



KDI ● **Knowledge and Data Integration**

Aligning Representation Diversity

Phase: 1. Introduction &
Representation Diversity

W1.L3.M1.T31

Contents

- 1 Alignment**
- 2 From Informal Language to LKGs**
- 3 From Informal Schemas to SKG**
- 4 From Informal Datasets to DKG**

Contents

- 1 Alignment**
- 2 From Informal Language to LKGs
- 3 From Informal Schemas to SKG
- 4 From Informal Datasets to DKG

Alignment

Diversity is unavoidable, then? Recall the representation diversity lies in three levels ...

- the **language** level: from informal language to concepts (what we called above identifiers), i.e. Language (level) Graphs (LKGs);
- the **knowledge** level: from informally defined knowledge to etypes + properties, i.e., Schema (level) Knowledge Graphs (LKGs);
- the **data** level: from informally defined data to entities + property values, i.e., Data (level) Knowledge Graphs (DKGs)

Alignments must be implemented in the three levels correspondingly.

Why KGs?

Representation diversity can be solved incrementally, where alignment at one level can be done independently of the other two.

- LKG offers the **terminologies** (with intended meanings) to name the entities and types in the domain.
- SKG offers the **representations** of the entity types and properties in the domain.
- DKG **populates** the entities to the types with specific values to the properties in the domain.

KG technology allows to unify into a single three-layer KG the three different forms of diversity

Contents

- 1 Alignment
- 2 From Informal Language to LKGs**
- 3 From Informal Schemas to SKG
- 4 From Informal Datasets to DKG

Language Knowledge Graph

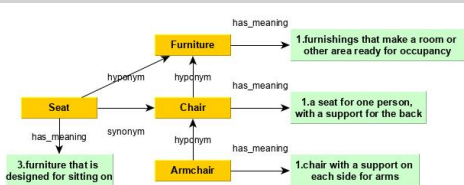
A LKG is a graph of alinguistic meanings, i.e. identifiers, e.g., #123, connected by properties correlating them, each identifier being associated to one or more words in one or more languages (thus generating synonymy and polisemy).

Definition (LKG)

$$LKG = L + P_{language}$$

where L is a set of identifiers, $P_{language}$ is a set of properties used to describe the language level relations in L , e.g., hyponymOf, hypernymOf, instanceOf.

Example of LKG:



From Informal Language to LKGs

Language terms are mapped to the unique graph of alinguistic identifiers. The goals are:

- To align terms **across** languages, such as 'monument' in English and 'monumento' in Italian.
- To align terms **within** the same language, such as 'car', 'vehicle', to 'automobile' and 'electric car' .

The alignment above is achieved by **mapping** language terms (used to denote etypes, properties, entity names and property values) to their unique identifiers and then by comparing them.

Problems to be solved: Word Sense Disambiguation (WSD), Named Entity Recognition and Detection (NER, NED).

Term Alignment Example

What does the word 'alignment' mean?

Example (alignment)

Senses from WordNet 3.1:

- (n) alliance, coalition, alignment, alinement (an organization of people (or countries) involved in a pact or treaty)
- (n) alignment (the spatial property possessed by an arrangement or position of things in a straight line or in parallel lines)
- (n) conjunction, alignment ((astronomy) apparent meeting or passing of two or more celestial bodies in the same degree of the zodiac)
- (n) alignment (the act of adjusting or aligning the parts of a device in relation to each other)

Term Alignment Example

What does the word 'alignment' mean?

Example (alignment)

Senses from WordNet 3.1:

- (n) alliance, coalition, alignment, alinement (an organization of people (or countries) involved in a pact or treaty)
- (n) alignment (the spatial property possessed by an arrangement or position of things in a straight line or in parallel lines)
- (n) conjunction, alignment ((astronomy) apparent meeting or passing of two or more celestial bodies in the same degree of the zodiac)
- (n) alignment (the act of adjusting or aligning the parts of a device in relation to each other)

Contents

- 1 Alignment
- 2 From Informal Language to LKGs
- 3 From Informal Schemas to SKG**
- 4 From Informal Datasets to DKG

Schema Knowledge Graph

An SKG is a graph where nodes are etypes and where links are relations (i.e., object properties). Etype nodes are further decorated with attributes (i.e., data properties).

Definition (SKG)

$$SKG = K + P_{knowledge}$$

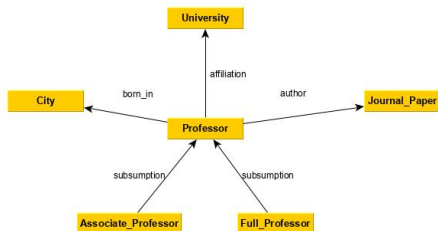
where K is a set of entity types, $P_{knowledge}$ is the set of properties used to describe and relate the etypes in K .

Logical properties:

equivalence, subsumption, ...

Model properties: authorOf, ..

A SKG is defined with respect to a LKG



From Informal Schemas to SKG

The translation of a set of informal schema to an SKG happens in two steps:

- The informally defined entity types and properties from different datasets and target queries are exploited to construct an **extended entity relationship (EER) model**, i.e, an informally defined schema, or informal schema.
- The EER model is then mapped to a **formal schema**, as close as possible to existing reference schemas (so called **telologies (ontologies)**).

Problems to be solved: term matching and alignment, being compliant with the meanings and properties induced by the reference LKG (SKG subsumption must be coherent with LKG hyponymy).

Contents

- 1 Alignment
- 2 From Informal Language to LKGs
- 3 From Informal Schemas to SKG
- 4 From Informal Datasets to DKG**

Data Knowledge Graph

A DKG is a graph where nodes are entities and where links are relations (i.e., object properties). Entity types are further decorated with attribute values (i.e., data property values).

Definition (DKG)

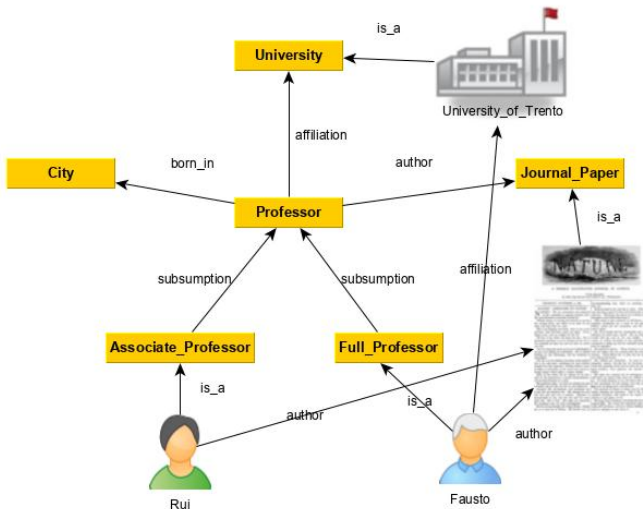
$$DKG = D + P_{data}$$

where D is a set of entities and P_{data} is the set of property values used to relate and describe the entities in D .

A DKG is defined with respect to a SKG, and via the SKG, with respect to a LKG.

Example of DKG

A DKG and corresponding SKG.



From Informal Datasets to DKG

The informally stated data in datasets are supposed to be integrated into a DKG. This requires a set of preliminary operations

- Cleaning amounts to detecting and fixing missing, invalid, incompatible. ..., format and data, and other pitfalls.
- Transforming is to change format of the informal data sets, usually in tabular form, often generated from unstructured or semi-structured data into the reference DKG format.

Problems to be solved: proper data cleaning and transformation, Natural Language Processing if needed, entity and property value matching and alignment, being compliant with the meanings induced by the reference LKG and the etypes and properties induced by the reference SKG.



W1.L3.M1.T31



Aligning Representation Diversity

Phase: 1. Introduction &
Representation Diversity