Schemata of Interoperable Database Systems: Beyond Data Modeling

Amit Sheth
Bell Communications Research, Inc., USA

Invited Talk

DS-5, 1992, Lorne, Australia.

Opinions expressed during this talk are those of the speaker and not of Bellcore.

Far (Schematically)
) Near (Semantically)

Sheth and Vipul Kashyap

Invited Paper,
DS-5, Lorne, Australia.

Components of SemPro

- Context
  - helps to capture the semantic aspect of the relationship between the two objects
- Abstraction
  - helps to capture the structural aspect of the relationship between the two objects
  - often expressed as a mapping between domains
- Domain

  sets of values from which the objects take their values
- State

  extension of an object recorded in a database

aspects of Heterogeneity

- STRUCTURAL TAXONOMY

rent ways
ght be

enumerate and classify the various types of structural and representational differences

- SEMANTIC TAXONOMY

antic
exist between

identify types of semantic similarities between two objects

TION

tween the two taxonomies, i.e. for each kind
rence, identify the possible semantic similarities.

Bellcore
Bell Communications Research

What is Semantics?

"the scientific study of the relations between signs and symbols and what they denote or mean".*

So we have *sign and symbols* => model, language

- this is structured, limited, incomplete

what they *mean or denote* => real world information

- real world semantics
- this is unstructured, unlimited, possibly complete

One way to make a model useful is to talk about modeling in a specific *domain of discourse* [i.e., context]- hence limiting the amount of information modeled (semantics captured).

To discuss semantics, we need to deal with **incomplete**, **inconsistent**, and **uncertain** information.

*Reference: William Woods, "What is a Link?."

erability and Semantics

ural/representational) difference
etween two objects is usually of interest
mantically related!

e addressed in the any type of
systems.

l federation

ma integration

ed federation

ing a view over multiple databases

nt data management

e multidatabase dependencies

e, we need to identify semantically related

Bellcore
Bell Communications Research

States of the Objects

- Actual states of the databases to which object belongs
  - Extension of an object recorded in a database (different from actual real world state of entity)
  - Different extensions of objects can correspond to same state wrt Real World Semantics

mains of the Objects

ects are defined to take values from
e thought of as types

nic (non-decomposable)
posite (composed of other atomic or
lomains
bject as subset of cross products of domains
utes of the object

Bellcore
Bell Communications Research

Abstractions used to relate object domains

- Mapping is a typical mechanism used to represent the relation between the two domains
- Mapping between semantically related domains may not exist (denoted by NONE)
- Some useful, well defined abstractions:
  - total 1-1 value mapping
  - partial many-one value mapping
  - generalization
  - aggregation
  - functional dependencies
- ANY, a special term to denote any of the above mappings
- NEG, a special term which denotes the absence of any mapping between semantically incompatible objects

Why Context?

An Abstraction by itself does not capture the semantics of a relationship.

Example: Consider two objects

STUDENT(Id#, Name, Grade)

DEPARTMENT(Num, Name, Address)

Domain(Id#) = Domain(Num) = {123, 456, 789}

- Thus, it is possible to construct a mapping between Id# and Num, two semantically unrelated objects.
- This mapping is not meaningful. Context can provide information in whether mapping is meaningful or of interest.

Context can be specified ... but may not be specified precisely....

- Using metadata information
  - May not be possible to define the Context precisely

  Semantics associated with an application's view of data [Siegel and Madnick]
- Using schema components
  - Partial context description can be used to determine whether the context(s) in which aemPro is asserted:
    - ALL, i.e. objects compared wrt all possible contexts
    - SAME, i.e. objects are compared wrt the same context
    - SOME, i.e. objects are compared wrt different contexts
    - SUB-CONTEXTS, i.e. constraining a context in some way
    - "Hard-coded"
    - NONE, i.e. objects cannot be compared wrt any context

    where the definition of the context is incorporated in the name of an object

    viz. TELECOMM-EMPLOYEE

s of Semantic Proximities

ivalence

emantic equivalence

xt = ALL

ıction = total 1-1 value mapping between D_1

2

ıantic equivalence

xt = ALL

ıction = total 1-1 value mapping between

S_1 and D_2 X S_2

y by semantic equivalence we mean domain
ivalence.

Bellcore
Bell Communications Research

Types of Semantic Proximities

- Semantic Relationship
  - Context = ALL
  - Abstraction = partial many-one mapping, generalization or aggregation
- Semantic Relevance
  - Context = SAME
  - Abstraction = ANY
- Semantic Incompatibility
  - Context = NONE
  - Abstraction = NEG

# of Semantic Proximities

nblance

ontext(O_1) U Context(O_2)

= NONE

LIATION OF

ame as D_2

Context) = role-of(O_2, Context)

PERSPECTIVES

),SALARY, ...) O_2 = BANK-EMP(ID, SALARY, ...)

ECOMMUNICATION

KING

O_1) U Context(O_2)

) = role-of(O_2, Context) = Subordinate

Uncertain Semantic Relations

- an aspect of semantics
- Previous approaches dealing with uncertainty
  - combining fuzzy and incomplete terminological knowledge with schema knowledge [Fankhauser et al]
  - – assignment of fuzzy strengths is arbitrary

    Example:

    Similarity between Person and Student = 0.6

    This relationship measure is proposed intuitively and we feel that

    it's proposal should be based on a consistent framework

- discrete probability distributions [BGMP90, TCY92]
  make the implicit independence assumption
  may not accurately reflect the Real World Semantics

  Example: Consider two objects as defined below

  INSTR1(Id1, HPhone, OPhone)

  INSTR2(Id2, Phone)

  M_1 : HPhone -> Phone

  M_2 : OPhone -> Phone

  M_1 and M_2 are not independent of each other and are related as

  Id1 X HPhone -> Id2 X Phone and Id1 X OPhone -> Id2 X Phone

- None of the above approaches consider the context information of the objects

- measure of belief
- analysis of underlying samples

Belief Measures based on Semantic Proximity

- Any assignment of belief measures should depend on
  - context(s) of the 2 objects
  - mapping(s) between the 2 objects
  - state(s) of the 2 objects
- Belief Measure as function of semantic proximity descriptor

  $V : semPro(O_1, O_2) \rightarrow [0,1]$
- Bounded Correctness Criteria:

  $0 < X_{Res} < X_{Rel} < X_{Relat} < X_{Eq} < 1$

Research Directions

ı formalism to express various schematic

of context and develop a structure for
tidatabase scenario

etween the object and the context in which
ed to another object

lology to combine the various semantic

t dependent uncertainty functions

completeness

Bellcore
Bell Communications Research