How to Represent Knowledge Diversity

Andreas Thalhammer[1], Ioan Toma[1], Rakebul Hasan[1], Elena Simperl[2], and
Denny Vrandecic[2]

[1] STI Innsbruck, University of Innsbruck, Austria, `firstname.lastname@sti2.at`
[2] AIFB, Karlsruhe Institute of Technology, Germany, `firstname.lastname@kit.edu`

Abstract. Information on the Web includes a huge diversity of opinions, viewpoints, sentiments, emotions, and biases. Accordingly, more and more methods, techniques and tools are available to extract these semantics from text. Representation and exchange of diversity-related information can be easily supported by the use of semantic technologies. For this, we introduce the Knowledge Diversity Ontology (KDO).

1 Introduction

With the advancements of natural language processing (NLP), more and more methods, techniques and tools are available that aim to extract meaningful knowledge from the Web. Some of them use semantic technologies, in particular ontologies to represent and exchange diversity-related information.

Of course, several ontologies have already been developed: the metadata specification DCMI,[1] an ontology for representing Web forums (SIOC),[2] a vocabulary for modeling provenance (OPMV),[3] and a model for linking knowledge organization systems (SKOS).[4] However, none of these vocabularies include concepts or properties for representing and linking opinions, viewpoints, sentiments, emotions, and biases. The Knowledge Diversity Ontology (KDO) introduces these components, but also reuses concepts and properties from existing vocabularies such as SIOC and FOAF.

The RENDER project[5] has its particular focus on enabling and retrieval of knowledge diversity. For this, we combine expertise in NLP (including named entity recognition), representation of knowledge diversity, and storage and retrieval. A driving force in this context are use cases provided by Telefonica,[6] Wikipedia,[7] and Google News.[8] Each of these companies provide the project with a different angle on knowledge diversity, which enables us to keep the ontology as general as possible and at the same time as specific as necessary.

---

[1] DCMI - `http://dublincore.org/`
[2] SIOC - `http://sioc-project.org/`
[3] OPMV - `http://open-biomed.sourceforge.net/opmv/ns.html`
[4] SKOS - `http://www.w3.org/2009/08/skos-reference/skos.html`
[5] `http://www.render-project.eu/`
[6] `http://www.telefonica.com/`
[7] `http://www.wikipedia.org/`
[8] `http://www.google.com/`

2 The Knowledge Diversity Ontology

In this section we first introduce a scenario that exemplifies one of the use cases that have to be addressed. Afterwards, the components of KDO[9] are introduced. The paper concludes with a KDO representation of the scenario.

2.1 Use Case Scenario

The following scenario is drawn from the Google News use case. Note that the described scenario is completely fictional.

Michael, a professional business analyst and broker, works for a bank in Frankfurt. In recent times, there are speculations about a probable takeover of an Italian automotive manufacturer (Bigat) by a German one (FMW). In the major business newspapers and rating agencies, there is speculation that this takeover will not happen because of doubts expressed in a public meeting in Turin by the chairman of the German company (Ferdinand Maier). Michael has assigned the task to write an own estimation about this issue. Therefore, he types "Bigat FMW maier" into the search field of "GoogleR" (Render powered version of Google). After execution, a selection of news articles that relate to these search terms is retrieved from various sources (e.g. www.bloomberg.com). Michael chooses only to select articles that contain "Ferdinand Maier" in the people set and "Bigat" and "FMW" in the organizations set. Moreover, as he knows that the doubts were expressed in Turin he also selects that this location should be mentioned in the relevant articles. Finally, he chooses to read only articles that were authored in the UK. Ten articles in English language are left and Michael chooses to read the article with the most positive sentiment score that is authored by James McDonald. In the authors opinion, "the takeover will happen and benefits both parties". The chosen article is written in English and it was published by BBC London on 23rd of March 2011.

2.2 KDO Components

The design of the KDO (see figure 1) builds on concepts and properties of existing ontologies, namely SIOC and FOAF: important elements are `sioc:Post` and `foaf:Agent`. The core concepts of the KDO are `kdo:Opinion`, `kdo:Sentiment`, `kdo:Polarity`, and `kdo:Bias`. An opinion can have one or more opinion expressions (`kdo:hasOpinionExpression`) and can be linked to an emotion (`kdo:hasEmotion`). The opinion as well as a post can mention named entities (`kdo:mentions`) and can have a sentiment (`kdo:hasSentiment`). Moreover, it is possible to express which opinion has been stated in a post by an agent (`kdo:OpinionStatement` connected through `kdo:hasOpinion`, `kdo:hasAgentOpinion`, and `kdo:isOpinionHeldBy`). Sentiments are linked to polarities by `kdo:hasPolarity`. There exist three instances of `kdo:Polarity`: `kdo:positivePolarity`, `kdo:neutralPolarity`, and `kdo:negativePolarity`.
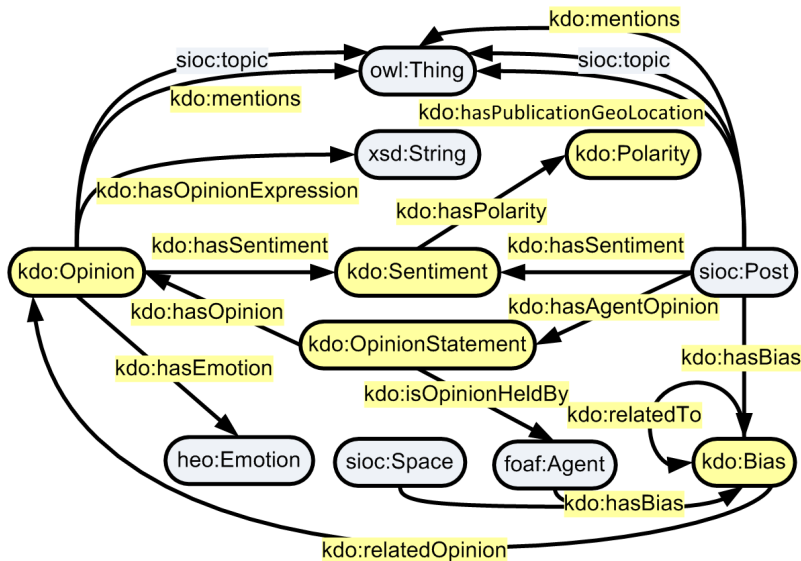
---

[9] http://kdo.render-project.eu/

Fig. 1. Visualization of the core concepts of the KDO.

A bias can be related to a SIOC post/space or an agent (`kdo:hasBias`). Also, each bias is related to one or more opinions (`kdo:relatedOpinion`) and different biases can be related to each other (`kdo:relatedTo`). SIOC covers blogs but not the similar concept of "NewsArticle". We include `kdo:NewsArticle` as a subclass of `sioc:Post` and accordingly `kdo:NewsWebsite` as a subclass of `sioc:Container`. DCMI covers the property "created" in terms of "something has been created at a point of time" but not a property that describes the location of the creation likewise. Therefore, we introduce the property "kdo:hasPublicationGeoLocation". Thus spatial and temporal information about various posts can be specified. Besides SIOC and FOAF, we make use of DCMI, PRO[10] and HEO.[11]

2.3 Representing the Use Case Scenario

The use case scenario of section 2.1 can be modeled with the KDO (Turtle syntax)[12] as follows:

In context to the chosen news article, two different persons can be identified. Note that one of which is the author and the other person is mentioned in the document:

```
<#FerdinandMaier> a foaf:Person ;
```

---

[10] PRO - `http://vocab.ox.ac.uk/pro/pro`

[11] HEO - `http://semedia.dibet.univpm.it/heo/heo.owl`

[12] `http://www.w3.org/TeamSubmission/turtle/`

```
      foaf:firstName "Ferdinand" ; foaf:lastName "Maier" .
<#JamesMcDonald> a foaf:Person ;
    foaf:firstName "James" ; foaf:lastName "McDonald" .
```

Similarly, the publisher of the document and the two mentioned car manufacturers are represented as `foaf:Organization`:

```
<#BBC> a foaf:Organization ; foaf:name "BBC" .
<#FMW> a foaf:Organization ; foaf:name "FMW" .
<#Bigat> a foaf:Organization ; foaf:name "Bigat" .
```

The opinion that is expressed in the document is denoted as follows:

```
<#Opinion1> a kdo:Opinion ;
    kdo:hasOpinionExpression "the takeover will happen
    and benefits both parties" .
```

The document itself relates to different blank nodes: the associated author and publisher that are linked through PRO, the sentiment, the authors opinion, and the SIOC container. Moreover, terms of DCMI are used to describe the creation date as well as the language.

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle ;
  pro:isRelatedToRoleInTime [ a pro:RoleInTime ;
    pro:withRole pro:author ; pro:isRoleHeldBy <#JamesMcDonald>] ;
  pro:isRelatedToRoleInTime [ a pro:RoleInTime ;
    pro:withRole pro:publisher ; pro:isRoleHeldBy <#BBC>] ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:positivePolarity] ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:isOpinionHeldBy <#JamesMcDonald> ;
    kdo:hasOpinion <#Opinion1>] ;
  sioc:hasContainer [ a kdo:NewsWebsite ;
    sioc:has_space <http://www.bloomberg.com/] ;
  dcterms:created "2011-03-23" ;
  kdo:hasPublicationGeoLocation <http://sws.geonames.org/2635167/> ;
  dcterms:language "en-UK" ;
  kdo:mentions <http://sws.geonames.org/3165524/> ;
  kdo:mentions <#FerdinandMaier> ; kdo:mentions <#FMW> ;
  kdo:mentions <#Bigat> .
```

2.4   Conclusions

In this paper we introduced the Knowledge Diversity Ontology (KDO) for representing diversity-related information on the Web. KDO covers features such as opinions, polarity, sentiments and biases, and reuses other relevant ontologies such as SIOC, FOAF and DublinCore. KDO follows a simple and extensible approach to modeling introducing a minimal set of concepts and properties that are needed to model diversity. To illustrate how KDO can be used a relevant example is described and modeled using the terminology provided by the ontology.