KDI : Knowledge and Data Integration

Representation Diversity

Phase: 1. Introduction & Representation Diversity

W1.L12.M1.T21

Contents

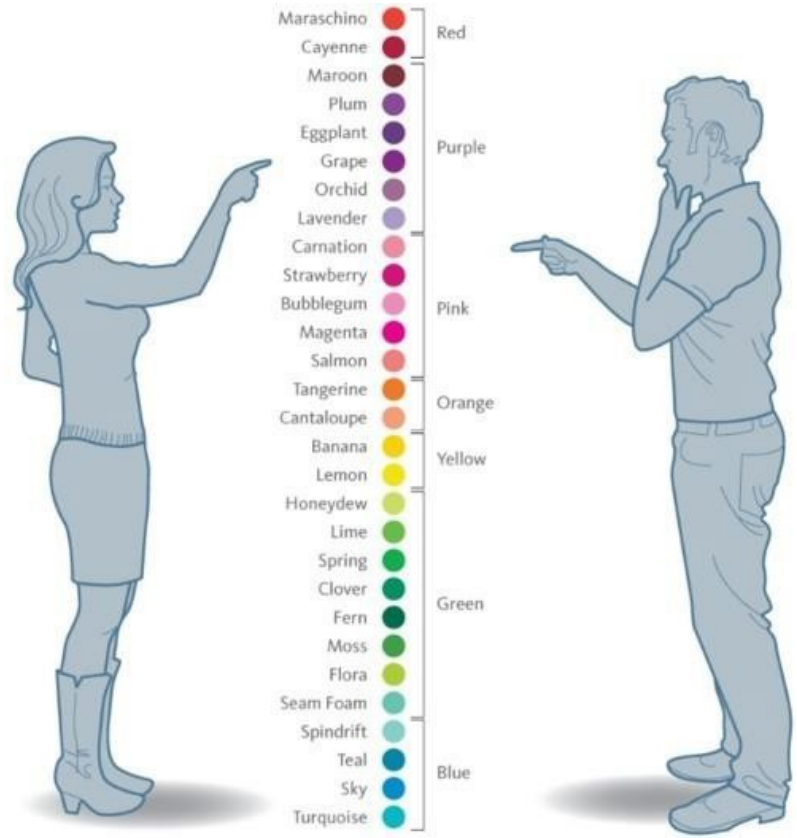
Contents

What is representation diversity?

We have Semantic Heterogeneity (e.g., in language, KBs, DBs) when there are differences in how the same real world phenomenon is represented.

Semantic heterogeneity arises whenever we have KBs and DBs developed by independent parties (in space and time).

We take Representation Diversity to mean semantic heterogeneity, as organized in the three components of language, knowledge and data.



Levels of Representation Diversity

Representation diversity occurs in

1. the different terms and meanings used in language;
2. the different entity types and the properties used;
3. the different entities and the property values used.

We categorize representation diversity in 3 levels:

- Language Diversity
- Knowledge Diversity
- Data Diversity

Representation diversity is unavoidable, at all three levels.

Contents

Diversity across Languages

There are various kinds of languages which can be used to represent the same entity in the world.

- Natural languages for different nationalities, such as *Italian, Chinese, English, etc.*
- Modeling languages for different tools, such as *OWL, RDF, XML, HTML, etc.*
- Programming languages for different syntax, such as *C, Java, Python, PHP, etc.*
- ...

Obviously, the representations of the same phenomenon is different across different languages.

Diversity within Languages

Even in the **same** language, there are multiple ways to represent the same entity, because the mappings between the word and the intended meanings within a language are many-to-many.

Linguistic phenomena such as polysemy, homographs, synonymity, hyponym, etc. are witnesses of these mappings.

Example

Polysemy: a word that has multiple meanings, e.g. **chair**: #1 a seat with a support for the back; #2 a position of professor, etc.
Synonym: a word that means the same (or nearly the same) as another, e.g. **big** means nearly the same as **large**.
Hyponym/ hypernym: a word that is particular than a more general word, e.g. **car** is a hyponym of **vehicle**.

Definition: Language Diversity

Recall the definition of language as

$$language = terms + meanings$$

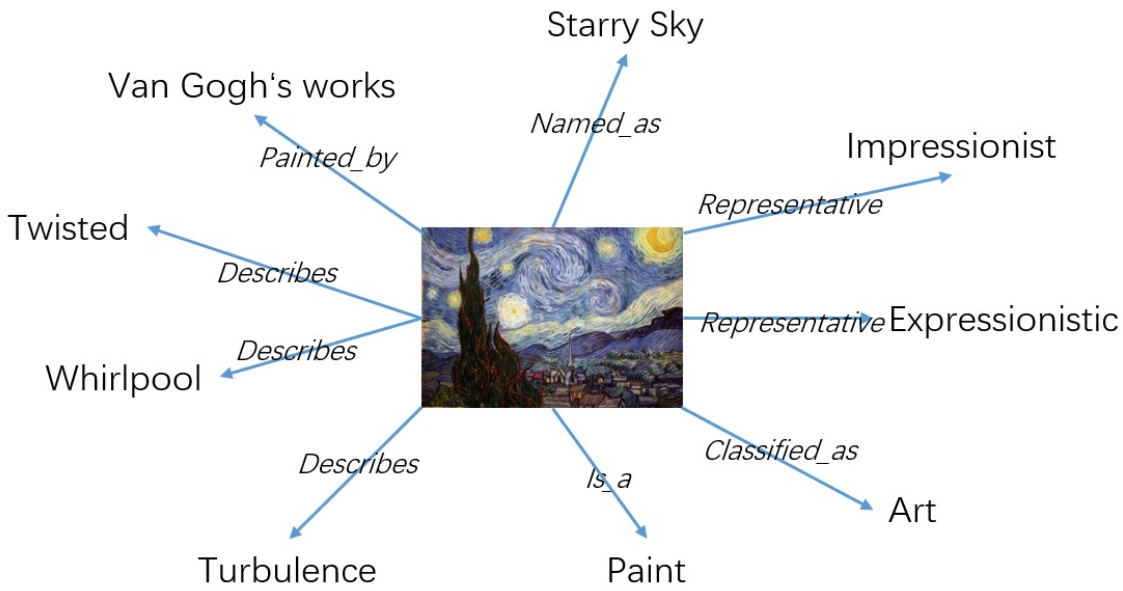
Definition (Language Diversity)

We have Language level representation diversity (short: language diversity) when we have two representations with different terms for the same meaning, or with different meanings for the same term.

Contents

Knowledge Diversity

Knowledge level diversity exists because the mapping between etypes and the properties used to describe etypes is many-to-many (depending on the focus).



For example, *Sarry Sky* as an instance of the entity type *painting*, can be associated with different properties such as *painted by, named as, representative, describe, classified,* etc.

Definition: Knowledge Diversity

Recall the definition of knowledge as

$$knowledge = entity\ types + properties$$

Definition (Knowledge Diversity)

We have knowledge level representation diversity (short: knowledge diversity) when we have two representations with different entity types with the same set of properties, or with different set of properties for the same entity type.

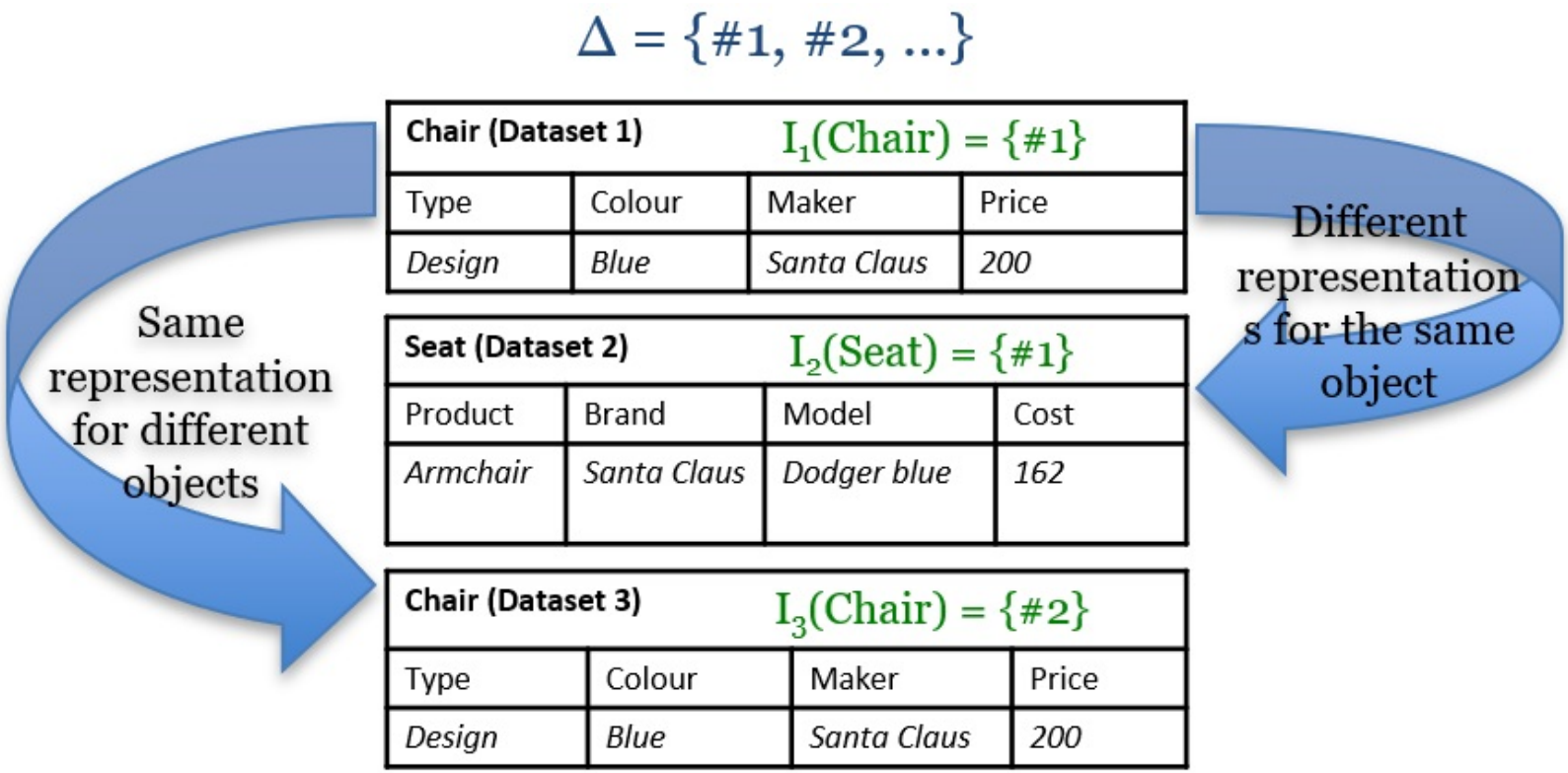
Contents

Data diversity

Data level diversity exists because the mapping between entties and the property values used to describe them is many-to-many (depending on the focus).

Data diversity

Data level diversity exists because the mapping between entties and the property values used to describe them is many-to-many (depending on the focus).



Definition: Data Diversity

Recall the definition of data as

$$data = entities + property\ values$$

Definition (Data Diversity)

We have data level representation diversity (short: data diversity) when we have two representations different entities with the same set of property values, or with different set of property values for the same entity.

KDI : Knowledge and Data Integration

W1.L12.M1.T21

Representation Diversity

Phase: 1. Introduction & Representation Diversity