Использование онтологий при анализе предметных областей и проектировании научных информационно-вычислительных систем

Фазлиев А.З.

История

Проект ATMOS (Intas-(2001-2004))

Проекты РФФИ (6 проектов/10 лет)

Атмосферная спектроскопия, атмосферная радиация, климат и погода, атмосферная химия

Атмосферный аэрозоль

Проект ATMOS

Структура предметной области

Модель процедурной предметной области

- Сеть задач
- Цепь задач

Типы задач

- Вычислительные
- Измерительные

Модель вычислительной задачи

Задача

(входные и выходные данные)

Данные задачи предметной области

1. Схема данных
2. Ограничения (модель предметной области)
3. Вычисления со значениями данных

Данные – структура, ограничения и вычисления

Данные задачи нахождения уровней энергии

1. Схемы данных (уровень энергии, квантовые числа (v_1, v_2, v_3, J, k_a, k_c), функция потенциальной поверхности, ...)
2. Ограничения ($w<45000$ cm^{-1}, $k_a+k_c=J \vee J+1$, симметрия молекулы, ...)
3. Вычисления со значениями данных (интегро-дифференциальные, алгебраические и т.д.)

Метаданные задачи

1. Схема метаданных
2. Ограничения на значения метаданных
3. Вычисления, связанные со значениями метаданных

В репозитариях метаданных схемы данных являются данными.

Метаданные задачи
нахождения уровней энергии

1. Схемы метаданных (интенсионалы данных, экстремальные значения данных, нарушения ограничений, …)
2. Ограничения на значения метаданных (нет)
3. Вычисления, связанные со значениями метаданных (нахождение min, max значений, проверка ограничений, …)

Модель молекулярной спектроскопии

1. Цепь прямых вычислительных задач (вычисление уровней энергии, вычисление частот переходов, вычисление параметров спектральных линий, вычисление спектральных функций)
2. Цепь обратных вычислительных задач (вычисление параметров спектральных линий, вычисление частот переходов, приписывание квантовых чисел, вычисление уровней энергии)

Концептуализация молекулярной спектроскопии Глобальные задачи

Молекула, атом, вакуум, излучение, уровень энергии, квантовое число, частота перехода, интенсивность, сдвиг линии, коэффициент поглощения, точность измерения частоты перехода, …

Иметь уровень энергии, иметь частоту перехода, ….

Концептуализация молекулярной спектроскопии
Локальные задачи

Min (max) значение уровня энергии (вакуумной частоты перехода, J, …), число уровней энергии (….), авторы, публикация, ….

Иметь J_{max}, иметь автора, …..

Онтологии задач

1. Схема для баз знаний
2. Ограничения на логику
3. Логические выводы, семантический поиск, ….

Явная спецификация разделяемой концептуализации

Онтология задачи нахождения уровней энергии

1. Схема для баз знаний (молекула-> уровень энергии -> квантовые числа)
2. Ограничения на логику (SHOIQ)
3. Автоматизированные логические выводы и семантический поиск

Различия (WonderWeb)

1. Схемы данных обеспечивают явную, но не автоматизируемую обработку семантики данных. Интеграция схем данных требует специальных усилий.
2. Онтологии являются логическими системами, включающими в себя семантику.
3. Онтологии более часто используются по сравнению со схемами данных
4. Онтологии децентрализованы по своей природе

Информационно-вычислительная система

1. Структура ИВС (слой данных и вычислений, информационный слой, слой знаний)
2. Ограничения (программное обеспечение, модель пользователя, …)
3. Сбор, хранение, доставка и предоставление информационных ресурсов и сервисов

Проект WonderWeb

Три поколения Web

1. Ручная запись HTML
2. Машинная генерация HTML
3. Машинная генерация семантических аннотаций (Semantic Web)

Проект WonderWeb

Требование к библиотеке

Минимальность

Строгость (точность)

Разделяемость

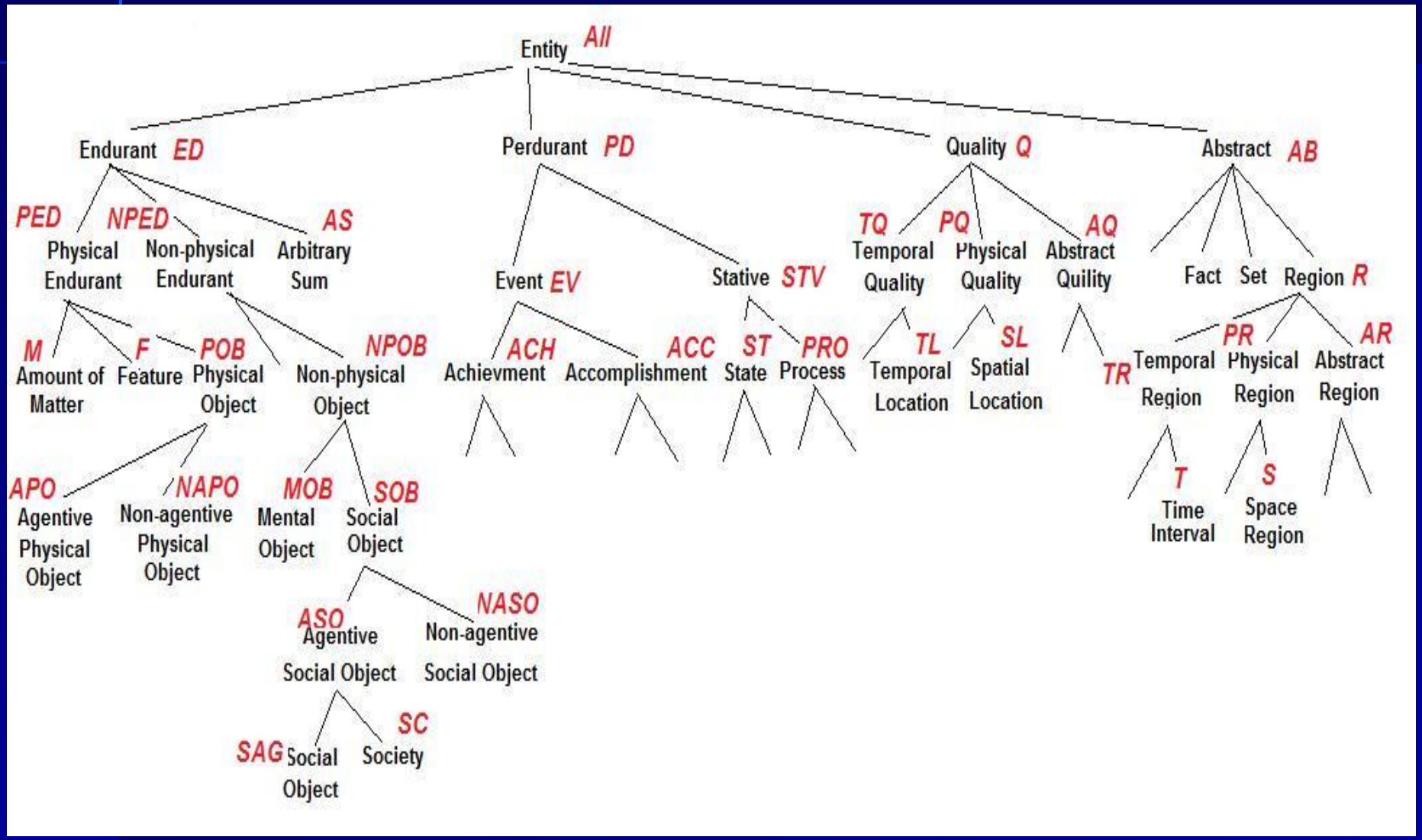
Проект WonderWeb
FOL

DOLCE – descriptive ontology for linguistic and cognitive engineering.

Particular – сущность не имеющая индивидов

Universal - сущность имеющая индивиды

Проект WonderWeb
Базовые категории DOLCE



Проект WonderWeb
Endurants and Perdurants

Endurants – сущности находящиеся во времени, но не имеющие зависящих от времени частей.

OWL: Индивид количественные значения свойств которого не изменяются во времени

Perdurants - сущности находящиеся во времени и имеющие зависящие от времени части

OWL: Индивид количественные значения свойств которого изменяются во времени

Проект WonderWeb
Qualities and quality regions

Qualities являются Particulars, **Properties** являются Universals.

OWL: Subject – Predicate - Object.
Пример: Predicate = Quality, Predicate-Object = Property

OWL: **Quality region** = Range

Пример: Цвет

Пространство и время – специальные **качества (!?)**

Проект WonderWeb
Qualities and quality regions

Два вида **неотъемлемости качества**: прямая и не прямая неотъемлемость.

Perdurants имеют определенную локализацию во времени, но их пространственная локализация определяется не прямым образом.

Endurants (физические) имеют явную локализацию в пространстве, а локализация во времени определяется неявно в зависимости от Perdurants в который они входят.

В онтологии DOLCE для качеств не определено отношение части.

Проект WonderWeb
Базисные аксиомы для примитивов

Parthood: "x is part of y"
$P(x,y) \rightarrow (AB(x) \vee PD(x)) \wedge (AB(y) \vee PD(y))$

Temporary Parthood: "x is part of y during t "
$P(x, y, t) \rightarrow (ED(x) \wedge ED(y) \wedge T(t))$

Constitution: "x constitutes y during t"
$K(x, y, t) \rightarrow ((ED(x) \vee PD(x)) \wedge (ED(y) \vee PD(y)) \wedge T(t))$

Participation: "x participates in y during t"
$PC(x, y, t) \rightarrow (ED(x) \wedge PD(y) \wedge T(t))$

Quality: "x is a quality of y"
$qt(x, y) \rightarrow (Q(x) \wedge (Q(y) \vee ED(y) \vee PD(y))$

Quale: "x is the quale of y (during t)"
$ql(x,y) \rightarrow (TR(x) \wedge TQ(y))$
$ql(x, y, t) \rightarrow ((PR(x) \vee AR(x)) \wedge (PQ(y) \vee AQ(y)) \wedge T(t))$

Решетка категорий

Категории верхнего уровня

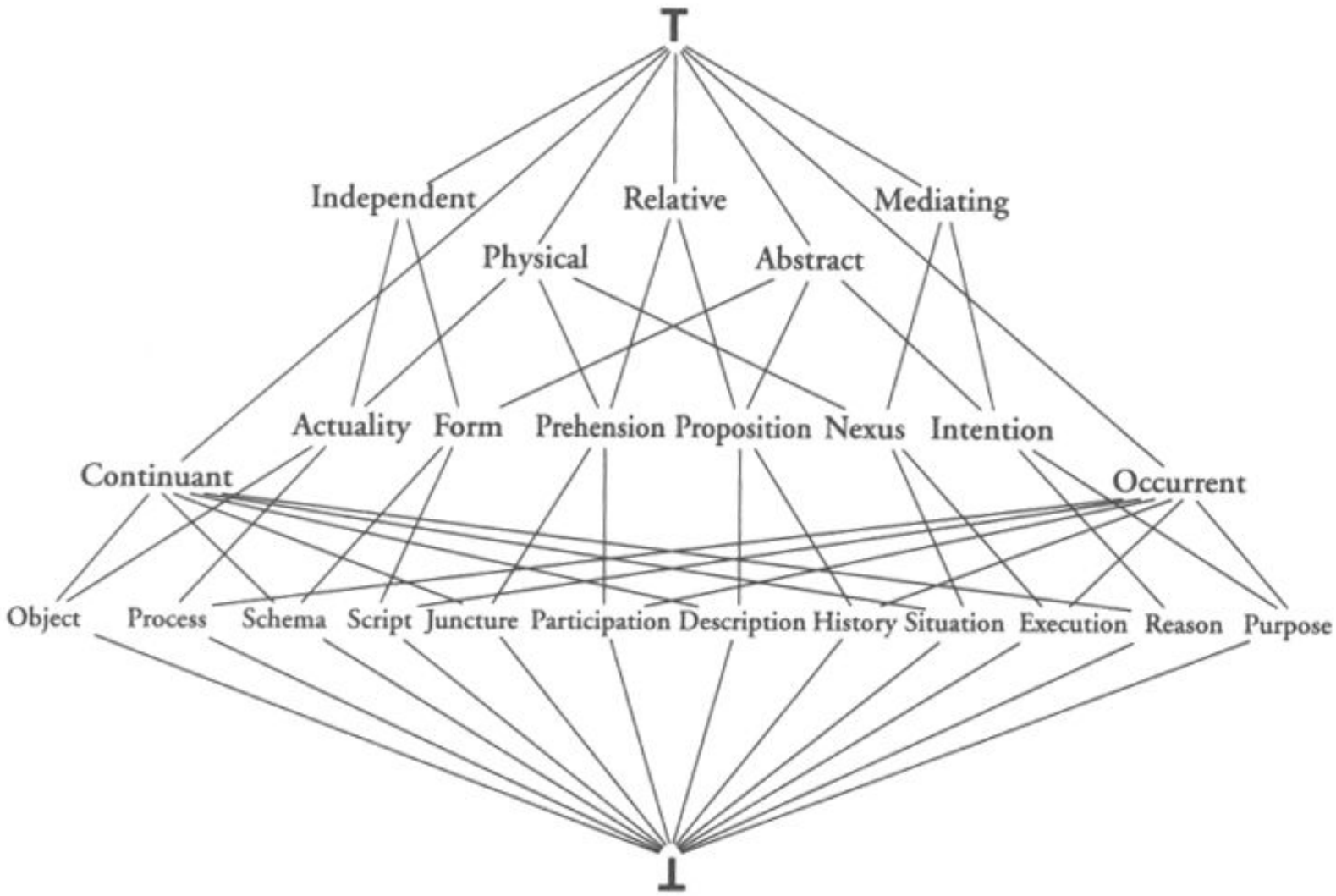


FIGURE 2.6 Lattice generated by the top three distinctions

Классификация ролей

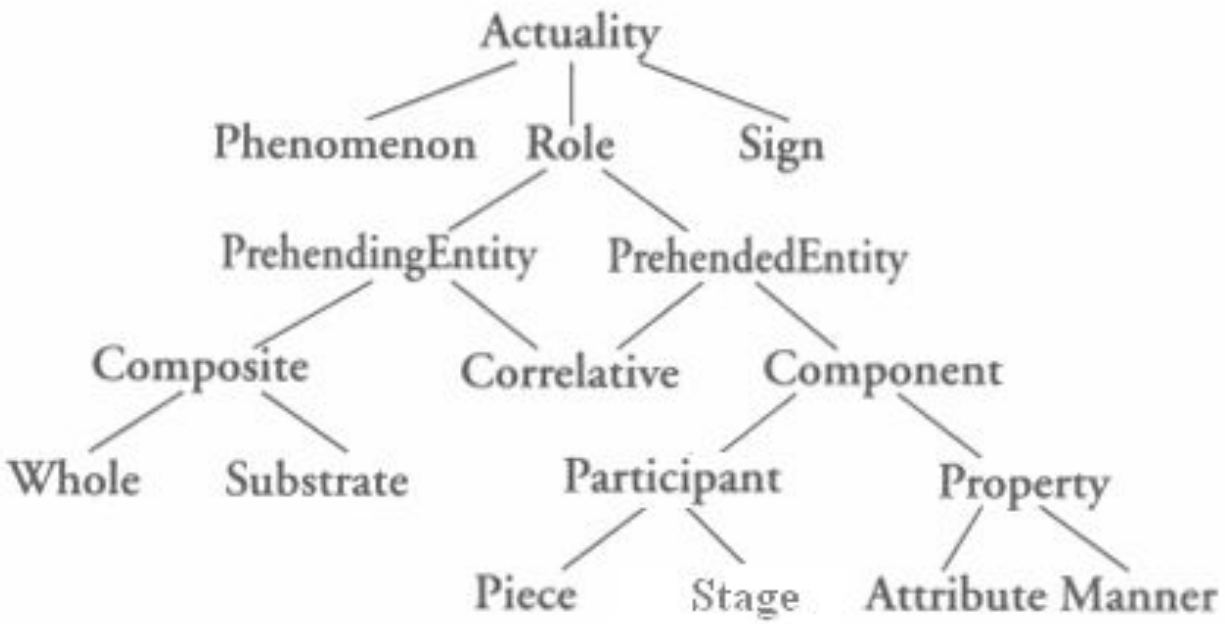


FIGURE 2.11 Classification of actual entities by structure, role, or sign

Категории форм

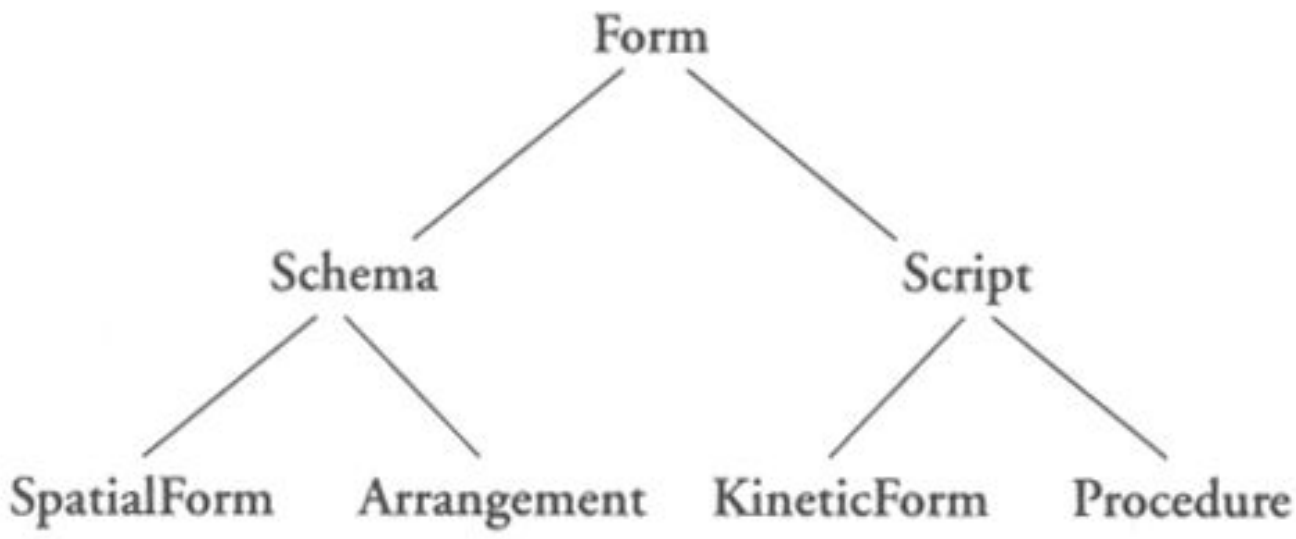


FIGURE 2.12 Temporal and spatial subdivisions of Form

Иерархии теорий

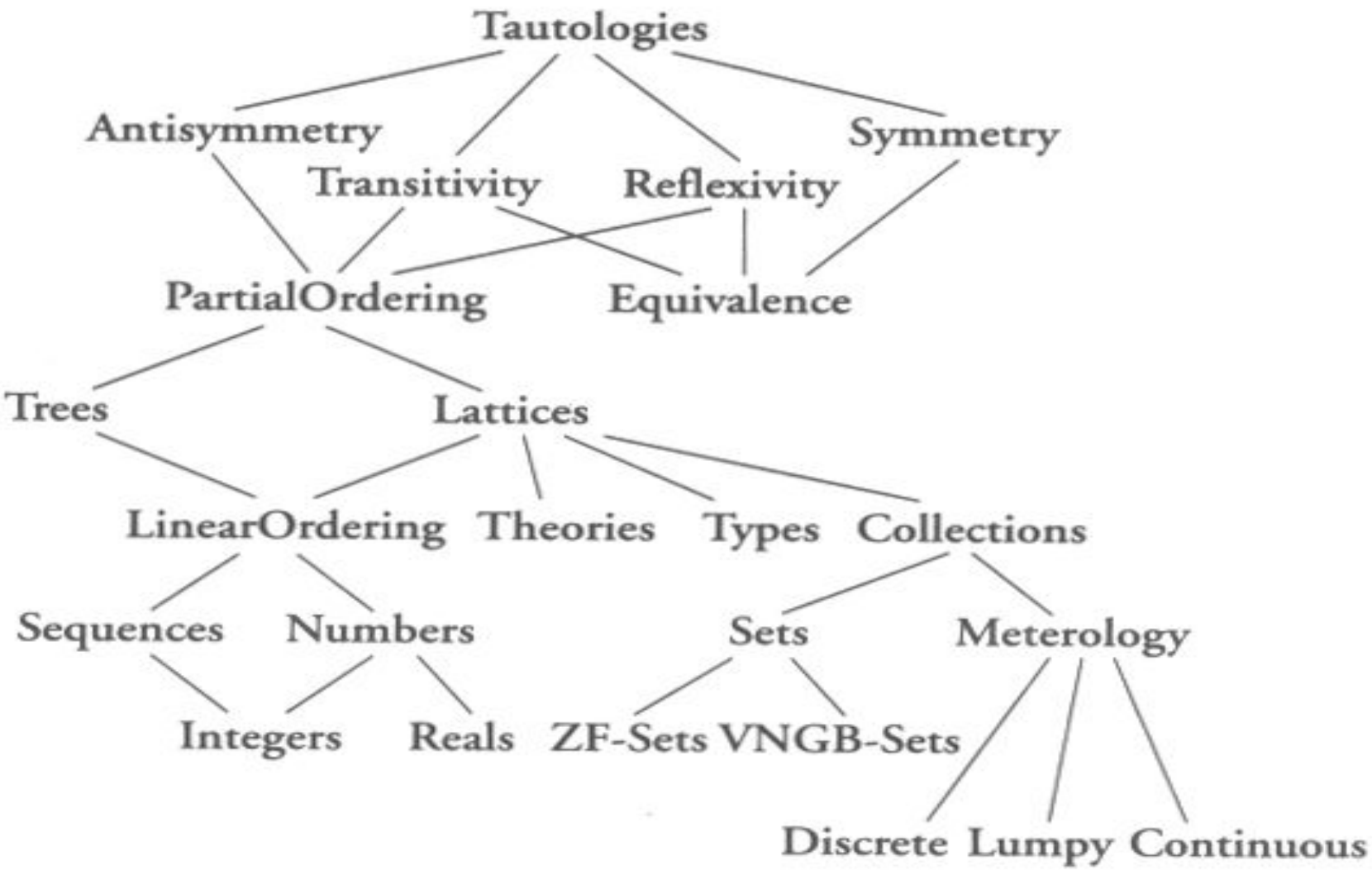


FIGURE 2.14 A generalization hierarchy of theories