined are keys



D (in U)



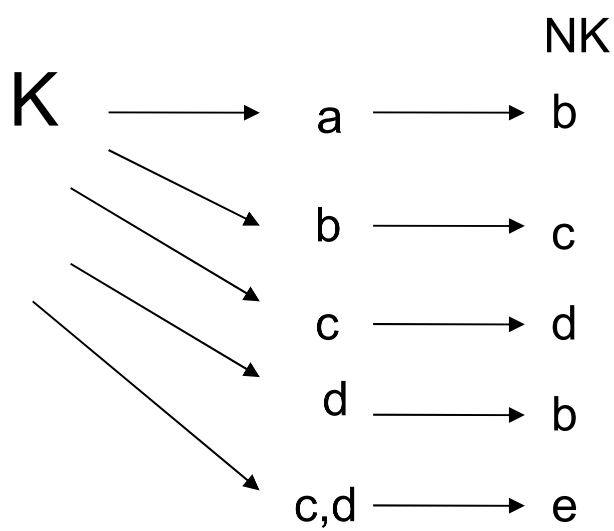
RLT

Relations are:
a,b
b,c
c,d
d,b
c,d,e

So you're now a database designer!

Keys as Entry Point

For relations can introduce Keys (K) as initial entry point:



K is lower bound
NK is non-key
NK collectively is upper bound

Could be tidied up, could replace d,b and c,d,e with d,b,e
This is done through database normalisation

So what is being done by database designer

RLT is not cartesian closed

But U is cartesian closed:

Has terminal object of Ø null

Has products (replace ',' by 'x' e.g. 'a,b' is 'a x b')

Has connectivity (exponentials)

The initial object a,b,c,d,e, that is a x b x c x d x e, can be used as a unique entry point 1_U, the identity functor

D is also Cartesian Closed

D is derived from U

Containing some of its objects and some of its arrows

So D is a subcategory of U

D is also cartesian closed as it preserves limits and colimits with

Initial object K (the product of the keys)

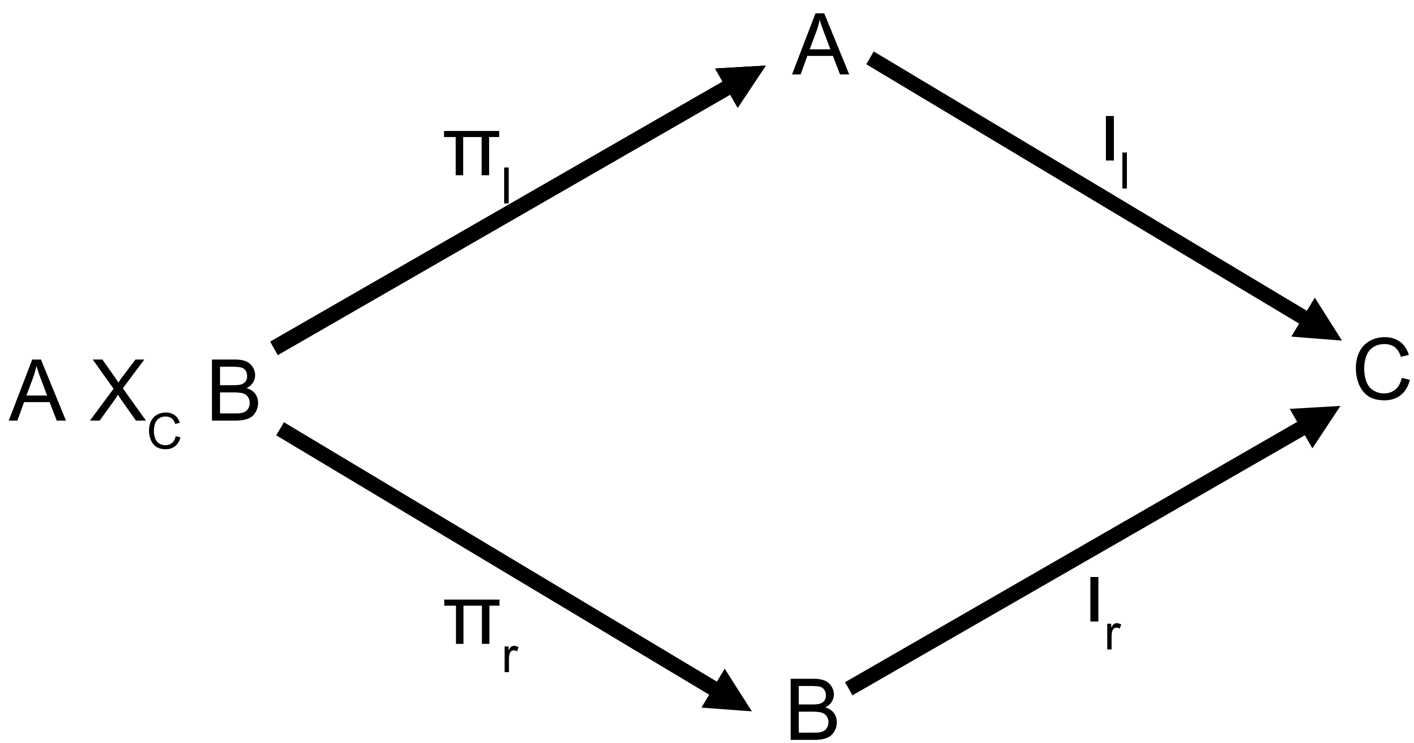
Terminal object NK (the sum of the non-keys)

The dual of any cartesian closed preorder, the cocartesian, is also closed, trivially.

LCCC

- In practice we use a variant of cartesian closed categories for detailed representation
  - Locally cartesian closed category
    - Product is replaced by a relationship
  - Product is all possible pairs
    - e.g. account number X borrower name (A X B)
  - Relationship is those pairs that satisfy a particular context
    - e.g. account number X borrower name in the context of cash owed ($A\ X_C\ B$)
  - In category theory this is a pullback (with adjointness properties)

Pullback



C is A+B+C

Other areas of computation

Machine architecture and processing

Reality for architecture is very simple:

Von Neumann – Serial (line ordering)

Parallel – partial order

Recursive – preorder

Trends in Computation

Von Neumann approach still very popular.

But used without refinement is very primitive.

Parallel/recursive methods used to some extent now, for instance quad/dual processors. Good for heavy simple tasks e.g. video processing.

Why does von Neumann continue?

Processing Languages

A programming language provides facilities for looping and recursion on serial architecture.

All languages now provide iteration and recursion except assemblers and machine code.

Language compilation and processing is very complex, involving many preorders and partial orders, giving complex RLT categories.

Theoretical work underpinning the languages is being done using locally cartesian closed categories (LCCC) at Cambridge and elsewhere.

Summary

Real-world structures employed in computation are very complex, whether for handling the outside world or for internal purposes.

Cartesian closed categories underpin the handling of such complexity.

A particular variant, locally cartesian closed categories, has been found to be effective in detailed work.