



RENDER
FP7-ICT-2009-5
Contract no.: 257790
www.render-project.eu

RENDER

Deliverable D3.1.1

Initial models for diversity-rich information

Editor:	Andreas Thalhammer, UIBK; Ioan Toma, UIBK
Author(s):	Andreas Thalhammer, UIBK; Ioan Toma, UIBK; Rakebul Hasan, UIBK
Deliverable Nature:	Report (R)
Dissemination Level: (Confidentiality)	Public (PU)
Contractual Delivery Date:	September 2011
Actual Delivery Date:	September 2011
Suggested Readers:	all consortium members, everybody interested in how to describe diversity information
Version:	1.1
Keywords:	Knowledge diversity ontology, diversity models

Disclaimer

This document contains material, which is the copyright of certain RENDER consortium parties, and may not be reproduced or copied without permission.

In case of Public (PU):

All RENDER consortium parties have agreed to full publication of this document.

In case of Restricted to Programme (PP):

All RENDER consortium parties have agreed to make this document available on request to other framework programme participants.

In case of Restricted to Group (RE):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement. However, all RENDER consortium parties have agreed to make this document available to <group> / <purpose>.

In case of Consortium confidential (CO):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the RENDER consortium as a whole, nor a certain party of the RENDER consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Full Project Title:	RENDER – Reflecting Knowledge Diversity
Short Project Title:	RENDER
Number and Title of Work package:	WP3 Diversity representation and processing
Document Title:	D3.1.1 - Initial models for diversity-rich information
Editor (Name, Affiliation)	Andreas Thalhammer, UIBK; Ioan Toma, UIBK
Work package Leader (Name, affiliation)	Ioan Toma, UIBK

Copyright notice

© 2010-2013 Participants in project RENDER

Executive Summary

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. This deliverable introduces the Knowledge Diversity Ontology (KDO), an ontology for representing diversity-related information on the Web. It includes concepts and properties for representing and linking opinions, viewpoints, sentiments, emotions, and biases. KDO heavily uses terminology provided by other well known existing vocabularies such as SIOC and FOAF.

To motivate the need and importance of having defined a knowledge diversity ontology we describe first a set of scenarios that are related to Wikipedia entries and edits, blogs, news articles, public and corporate forums, surveys, emails, and tweets. These scenarios are inspired by all the project use cases and are used to derive a set of competency questions that help us identify the key concepts and properties of the Knowledge Diversity Ontology.

The design of KDO builds on concepts and properties of existing ontologies, namely SIOC and FOAF using elements such as `sioc:Post` and `foaf:Agent` from these ontologies. Concepts such as `kdo:Opinion`, `kdo:OpinionStatement`, `kdo:Sentiment`, `kdo:Polarity`, and `kdo:Bias` are central to our ontology. 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` which can be positive, neutral or negative.

As previously stated, KDO reuses many existing ontologies. This deliverable identifies the ontologies that are relevant in the context of modelling diversity, describes the fragments of these ontologies that are actually used, and illustrates using snippet examples from the modelled scenarios, the interplay of KDO and reused ontologies to model diversity. To improve the reader's understanding, three of the motivating scenarios, of which each relates to one of the project use cases, are formally described in RDF using KDO and the reused ontologies. We also provide a SPARQL formalization of the competency questions that can be tested with knowledge bases that use the KDO terminology.

Table of Contents

Executive Summary	3
Table of Contents	4
List of Figures	5
Abbreviations	7
Definitions	8
1 Introduction	9
2 Motivating Scenarios.....	10
2.1 News Articles.....	10
2.1.1 Scenario 1	10
2.1.2 Scenario 2	11
2.2 Collaborative Knowledge Bases	11
2.2.1 Scenario 3	12
2.2.2 Scenario 4	12
2.3 Customer Feedback	12
2.3.1 Scenario 5	13
2.3.2 Scenario 6	13
2.3.3 Scenario 7	13
3 Scope of the Knowledge Diversity Ontology.....	15
3.1 Competency Questions	16
4 The Knowledge Diversity Ontology (KDO).....	17
4.1 The Core Ontology	17
4.1.1 KDO Classes.....	18
4.1.2 KDO Properties	21
4.2 Reused Ontologies and Standards	25
4.2.1 The Friend of a Friend (FOAF) Vocabulary.....	25
4.2.2 The Dublin Core Metadata Initiative	26
4.2.3 The Semantically-Interlinked Online Communities (SIOC) Ontology	27
4.2.4 PRO, the Publishing Role Ontology.....	29
4.2.5 Human Emotions Ontology (HEO)	29
4.2.6 GeoNames	30
4.2.7 dmoz (open directory project).....	30
5 Using the Knowledge Diversity Ontology.....	31
5.1 Modelling Motivating Scenarios with KDO	31
5.2 Answering Competency Questions with SPARQL	34
6 Conclusion	37
References.....	38
Annex A KDO formulated in OWL.....	39

List of Figures

Figure 1: The Knowledge Diversity Ontology – classes and properties.....	18
Figure 2: Main classes and properties in SIOC	28
Figure 3: The individuals of the pro:Role class representing the publishing roles.....	29
Figure 4: The classes defined by HEO (source [3])	30

List of Tables

Table 1: Competency Questions.....	16
Table 2: DCMI metadata terms	26
Table 3: SPARQL queries answering the competency questions	34

Abbreviations

DCMI	Dublin Core Metadata Initiative
FOAF	Friend Of A Friend
HEO	Human Emotions Ontology
HTTP	Hypertext Transfer Protocol
HTML	Hypertext Markup Language
OWL	Web Ontology Language
KDO	Knowledge Diversity Ontology
NER	Named Entity Recognition
PRO	Publishing Role Ontology
REST	Representational State Transfer
RDF	Resource Description Framework
SIOC	Semantically-Interlinked Online Communities
URL	Uniform Resource Locator
XML	Extensible Markup Language

Definitions

A glossary which introduces the terminology that is used in this deliverable can be found in Hasan et al. "Towards a Knowledge Diversity Model" [2].

1 Introduction

The Web is a facilitator and catalyst for the publication, use and exchange of information, fostering a global network of news, stories and statements which represent an amazing diversity of opinions, viewpoints, mind sets and backgrounds. An important component for capturing inter and intra-document diversity is a comprehensive knowledge model for representing diversity that reflects the plurality of opinions and viewpoints.

The Knowledge Diversity Ontology (KDO) aims at providing a vocabulary to describe different aspects of knowledge diversity such as opinion, sentiment, and emotion. The core part of the ontology allows describing concepts such as bias, opinion, sentiment, polarity, etc. Existing ontologies are reused to describe concepts such as individuals, organizations, locations, emotions, and publications. For example: FOAF¹ ontology is used to describe people and organizations; additionally, DCMI², PRO³, SIOC⁴, and HEO⁵ are reused. The KDO will be accessible in two ways: cognitively to human users as well as computationally to machines. The semantic representation will be then leveraged for improving the selection and ranking of content, and the presentation to users. In RENDER, selection and ranking will go beyond widely adopted approaches based on popularity or personalization, and take opinions and viewpoints into account when computing the relevance of results.

In the RENDER project, as a first step to capture knowledge diversity, the considered content such as articles, (micro-) blog entries, emails, surveys, or news feeds (provided by our use case partners of WP5) are processed with state-of-the art natural language processing technology that is enhanced in WP2. The extracted knowledge is then transformed into a semantic representation (the KDO) and collected in special data stores that are developed in WP1. A toolkit (that two of our partners develop in WP4) will exemplify the benefits of the integration of work packages one, two and three.

The rest of this document is organized as follows. Section 2 contains a collection of motivating scenarios in which diversity is a central feature and which motivate the need and importance of having a well defined knowledge diversity ontology. Section 3 defines the scope of the KDO and a set of competency questions that are used to identify query scenarios that have to be considered and addressed. Section 4 contains the description of the ontology itself including descriptions of new classes and properties on the one hand and reused ontologies on the other. Section 5 exemplifies how the ontology can be used with respect to modelling and also querying. Section 6 concludes the deliverable.

¹ <http://xmlns.com/foaf/spec/>

² <http://dublincore.org/documents/dcmi-terms/>

³ <http://purl.org/spar/pro/>

⁴ <http://sioc-project.org/>

⁵ <http://semedia.dibet.univpm.it/heo/heo.owl>

2 Motivating Scenarios

This section contains a collection of scenarios that motivate the need and importance of having defined a knowledge diversity ontology. These scenarios are related to Wikipedia entries and edits, blogs, news articles, public and corporate forums, surveys, emails, and tweets. Moreover, we ground the competency questions of section 3.1 on these story lines.

Please note that the organizations and people described in the scenarios are completely fictional.

The rest of this section is structured as follows: The news articles use case defined by Google, for which a motivating scenario is presented in section 2.1, relates to diversity mining in order to present a variety of opinions about the same topic. The collaborative knowledge bases use case of Wikimedia, outlined in section 2.2, exemplifies opinion mining in environments where a variety of opinions have to be integrated. Last but not least, scenarios for the customer feedback use cases specified by Telefonica are explained in section 2.3.

2.1 News Articles

News articles published by newspapers and their Web counterpart are usually created by well-qualified journalists and cover a large variety of topics. These professionally created news articles along with articles created by Web users such as blogs offer a rich diversity of topics, opinions and viewpoints; e.g. two articles addressing the same event can represent very different viewpoints that expose different levels of bias.

In the RENDER project, the diversity-aware selection and ranking of the news articles in the Google News case study relates to this scenario. The news aggregator service of Google (Google News) indexes more than ten thousands of news Web sites which originate from more than forty regional issues and more than 15 different language areas. The considered news content is created by professional journalists and by Web users and offers a rich diversity of information. Current ranking algorithms result in news summaries that are dominated by popular viewpoints or opinion holders such as large news agencies. Alternative opinions or arguments from smaller publishers often disappear and do not reach the interested audience. Consequently, even though Google aims for wide and comprehensive news coverage, the selection of presented viewpoints is often biased. Representation of different aspects of diversity in the news articles will enable the development of diversity-aware ranking of news posts for covering the most diverse viewpoints on a particular topic. Moreover, the data can be enriched with other information from different sources like blogs, tweets, and wiki pages. The representation and processing of the diversity-aware information is expected to considerably increase the interconnection between diversifying news and discussions on the Web.

In the following we demonstrate two scenarios which outline the advantages of mining diversity in news and blog articles. Both scenarios are aligned with example 1.1 of D2.2.1 which provides an overview of data attributes that can be extracted from news articles. The first scenario outlines the view of a broker who wants to get an overview of recent news and blog articles about the potential takeover of a publically tradable company by another one. In the second scenario a journalist wants to find very biased opinions about a political topic that she wants to discuss more neutrally in her own article.

2.1.1 Scenario 1

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”. 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 GGC London on 23rd of March 2011.

2.1.2 Scenario 2

Katharina is a professional journalist holding a contract with Singer AG, Hamburg, Germany. Recently, she received a request to write an article about the application of the city of Munich for the Olympic Winter Games 2018. This topic is very controversially discussed, as politicians and landowners are not able to come to an agreement about the locations of the areas of competition. The newspaper, in which Katharina is publishing, has a high interest in releasing non-polarized articles. Therefore, Katharina tries to get an insight into the different opinions related to the ongoing discussions using the tool “GoogleR”. Katharina types in the search field “olympic winter games 2018 munich”. When the results are displayed, Katharina selects the topic “Olympic Winter Games 2018” which only contains news articles about this specific topic. The results are updated. Of course, non-local newspapers do not focus on the discussions in Munich as much as local ones do. So she has the option to select only “Munich” as the origin of relevant articles and she can also group the articles by source (e.g. www.sueddeutsche.de). The results are updated again. Also, as the discussion has evolved to a certain point, Katharina chooses to browse only articles that were published within the last two weeks. The results are updated. Articles with a neutral sentiment score should be omitted so Katharina switches the checkbox “neutral” off. Finally, Katharina can reorder the articles according to positive and negative sentiments and chooses to read an angry blog post in German with the most negative sentiment score that backs the viewpoints of the landowners. In order to stay neutral, she also reads an article published by “Morgenzeitung München” that advertises the benefits of “München 2018” for the region.

2.2 Collaborative Knowledge Bases

An example of a collaborative knowledge base is the widely known Wikipedia, an online encyclopaedia created by the collaborative contribution of authors and editors from around the world with diverse backgrounds. It is often the case that the individual viewpoints of the contributors are expressed in the knowledge created in this fashion. This introduces the problem of biased coverage of different topics. Furthermore, often, inconsistencies of knowledge can be found throughout different editions of the articles, i.e. articles that exist in different languages. In the RENDER project, the Wikipedia case study relates to this scenario.

Despite efforts for a balanced coverage at Wikipedia, biases influenced by the individual views of the more than 100,000 volunteer contributors have been introduced. The increasing complexity of the control processes for creating and editing articles that are put in place to overcome the problem of biases, negatively impacts the growth of Wikipedia. Edit conflict resolution, arbitration committees, banning policies, a complex hierarchy of contributors, editors, and administrators is slows down the process of article creation and diversification. Effectively, recent statistics show that the number of new articles has been decreasing dramatically over the past years, while the number of edits is still growing steadily⁶. Shifting missing content from one language version of a Wikipedia article to another, or the detection of diverse viewpoints within a topic or article are needed to support the editorial team in managing and

⁶ <http://en.wikipedia.org/wiki/Wikipedia:Statistics>

encouraging large-scale participation and growth. Diversity-empowered services such as quality or reliability assessment of an article or a specific statement, conflict resolution, anomaly detection, and cross-lingual consistency checking are expected to considerably improve the way information is currently managed in Wikipedia.

For Wikipedia there are four use cases specified in RENDER⁷. We combine these use cases in two scenarios. The first scenario in the Wikipedia use case focuses on a Wikipedia editor who wants to get a collection of articles that are very biased (UC1) and that leave space for extension (UC2). In the second scenario a journalist wants to find out about how the topic of a Wikipedia article is described in various languages (UC1), changed over time in order to estimate the direction (UC3), and obtain statistics about how often edits of a particular institution took place (UC4).

2.2.1 Scenario 3

Thomas is a frequent user of the Czech Wikipedia and he likes to edit and create articles in his first language (Czech). As he studies medicine in his fourth year, he wants to focus his contributions in his area of specialization which is immunosuppressants. With the help of Wikipedia's tool "Editors Guide" he discovers that only few articles in this area exist in Czech language and moreover, the main article "Imunosupresivum" is very biased (only one opinion with a negative sentiment and eight versions of seven different users exist) and was last edited on 2nd of February in 2009. The seven current editors actually belong to a group of users that all like to edit articles related to medicine. The last edit was from user "drogomir" and from the area of Plzeň. As the same article is much more extensive in different languages such as English and German, a new structure for the Czech version is suggested automatically. Several articles that relate to the same topic and do not exist in Czech language are suggested to be created by him.

2.2.2 Scenario 4

The European sovereign debt crisis has been discussed unequally in different countries of the European Union. Nicola works for the news agency BNN International and wants to investigate the different opinions of Wikipedia entries that relate to the most affected countries. With the help of Wikipedia's tool for bias detection that mines several internal, external and multilingual sources, she detects that the article in Spanish is biased focusing on Ireland and Greece as the main bailout candidates while listing Spain at the section for "Other affected countries". The tool also detects that the topic of all language versions of the same article changed over time: while in May 2010 there was a strong focus on Greece as a bailout candidate, it appears that more and more contributions for the other affected countries were added (a collection of mentioned locations can be extracted). The tool supports to focus on the German version and Nicola discovers that certain edits with a negative sentiment, from the user "berkel", origin in governmental IP ranges from the area of Berlin.

2.3 Customer Feedback

Recent growth of the social media platforms contributed significantly to the amount of content in the Web. These user generated contents contain very diversity-rich information. The social media platforms include channels such as social networking sites like Facebook, micro blogging sites like Twitter, blogs, discussion forums and question answering websites etc. Users express their opinion and viewpoints about various topics through these channels. Topics can range from political attitudes over personal interests to opinions about specific products or services. As these media are designed to be used in a social manner, comments and answers from other users are likely to represent different viewpoints. Many companies consider comments on their products or services in these channels as a central element of nowadays customer

⁷ Use Cases Wikipedia, Presentation 19-20 April 2011, Ljubljana

relationships management (CRM). Of course, users of social media usually are just a fraction of all customers. More traditional sources of customer feedback, such as surveys, call centre transcriptions, and customer emails, help to give a complete overview of customer opinions.

In the RENDER project, the Telefonica case study relates to the integration of internal and external sources of customer feedback. Moreover, various new directions of opinion mining are defined by this area of application. In deliverable D5.3.1, three use cases that exemplify the need and opportunities of source integration in combination with automatic opinion mining are described. The storyboards of these use cases serve as basis for the following motivating scenarios that relate to opinion mining over customer feedback.

2.3.1 Scenario 5

The customer relationship management (CRM) of the company Movistar wants to know the different opinions about a new tariff, called “Tarifa Autónomos”, which was introduced one month ago. The analysis focuses on the most recent Twitter data crawl as the source of information. Ana, a CRM employee of Movistar, first checks whether the new tariff is contained in the trend topics. Afterwards, in the opinion search she selects Twitter as the source and enters “Tarifa Autónomos” in the keyword search field. She selects the last month as a date period and also Spanish as the language of the sources. After execution, Ana receives positive and negative comments that have attached percentage values according the number of all comments that relate to the keywords. The result screen also offers various ways of clustering the results according to source-specific features which, in the case of Twitter, can be date, geolocation, or topic. On the detailed results page, Ana can take a look at the comment enriched with information about the user, the sentiment orientation the opinion embodies, the emotion of the sentiment, and additional data that the particular source of information provides about the comment. Finally, Ana corrects the sentiment orientation of one of the comments.

2.3.2 Scenario 6

The company O2 Ireland has started to have presence on Twitter. John, from the department of Corporate Reputation, is responsible for analysing customer comments of various channels of feedback. For this, he needs to consider internal as well as external sources. The data also includes information about the last update of the dataset (“Data collection information”) of each source. John discovers that the Twitter source includes Tweets until the last week and other online sources don’t have new data within the last two weeks. John decides to select Twitter and email data from 2 month ago to one week ago. He wants to check on the most important topics in English from the two data sources. The minimum number of opinions that define a “trend topic” can be configured by the user. In the trends search, grouping by source type or user group is possible. John chooses to group by online users. John is also able to select a score range before executing the search (and he selects the default range which is 0 to 100 – indifferent score). At the execution of the search, the sentiment analysis omits neutral comments and computes the percentages of negative and positive comments and displays them. John is able to cluster the results according to common features of Twitter and email (date, location, and trend topic). On the “Detailed Results” page, John is able to get more information about the user, sentiment orientation plus score, and the source types.

2.3.3 Scenario 7

The Business Intelligence (BI) unit of O2UK focuses on the analysis of data interaction between the customers and the CRM. Jane, a BI employee, uses the Opinion Mining tool to see information about available data of the email and survey sources (email: until last week; surveys: last month). She decides to define a search within the email and survey sources of the last month. Jane has a look at the trend topics of

topics in English. Jane finds the topic “Price” interesting and she decides to search negative comments (score 0-40) in English language that were provided in the last month grouped by source type. After execution of the search, Jane has the option to cluster the results by features common to the data sources. In this scenario (surveys⁸ and emails⁹ in English of the last month), the overlapping features are: product type, date, and location. On the “Detailed Results” page, Jane is able to get more information about the user, sentiment orientation plus score, and the source types.

⁸ Features of the survey source: product name, contract type, location, store, operator id, and date

⁹ Features of the email source: contract type, date, and location

3 Scope of the Knowledge Diversity Ontology

The Web is an unprecedented enabler for publishing, using and exchanging information at global scale. Virtually any topic is covered by a large diversity of opinions, viewpoints, mind sets and backgrounds. As an example, a news article published in a news publisher website can mention different locations, people, and organizations. An article can have a positive or negative sentiment. An article can present an entity, such as an organization, negatively and another entity positively. The purpose of Knowledge Diversity Ontology (KDO) is to represent these different aspects of diversity in the Web. Applications that focus on an improvement of organization, navigation, and presentation of diversity-aware Web information will use the KDO.

Deliverable 2.2.1 (Prototype of the Fact Mining Toolkit) discusses the dimensions of diversity that can be mined in RENDER. These are split up in two groups. The dimensions of the **content diversity** that can be detected by analyzing texts in articles published in mainstream news sites, blogs, forums, Twitter, etc. are:

- **Topic:** the topics of the items.
- **Social:** people and organization involved in the publication and authoring of the articles and their social influence; mentioned and/or connected people and organizations in the texts.
- **Geographical:** geographical locations addressed in the articles; geographical locations of the authors.
- **Opinion:** opinions expressed in the texts.
- **Sentiment:** sentiment of a particular article or a particular entity.
- **Bias:** difference in sentiment, vocabulary and mentioned entities in the articles about the same event.
- **Knowledge:** difference in fact coverage, facts themselves and entity relations.
- **Cross-lingual and multi-lingual:** diversity of the other diversity dimensions across languages.
- **Context:** background knowledge, temporal trends and contrast between different sources (e.g., mainstream news vs. blogs vs. Twitter).

The understanding of diversity based on the dimensions of content diversity is considered to drive the development of the KDO. Based on this, the main concepts regarding knowledge diversity are topics, opinions, sentiments, biases, documents, people and organizations etc. The KDO interlinks these main concepts. Detecting absolute truth using the current technology is not an easy task. Therefore in RENDER, statements are considered as opinions expressed in digital resources such as Web documents and authored by agents such as people.

In D2.2.1 also the activities of users, such as passive readers and prosumers, are introduced to reflect the **usage diversity**:

- **Demography:** demographical information of users reading the contents.
- **Topic:** topics of interest of the users reading the contents.
- **Geography:** geographical location of the reader.
- **Access method:** means of accessing the contents by the users (e.g., web browsers, mobile browsers, email, phone).
- **Social context:** social graph of the reader community and possible influence within the communities.
- **Time:** time of the users accessing the content and changes of the user activities through the time.

This different kind of knowledge is (currently) not addressed by the KDO as the use cases (its main foundation) do not specifically cover the recording of user information and behaviour.

3.1 Competency Questions

The concepts describing the different aspects of diversity of information on the Web will appear in the KDO. The competency questions for the KDO emerge from a reflection of the motivating scenarios of section 2 to the dimensions of diversity. As a result, the KDO should enable the creation of knowledge bases that can answer the following questions:

Table 1: Competency Questions

Number	Question	Related Scenarios
1	What is the URL of a document?	All
2	What is the Source of a document (e.g. www.bloomberg.com)?	All
3	What is the type of a document (news article, blog, wiki article, etc.)?	All
4	Which topics, keywords, locations, organizations, and people occur in a Web document?	All
5	What is the language of a document?	All
6	What is the publication date of a document?	All
7	Who are the contributors (with roles) of a document (e.g. author, editor, publisher, etc.)?	All
8	What is the location of contribution of a document (e.g. London)?	1, 2, 3, 4, 5
9	What are the groups of a particular contributor ?	3, 6
10	What are the biases of a document?	3, 4
11	What other versions of the same document do exist (i.e. language, earlier version, etc.)?	3, 4
12	What are the opinions in a document?	1, 3, 5, 6, 7
13	Who is the holder of an opinion ?	1, 3, 5, 6, 7
14	What is the sentiment of a particular opinion ?	3, 4, 5, 6, 7
15	What is the sentiment of a document?	1, 2, 6, 7
16	What is the polarity of a sentiment ?	1, 2, 4, 5, 6, 7
17	What are the emotions of a sentiment ?	2, 5
18	What are the different expressions of a particular opinion ?	5, 6, 7

4 The Knowledge Diversity Ontology (KDO)

The Knowledge Diversity Ontology aims at providing a vocabulary that describes different dimensions of knowledge diversity of the Web. To support the representation of diversity information, the conceptual model of the Knowledge Diversity Ontology includes concepts and relations that were identified and modelled by focusing on the use case scenarios and competency questions described in the previous sections.

The ontology includes concepts such as **post**, representing human-created documents; **agent**, representing people, organizations or groups; **opinion**, representing the opinions in the documents; **opinion statement**, for relating opinions in a ternary relationship to posts and agents, **sentiment**, representing the sentiments in a document or opinion; **emotion**, representing the emotions that relate to opinions; **polarity**, representing sentiment orientation; **bias**, representing biases related to documents, agents and spaces; **role**, representing the role of a contributor agent in a document. An opinion can have one or more opinion expressions and can be linked to an emotion. The opinion as well as a post can mention named entities and can have a sentiment. Moreover, it is possible to express which opinion has been stated in a post by an agent. Sentiments are linked to polarities which can be neutral, negative, and positive. A bias can be related to a SIOC post/space or an agent. Also, each bias is related to one or more opinions and different biases can be related to each other. 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**.

KDO reuses vocabulary of FOAF, DCMI, PRO, HEO, and SIOC. To refer to specific topics or locations, we suggest to use **dmoz**¹⁰ or **GeoNames**¹¹ URIs respectively. In the following two sections we describe the core part of the RENDER ontology and how existing ontologies are reused.

4.1 The Core Ontology

The conceptual model of the Knowledge Diversity Ontology illustrated in Figure 1. The Knowledge Diversity Ontology is built around the concepts **agent (foaf:Agent)**, **space (sioc:Space)**, **post (sioc:Post)**, **opinion (kdo:Opinion)**, **opinion statement (kdo:OpinionStatement)**, **sentiment (kdo:Sentiment)**, **polarity (kdo:Polarity)**, **bias (kdo:Bias)**, **emotion (heo:Emotion)**, and more generally, things (**owl:Thing**). An agent is represented by the **foaf:Agent** class. Agents are the opinion holders which is represented by **kdo:isOpinionHeldBy**. Agents play various publishing roles such as author and editor etc., which is represented by **pro:RoleInTime**, a class representing the ternary relationship between a document, an agent and the publisher role the agent holds. Human-created documents are represented by the **sioc:Post** class that defines a variety of subclasses such as (micro-) blog and forum posts and Wiki articles that also relate to their respective information containers (**sioc:Container**). Unfortunately the similar concept of news article is not covered as a subclass of **sioc:Post** and therefore we extend the subclass set by **kdo:NewsArticle**. Accordingly, we introduce **kdo:NewsWebsite** as a subclass of **sioc:Container**.

A document can have a geographical publication location which is modelled by the property **kdo:hasPublicationGeoLocation**. A document can contain multiple opinions. The concept of opinion is represented by the **kdo:Opinion** class. The ternary relationship between the documents, agents, and opinions is accumulated by the **kdo:OpinionStatement** class. Opinions consist of textual opinion expressions, represented by the **rdf:hasOpinionExpression** data property. Also, opinions can have emotions attached which is modelled by the **kdo:hasEmotion** property. The class **kdo:Opinion** as well as **sioc:Post** can mention (**kdo:mentions**) or focus (**sioc:topic**) on things (**owl:Thing**). Sentiments are represented by the **kdo:Sentiment** class. The **kdo:hasSentiment** property is used to attach a sentiment to a post or an opinion. In our model a sentiment has a polarity attached to it which can be positive, negative or neutral.

¹⁰ <http://www.dmoz.org/>

¹¹ <http://www.geonames.org>

Posts, agents, and spaces (e.g. web sites) can have a bias which is modelled by the **kdo:hasBias** property. Different biases can be related to each other (**kdo:relatedTo**) and do also relate to opinions (**kdo:relatedOpinion**).

In rest of this subsection, we describe each of the classes and properties in the Knowledge Diversity Ontology.

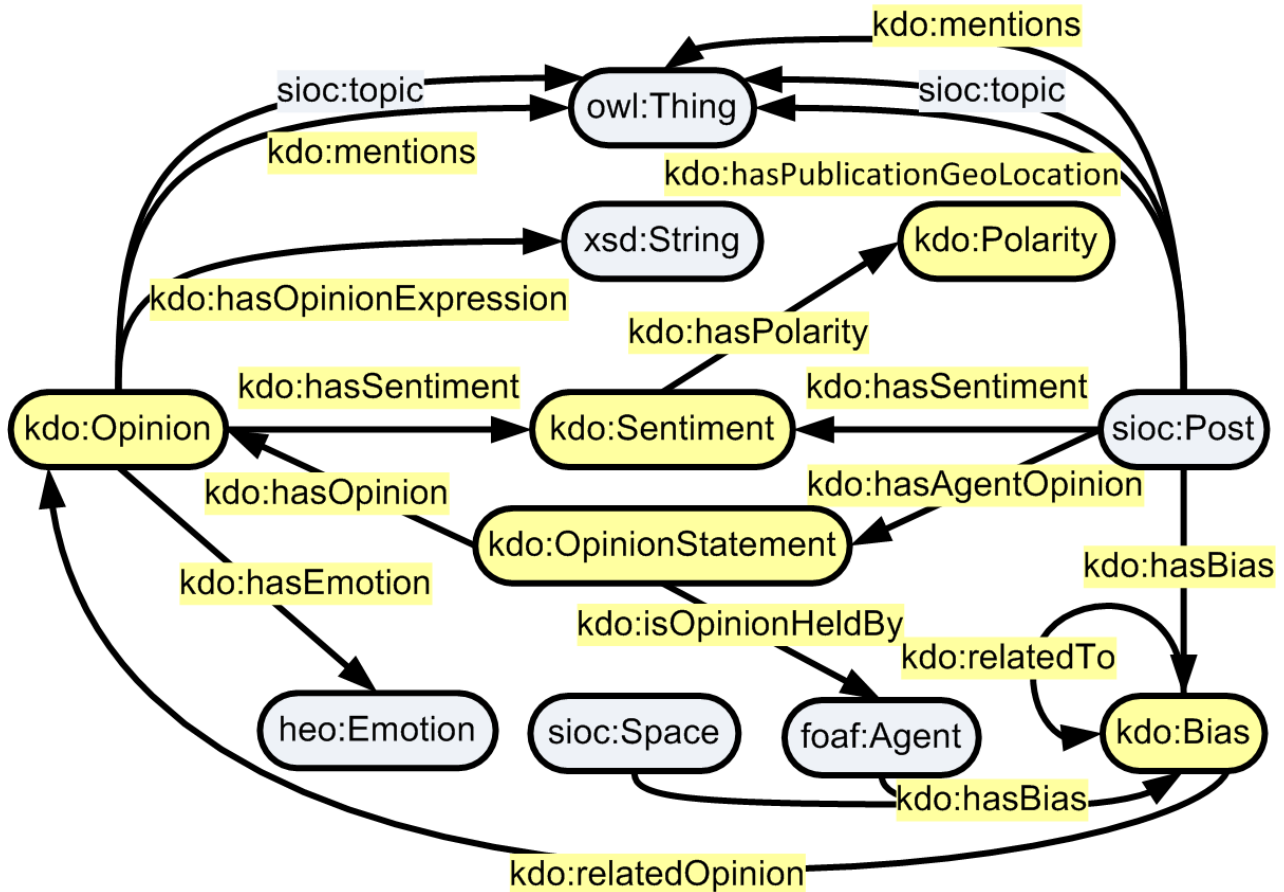


Figure 1: The Knowledge Diversity Ontology – classes and properties

4.1.1 KDO Classes

We describe each of the classes below:

4.1.1.1 Class: Opinion

URI	http://kdo.render-project.eu/kdo#Opinion
Label	Opinion
Subclass of	-

The **kdo:Opinion** class is used to model the concept of opinion. This concept is usually used together with OpinionStatement (**kdo:hasOpinion**) which defines the ternary relationship between opinion, agent, and post. An opinion can be expressed on a topic (**sioc:topic**) and can mention (**kdo:metion**) different entities. An opinion can involve (**kdo:hasEmotion**) a particular emotion. An opinion can have a particular sentiment (**kdo:hasSentiment**). Opinions can have various expressions (**kdo:hasOpinionExpression**).

An example is provided in the following:

```
<#Opinion1> a kdo:Opinion ;
  kdo:hasOpinionExpression "the takeover will happen and benefits both parties" .
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:isOpinionHeldBy <#JamesMcDonald> ;
    kdo:hasOpinion <#Opinion1> ] .
```

4.1.1.2 Class: OpinionStatement

URI	http://kdo.render-project.eu/kdo#OpinionStatement
Label	OpinionStatement
Subclass of	-

The **kdo:OpinionStatement** class is used to model the ternary relationship between agents, posts, and opinion. As this concept is typically attached to a post (e.g. as a blank node¹²), it is used as range of **kdo:hasAgentOpinion**. It is the domain of the properties **kdo:isOpinionHeldBy** and **kdo:hasOpinion**.

An example is provided in the following:

```
<http://de.wikipedia.org/wiki/Euro-Krise> a sioc:WikiArticle;
  kdp:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#BerkelWikiOpinion> ;
    kdo:isOpinionHeldBy <#berkel> ] .
```

4.1.1.3 Class: Sentiment

URI	http://kdo.render-project.eu/kdo#Sentiment
Label	Sentiment
Subclass of	-

The Sentiment class represents the sentiments expressed by the opinions and posts. A sentiment can have (**kdo:hasPolarity**) different types of polarity (positive, negative, and neutral). It is in the range of **kdo:hasSentiment**.

An example is provided in the following:

```
<#BerkelWikiOpinion> a kdo:Opinion ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:negativePolarity ] .
```

4.1.1.4 Class: Bias

URI	http://kdo.render-project.eu/kdo#Bias
Label	Bias
Subclass of	-

¹² http://en.wikipedia.org/wiki/Blank_node

The **kdo:Bias** class represents bias. A particular bias can have a selection of different **kdo:Opinion** instances related (**kdo:relatedOpinion**). Moreover, it can be related to other biases (**kdo:relatedTo**). It is in the range of **kdo:hasBias** which are used in context of agents, posts, and spaces.

An example is provided in the following:

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle .
  hasBias [ a kdo:Bias ;
    kdo:relatedTo <#anotherBias> ;
    kdo:relatedOpinion <#anOpinion> ] .
```

4.1.1.5 Class: Polarity

URI	http://kdo.render-project.eu/kdo#Polarity
Label	Polarity
Subclass of	-

The **kdo:Polarity** class is used to express the polarity of sentiments. The predefined instances of this class are: **kdo:neutralPolarity**, **kdo:positivePolarity**, **kdo:negativePolarity**.

An example is provided in the following:

```
<#BerkelWikiOpinion> a kdo:Opinion ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:negativePolarity ] .
```

4.1.1.6 Class: NewsArticle

URI	http://kdo.render-project.eu/kdo#NewsArticle
Label	News article
Subclass of	sioc:Post

The **kdo:NewsArticle** class represents news articles. The news article class is defined because there is no class to represent news articles in the SIOC ontology. The NewsArticle is a subclass of the **sioc:Post** class which is itself a subclass of the **foaf:Document**.

An example is provided in the following:

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle .
```

4.1.1.7 Class: NewsWebsite

URI	http://kdo.render-project.eu/kdo#NewsWebsite
Label	News web site
Subclass of	sioc:Container

The **kdo:NewsWebsite** class represents news publisher containers that are commonly related to a single web site. The NewsWebsite class is defined because there is no container class to represent the news publisher web sites in the SIOC ontology. The NewsWebsite class is defined as a subclass of the **sioc:Container**.

An example is provided in the following:

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle ;
    sioc:hasContainer [ a kdo:NewsWebsite ;
        sioc:has_space <http://www.bloomberg.com/ ] .
```

4.1.2 KDO Properties

The properties of KDO are described below:

4.1.2.1 Property: mentions

URI	http://kdo.render-project.eu/kdo#mentions
Label	mentions
Domain	kdo:Opinion, sioc:Post
Range	owl:Thing

Denotes a thing (**owl:Thing**) that is mentioned in a post (**sioc:Post**) or (**kdo:Opinion**).

An example is provided in the following:

```
<http://de.wikipedia.org/wiki/Euro-Krise> a sioc:WikiArticle;
    kdo:mentions <http://sws.geonames.org/390903/> ;
    kdo:mentions <http://sws.geonames.org/2510769/> ;
    kdo:mentions <http://sws.geonames.org/2264397/> .
```

4.1.2.2 Property: hasOpinionExpression

RI	http://kdo.render-project.eu/kdo#hasOpinionExpression
Label	has opinion expression
Domain	kdo:Opinion
Range	xsd:String

Denotes a textual representation of an opinion.

An example is provided in the following:

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

4.1.2.3 Property: hasEmotion

URI	http://kdo.render-project.eu/kdo#hasEmotion
-----	---

Label	has emotion
Domain	kdo:Opinion
Range	heo:Emotion

The **kdo:hasEmotion** can be used to associate an **heo:Emotion** to a **kdo:Opinion**.

An example is provided in the following:

```
<#TarifaOpinion1> a kdo:Opinion ;
  kdo: hasOpinionExpression "Yo recomiendo la tarifa para autónomos" ;
  kdo:hasEmotion heo:excitement .
```

4.1.2.4 Property: hasSentiment

URI	http://kdo.render-project.eu/kdo#hasSentiment
Label	has sentiment
Domain	kdo:Opinion, sioc:Post
Range	kdo:Sentiment

The **kdo:hasSentiment** property is used to specify which **kdo:Sentiment** is expressed by an **kdo:Opinion** or **sioc:Post**.

An example is provided in the following:

```
<#TarifaOpinion1> a kdo:Opinion ;
  kdo: hasOpinionExpression "Yo recomiendo la tarifa para autónomos" ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:positivePolarity ] .
```

4.1.2.5 Property: hasPolarity

URI	http://kdo.render-project.eu/kdo#hasPolarity
Label	has polarity
Domain	kdo:Sentiment
Range	kdo:Polarity

The **kdo:hasPolarity** is used to specify the polarity of a **kdo:Sentiment**. It has three individuals: **kdo:positivePolarity** representing the positive orientation, **kdo:negativePolarity** representing the negative orientation, **kdo:neutralPolarity** representing the neutral.

An example is provided in the following:

```
<#TarifaOpinion1> a kdo:Opinion ;
  kdo: hasOpinionExpression "Yo recomiendo la tarifa para autónomos" ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:positivePolarity ] .
```

4.1.2.6 Property: hasOpinion

URI	http://kdo.render-project.eu/kdo#hasOpinion
Label	has opinion
Domain	kdo:OpinionStatement
Range	kdo:Opinion

Relates a **kdo:OpinionStatement** to a **kdo:Opinion**.

An example is provided in the following:

```
<http://twitter.com/#!/some_user1/status/94027610476126208> a sioc:Microblog ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#TarifaOpinion1> ;
    kdo:isOpinionHeldBy <#tom294> ] .
```

4.1.2.7 Property: isOpinionHeldBy

URI	http://kdo.render-project.eu/kdo#isOpinionHeldBy
Label	is opinion held by
Domain	kdo:OpinionStatement
Range	foaf:Agent

The **kdo:isOpinionHeldBy** property is used to specify the opinion holder an opinion. An opinion holder is represented by **foaf:Agent** class.

An example is provided in the following:

```
<http://twitter.com/#!/some_user1/status/94027610476126208> a sioc:Microblog ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#TarifaOpinion1> ;
    kdo:isOpinionHeldBy <#tom294> ] .
```

4.1.2.8 Property: hasAgentOpinion

URI	http://kdo.render-project.eu/kdo#hasAgentOpinion
Label	has agent opinion
Domain	sioc:Post
Range	kdo:OpinionStatement

Relates an opinion expressed by an agent to the post in which it was stated.

An example is provided in the following:

```
<http://twitter.com/#!/some_user1/status/94027610476126208> a sioc:Microblog ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#TarifaOpinion1> ;
    kdo:isOpinionHeldBy <#tom294> ] .
```

4.1.2.9 Property: hasBias

URI	http://kdo.render-project.eu/kdo#hasBias
Label	has bias
Domain	sioc:Space, sioc:Post, foaf:Agent
Range	kdo:Bias

The **kdo:hasBias** property is used associate a **kdo:Bias** with **sioc:Space**, **sioc:Post**, and **foaf:Agent**.

An example is provided in the following:

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle .
  hasBias [ a kdo:Bias ;
    kdo:relatedTo <#anotherBias> ;
    kdo:relatedOpinion <#anOpinion> ] .
```

4.1.2.10 Property: relatedTo

URI	http://kdo.render-project.eu/kdo#relatedTo
Label	related to
Domain	kdo:Bias
Range	kdo:Bias

Relates a bias to another one.

An example is provided in the following:

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle .
  hasBias [ a kdo:Bias ;
    kdo:relatedTo <#anotherBias> ;
    kdo:relatedOpinion <#anOpinion> ] .
```

4.1.2.11 Property: relatedOpinion

URI	http://kdo.render-project.eu/kdo#relatedOpinion
Label	related opinion
Domain	kdo:Bias
Range	kdo:Opinion

Relates a bias to an opinion.

An example is provided in the following:

```
<http://www.bloomberg.com/article/c2lj3h3j4j4> a kdo:NewsArticle .
  hasBias [ a kdo:Bias ;
    kdo:relatedTo <#anotherBias> ;
    kdo:relatedOpinion <#anOpinion> ] .
```

4.1.2.12 Property:hasPublicationGeoLocation

URI	http://kdo.render-project.eu/kdo#hasPublicationGeoLocation
Label	has publication geographic location
Domain	sioc:Post
Range	owl:Thing

The **kdo:hasPublicationGeoLocation** property is used to specify the location where a particular document has been published. As more than one geo vocabulary exists we define the range as owl:Thing.

An example is provided in the following:

```
<http://twitter.com/#!/some_user1/status/94027610476126208> a sioc:Microblog ;
  kdo:hasPublicationGeoLocation <http://sws.geonames.org/3336903/> .
```

4.2 Reused Ontologies and Standards

In this section, we describe the reused ontologies and provide a guideline on the subset of the concepts and properties that are recommended for use with the KDO.

4.2.1 The Friend of a Friend (FOAF) Vocabulary

The aim of the FOAF project is to link people and information using the Web. FOAF provides a vocabulary to describe people and their relationships. In spirit of AAA principle: Anyone can say Anything about Any topic, usually the topics are about people in FOAF. Other things related to what one might say about people, such as Organizations, that people belong to; Documents, that describe people and other topics described by people; Images, that illustrate people, are also part of the core FOAF vocabulary. FOAF vocabulary is mainly organized around two main classes: **foaf:Person** and **foaf:Document**.

4.2.1.1 Agents

Some of the things that can be said about people also hold for other things such as organizations, and groups. Therefore, **foaf:Agent** is defined as a general class of **foaf:Person**, **foaf:Organization**, and **foaf:Group**. In the RENDER project, an opinion is held by an opinion holder, which is a person or an organization or a group of people, can be described as an agent. Furthermore, FOAF provides properties such as **foaf:name**, describing the name of something; **foaf:age**, describing the age of an agent; **foaf:birthday**, describing the date of birth of an agent; **foaf:gender**, describing the gender of an agent; **foaf:based_near**, describing the geographical location of something that is based near; **foaf:topic_interest**, describing a topic of interest to a person; and **foaf:knows**, relating a person to another person that he or she knows; etc. which allow to add demographic details to agents such as people and establish relationships between people. In RENDER, these features of adding demographic information and establish

relationships between people are useful for cases such as describing usage diversity dimension of demography, topic, geography, and social context.

4.2.1.2 Documents

FOAF represents documents as things which are broadly conceived as documents. Currently, FOAF does not distinguish between physical and electronic documents, or between copies of a work and the abstraction those copies embody. In RENDER, the web contents such as news articles, blog posts, wiki article, or microblog posts are represented as electronic documents, represented by the **foaf:Document** class. Furthermore, different types of documents are represented using the type module of the Semantically-Interlinked Online Communities (SIOC) Ontology¹³ which is discussed in section 4.2.3.

Other FOAF concepts and properties that are relevant for specifying diversity information include for example **foaf:Topic** use to specify the topic of a document. It is defined as a property having as domain any **foaf:Document** and as range **owl:Thing**. The topics of other entities, other than documents, can be expressed using the property **sioc:topic** from SIOC ontology (see section 4.2.3).

4.2.2 The Dublin Core Metadata Initiative

The Dublin Core Metadata Initiative provides terms for expressing metadata using RDF. Table 2 contains the Dublin Core terms that can be used to describe metadata in the documents:

Table 2: DCMI metadata terms

Term	Type	Domain	Range	Description
dcterms:creator	Property	rdfs:Resource	dcterms:Agent	An entity primarily responsible for making the resource. Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
dcterms:contributor	Property	rdfs:Resource	dcterms:Agent	An entity responsible for making contributions to the resource. Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
dcterms:publisher	Property	rdfs:Resource	dcterms:Agent	An entity responsible for making the resource available. Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
dcterms:description	Property	rdfs:Resource	rdfs:Resource	An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
dcterms:identifier	Property	rdfs:Resource	rdfs:Literal	An unambiguous reference to the resource within a given context.

¹³ <http://rdfs.org/sioc/spec/>

				Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.
dcterms:source	Property	rdfs:Resource	rdfs:Resource	<p>A related resource from which the described resource is derived. The described resource may be derived from the related resource in whole or in part. Use only when the described resource is the result of digitization of non-digital originals.</p> <p>For example a digital resource can have as source a non-digital source identified as follows</p> <pre>< dcterms:source>Ecology Letters (1461023X) vol.4 (2001) </dcterms:source> < dcterms:source>ISSN: 0928-0987 </ dcterms:source></pre>
dcterms:title	Property	rdfs:Resource	rdfs:Resource	A name given to the resource.
dcterms:created	Property	rdfs:Resource	rdfs:Literal	Date of creation of the resource.
dcterms:language	Property	rdfs:Resource	dcterms:LinguisticSystem	A language of the resource. Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].

4.2.3 The Semantically-Interlinked Online Communities (SIOC) Ontology

The SICO ontology provides a vocabulary to describe and interlink online community sites such as blogs, message boards, wikis, etc. Figure 2 which is taken from [1] shows the main terms in the SIOC core ontology.

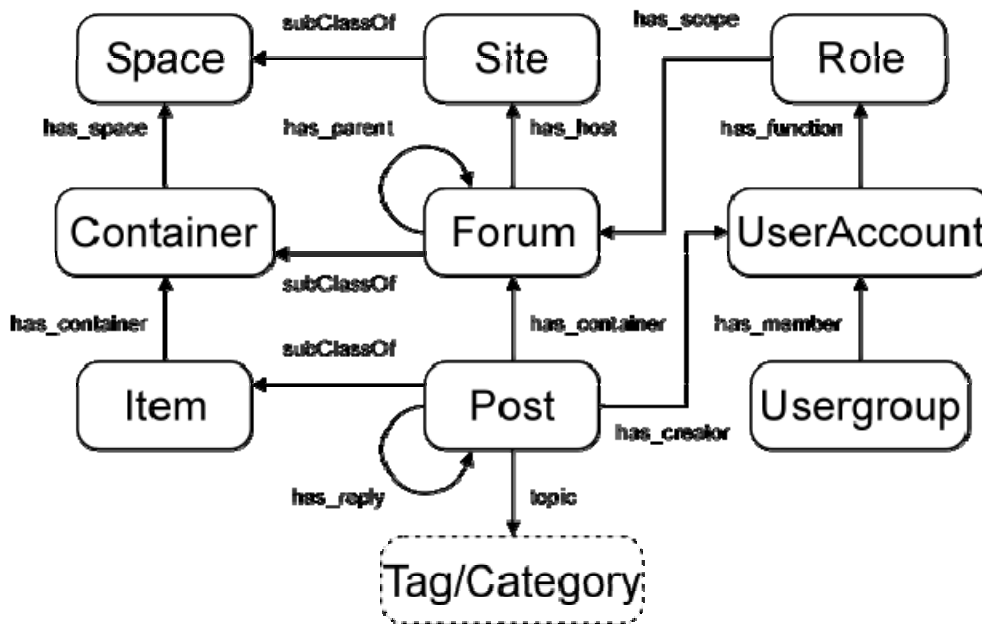


Figure 2: Main classes and properties in SIOC

In short, Users create the content Items (e.g., Posts) containing in the Containers (e.g., Posts in a Forum) on the data Spaces (e.g., Sites). The SIOC Type module extends SIOC core ontology to describe various types of content items on the Web 2.0 platform. For example:

- The **Container** class representing the containers is extended (subclassed) to AddressBook, AnnotationSet, AudioChannel, BookmarkFolder, Briefcase, EventCalendar, ImageGallery, ProjectDirectory, ResumeBank, ReviewArea, SubscriptionList, SurveyCollection, VideoChannel and Wiki to represent different type of containers.
- The **Forum** class representing a discussion forum/area is extended to ArgumentativeDiscussion, ChatChannel, MailingList, MessageBoard and Weblog to represent different type of discussion forums.
- The **Post** class representing an article or message that can be posted in a container is extended to BlogPost, BoardPost, Comment, InstantMessage, MailMessage, WikiArticle to represent different kind of posts.
- The **Item** representing something which can be in a container is extended to Poll to represent online polls or surveys.

In the RENDER project, the various types of Web documents can be represented using the SIOC type module. However, there is no term to represent news items in SIOC ontology. This problem is solved by adding the **kdo:NewsArticle** class, representing the news article posts, as a subclass of the **sioc:Post** class and adding the **kdo:NewsWebsite** class, representing the new publisher websites, as a subclass of the **sioc:Container** class.

In the RENDER project, absolute truth is not considered. The statements are considered as the opinions of the opinion holders. Therefore, it is important to represent and exploit the origin of the statements, namely the provenance information. A combination of the FOAF vocabulary, Dublin Core Terms and SIOC can be used to describe provenance information such as creator or contributor of the Web documents.

The topics of any entity can be expressed using the property **sioc:topic** which has as domain **owl:Thing** and as range a URI.

4.2.4 PRO, the Publishing Role Ontology

The Publishing Role Ontology contains 31 roles to describe the publishing roles. The Figure 3 shows the individuals of the **pro:Role** class representing the 31 publishing roles. Moreover, pro defines properties that enable the linkage of **pro:Role** with **foaf:Agent** and **foaf:Document** using a forth class that is called **pro:RoleInTime**. The relevant properties are **pro:isRelatedToRoleInTime**, **pro:withRole**, **pro:isRoleHeldBy**, and their respective inverses (**owl:inverseOf**).

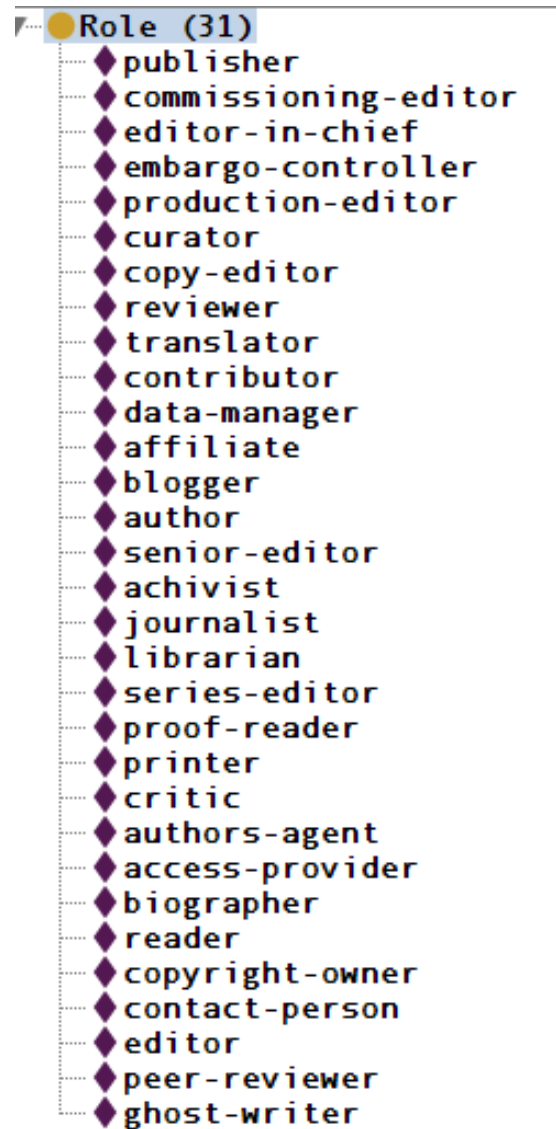


Figure 3: The individuals of the pro:Role class representing the publishing roles

4.2.5 Human Emotions Ontology (HEO)

The Human Emotions Ontology (HEO)¹⁴ has been developed by Marco Grassi of Università Politecnica delle Marche [3]. It includes various categories for emotion, such as the archetypal category and the Douglas Cowie category. As visualized in Figure 4, the class Emotion states a central element. In KDO, we reuse this class and different instances to describe emotions that are encompassed with opinions.

¹⁴ <http://semedia.dibet.univpm.it/heo/heo.owl>

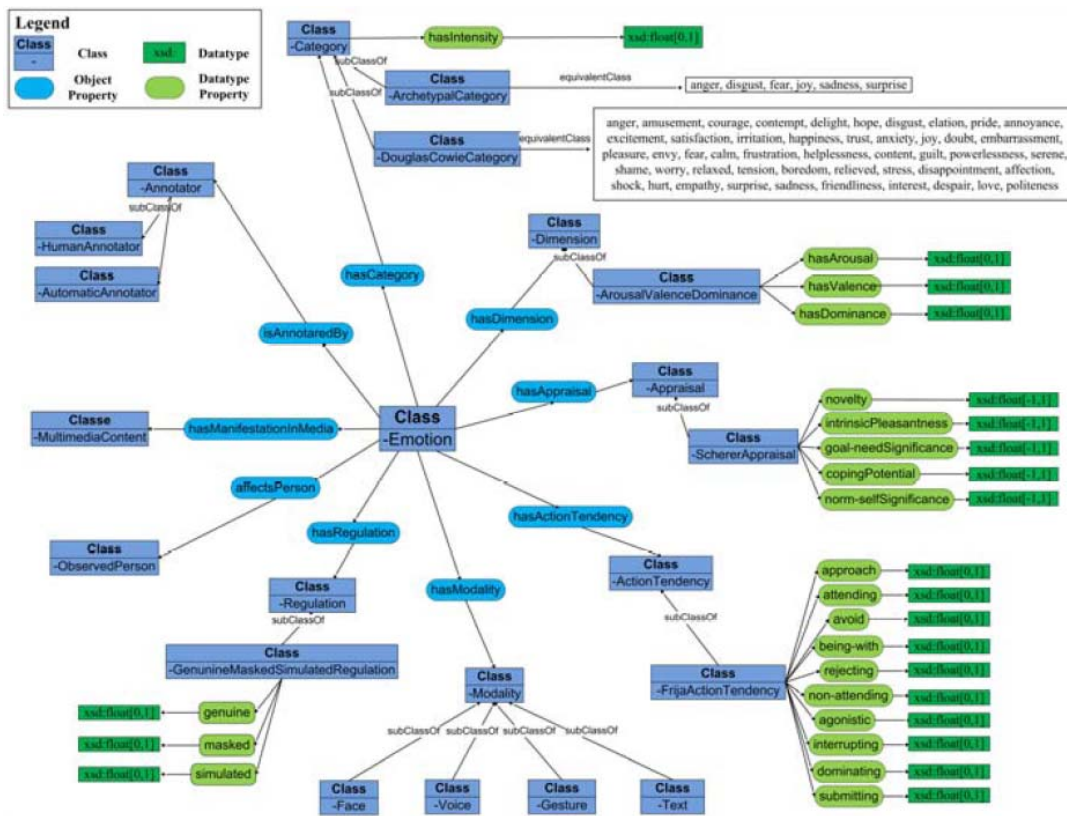


Figure 4: The classes defined by HEO (source [3])

4.2.6 GeoNames

GeoNames is a worldwide geographical database available for free under Creative Common attribution. GeoNames contains 6.2 million geonames with unique URIs. GeoNames URIs can be used to refer to locations in RENDER. GeoNames also provides a RDF representation of each node (e.g. for the country Bulgaria it is retrievable from <http://sws.geonames.org/732800>). Other geographical ontologies can be used for the same purpose.

4.2.7 dmoz (open directory project)

The open directory project dmoz describes real world things and categorizes them in a taxonomy. The defined URIs can be used to refer to specific things, such as the Olympic Winter Games 2018 (http://www.dmoz.org/Sports/Events/Olympics/Winter_Games/2018/). In the RENDER project, this open directory can be used to refer to particular entities that are extracted through named entity recognition (NER)¹⁵.

¹⁵ http://en.wikipedia.org/wiki/Named_entity_recognition

5 Using the Knowledge Diversity Ontology

The KDO can be easily used to model various scenarios that relate to the use cases defined by Google News, Telefonica and Wikipedia. In order to exemplify this, in subsection 5.1, we show how to model the data of the motivating scenarios (see section 2). Moreover, based on the modelled data, we are able to formulate SPARQL¹⁶ queries in subsection 5.2 which enable answering the competency questions of section 3.1.

5.1 Modelling Motivating Scenarios with KDO

In the following, we exemplify how to model the data defined by the storylines of the motivating scenarios (see section 2). We use the Turtle¹⁷ RDF serialization syntax and omit the prefixes of the different namespaces.

First, we model **Scenario 1** (cf. section 2.1.1) that relates to the first of the “News Articles” use cases:

```
<#FerdinandMaier> a foaf:Person ;
  foaf:firstName "Ferdinand" ;
  foaf:lastName "Maier" .

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

<#GGC> a foaf:Organization ;
  foaf:name "GGC" .

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

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

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

<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 <#GGC> ] ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:positivePolarity ] ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:isOpinionHeldBy <#JamesMcDonald> ;
    kdo:hasOpinion <#Opinion1> ] ;
```

¹⁶ W3C Recommendation - SPARQL Query Language for RDF: <http://www.w3.org/TR/rdf-sparql-query/>

¹⁷ Turtle - Terse RDF Triple Language: <http://www.w3.org/TeamSubmission/turtle/>

```

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> .

```

Listing 1: Scenario 1 modelled with the KDO

For the use case that relates to the collaborative knowledge bases, we chose **Scenario 4** (cf. section 2.2.2):

```

<#SpanishWikiOpinion> a kdo:Opinion ;
  kdo:hasOpinionExpression "The financial crisis mainly affects Greece and Ireland" .

<http://es.wikipedia.org/wiki/Crisis_del_euro_en_2010> a sioc:WikiArticle;
  dcterms:hasVersion <http://de.wikipedia.org/wiki/Euro-Krise> ;
  dcterms:hasVersion <http://en.wikipedia.org/wiki/European_sovereign_debt_crisis> ;
  dcterms:created "2011-07-10" ;
  dcterms:language "es-ES" ;
  kdo:mentions <http://sws.geonames.org/2963597/> ;
  kdo:mentions <http://sws.geonames.org/390903/> ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#SpanishWikiOpinion > ] .

<#berkel> a foaf:Agent ;
  foaf:nick "berkel" .

<#BerkelWikiOpinion> a kdo:Opinion ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:negativePolarity ] .

<http://de.wikipedia.org/wiki/Euro-Krise> a sioc:WikiArticle;
  dcterms:hasVersion <http://en.wikipedia.org/wiki/European_sovereign_debt_crisis> ;
  dcterms:hasVersion <http://es.wikipedia.org/wiki/Crisis_del_euro_en_2010> ;
  dcterms:created "2011-07-10" ;
  dcterms:language "de-DE" ;
  pro:isRelatedToRoleInTime [ a pro:RoleInTime ;
    pro:withRole pro:editor ;
    pro:isRoleHeldBy <#berkel> ;
  ];
  kdp:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#BerkelWikiOpinion> ;
    kdo:isOpinionHeldBy <#berkel> ] ;
  kdo:mentions <http://sws.geonames.org/390903/> ;
  kdo:mentions <http://sws.geonames.org/2510769/> ;
  kdo:mentions <http://sws.geonames.org/2264397/> .

```



```

<http://de.wikipedia.org/w/index.php?title=Euro-Krise&oldid=73933632> a sioc:WikiArticle;
  dcterms:hasVersion <http://de.wikipedia.org/wiki/Euro-Krise> ;
  dcterms:created "2010-05-10" ;
  dcterms:language "de-DE" ;
  kdo:mentions <http://sws.geonames.org/390903/> .

<http://es.wikipedia.org/w/index.php?title=Crisis_del_euro_en_2010&oldid=37299051> a sioc:WikiArticle;
  dcterms:hasVersion <http://es.wikipedia.org/wiki/Crisis_del_euro_en_2010> ;
  dcterms:created "2010-05-24" ;
  dcterms:language "es-ES" ;
  kdo:mentions <http://sws.geonames.org/390903/> .

```

Listing 2: Scenario 4 modelled with the KDO

For the use case that relates to the customer feedback, we chose **Scenario 5** (cf. section 2.3.1):

```

<#max_1977_1> a foaf:Agent ;
  foaf:nick "max_1977_1" .

<#tom294> a foaf:Agent ;
  foaf:nick "tom294" .

<#onlineUsers> a foaf:Group ;
  rdfs:label "Online Users" ;
  foaf:member <#max_1977_1> ;
  foaf:member <#tom294> .

<#TarifaOpinion1> a kdo:Opinion ;
  kdo:hasOpinionExpression "Yo recomiendo la tarifa para autónomos" ;
  kdo:hasEmotion heo:excitement
  kdo:hasSentiment [ a kdo:Sentiment;
    kdo:hasPolarity kdo:positivePolarity ] .

<#TarifaOpinion2> a kdo:Opinion ;
  kdo:hasOpinionExpression "No contratéis la tarifa para autónomos, es un robo!" ;
  kdo:hasEmotion heo:anger ;
  kdo:hasSentiment [ a kdo:Sentiment ;
    kdo:hasPolarity kdo:negativePolarity ] .

<http://twitter.com/#!/some_user1/status/94027610476126208> a sioc:Microblog ;
  dcterms:created "2011-04-24" ;
  dcterms:language "es-ES" ;
  kdo:hasPublicationGeoLocation <http://sws.geonames.org/3336903/> ;
  sioc:topic <http://www.dmoz.org/Regional/Europe/Spain/Communications/Wireless/Tarifa_Autónomos> ;
  kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
    kdo:hasOpinion <#TarifaOpinion1> ;
    kdo:isOpinionHeldBy <#tom294> ] ;
  pro:isRelatedToRoleInTime [ a pro:RoleInTime ;

```

```

pro:withRole pro:author ;
pro:isRoleHeldBy <#tom294>].

<http://twitter.com/#!/some_user2/status/6524389758437534> a sioc:Microblog ;
dcterms:created "2011-04-27" ;
dcterms:language "es-ES" ;
kdo:hasPublicationGeoLocation <http://sws.geonames.org/3336903/> ;
kdo:hasTopic <http://www.dmoz.org/Regional/Europe/Spain/Communications/Wireless/Tarifa_Autónomos> ;
kdo:hasAgentOpinion [ a kdo:OpinionStatement ;
  kdo:hasOpinion <#TarifaOpinion2> ;
  kdo:isOpinionHeldBy <#max_1977_1> ] ;
pro:isRelatedToRoleInTime [ a pro:RoleInTime ;
  pro:withRole pro:author ;
  pro:isRoleHeldBy <#max_1977_1>].
    
```

Listing 3: Scenario 5 modelled with the KDO

5.2 Answering Competency Questions with SPARQL

In the following table, we exemplify for each of the competency questions defined in section 3.1 how the questioned item can be retrieved using a SPARQL query that is based on the KDO.

Table 3: SPARQL queries answering the competency questions

Question Number	SPARQL
1	<pre> SELECT ?uri WHERE { <document> rdf:type sioc:Post } </pre>
2	<pre> SELECT ?source WHERE { <document> sioc:hasContainer ?container. ?container sioc:has_space ?source } </pre>
3	<pre> SELECT ?type WHERE { <document> rdf:type ?type} </pre>
4	<pre> SELECT ?mentioned WHERE { <document> kdo:mentions ?mentioned } / SELECT ?topic WHERE { <document> sioc:topic ?topic } </pre>

5	<pre>SELECT ?language WHERE { <document> dcterms:language ?language }</pre>
6	<pre>SELECT ?date WHERE { <document> dcterms:created ?date }</pre>
7	<pre>SELECT ?contributor, ?role WHERE { <document> pro:isRelatedToRoleInTime ?roleInTime. ?roleInTime pro:withRole ?role. ?roleInTime pro:isRoleHeldBy ?contributor }</pre>
8	<pre>SELECT ?location WHERE { <document> kdo:hasPublicationGeoLocation ?location. }</pre>
9	<pre>SELECT ?contributor, ?group WHERE { <document> pro:isRelatedToRoleInTime ?roleInTime. ?roleInTime pro:isRoleHeldBy ?contributor. ?group foaf:member ?contributor }</pre>
10	<pre>SELECT ?bias WHERE { <document> kdo:hasBias ?bias }</pre>
11	<pre>SELECT ?version WHERE { <document> dcterms:hasVersion ?version }</pre>
12	<pre>SELECT ?opinion WHERE { <document> kdo:hasAgentOpinion ?opinionStatement. ?opinionStatement kdo:hasOpinion ?opinion }</pre>
13	<pre>SELECT ?opinionHolder WHERE { ?opinionStatement kdo:hasOpinion <opinion>. ?opinionStatement kdo:isOpinionHeldBy ?opinionHolder }</pre>
14	<pre>SELECT ?sentiment WHERE { <opinion> kdo:hasSentiment ?sentiment }</pre>
15	<pre>SELECT ?sentiment WHERE { <document> kdo:hasSentiment ?sentiment }</pre>

	}
16	SELECT ?polarity WHERE { <sentiment> kdo:hasPolarity ?polarity }
17	SELECT ?emotion WHERE { <sentiment> kdo:hasEmotion ?emotion }
18	SELECT ?expression WHERE { <opinion> kdo:hasOpinionExpression ?expression }

6 Conclusion

This deliverable provides an initial version of the Knowledge Diversity Ontology (KDO) and how it has been derived. We formulated motivating scenarios that are based on the use cases that our partners Wikimedia, Google, and Telefonica provided. Based on this, we derived competency questions that are meant to be answered with the help of KDO. The KDO itself models basic concepts and properties that are needed to represent diversity of knowledge. Moreover, we demonstrated the interplay of the introduced components in combination with other vocabularies such as FOAF, SIOC, DCMI, PRO, and HEO: an excerpt of the motivating scenarios has been modelled. All the competency questions were finally formulated as SPARQL queries.

KDO reuses many existing ontologies. This deliverable helps to identify the ontologies that are relevant in the context of modelling diversity, describing the fragments of these ontologies that are actually used, and illustrates using snippet examples from the modelled scenarios, the interplay of KDO and reused ontologies to model diversity.

The knowledge diversity ontology defined in this deliverable will serve as the underlying conceptual model for the Render services and Render tools extensions that are developed in the other work packages. Representing diversity using KDO and using the services and tools extensions to link diversity to information items produced or handled by these services and tools will bring benefits when doing search, filtering, ranking or recommendation. For example, a connection the FactForge developed by partners from WP1, will enable users to formulate more complex queries using more datasets. More complex examples on how to use the KDO ontology will be provided in the upcoming deliverables in WP3.

References

- [1] Breslin, J. G., Passant, A., & Decker, S. (2009). Interlinking online communities. In *The Social Semantic Web* (pp. 197-250). Springer Berlin Heidelberg.
- [2] R. Hasan, K. Siorpaes, R. Krummenacher, and F. Flöck, "Towards a Knowledge Diversity Model," in *First International Workshop on Knowledge Diversity on the Web (DiversiWeb 2011), 20th World Wide Web Conference (WWW 2011)*, 2011.
- [3] Grassi, M (2009). *Developing HEO Human Emotions Ontology*. Lecture Notes in Computer Science, Vol. 5707/2009. Springer-Verlag Berlin Heidelberg, 2009

Annex A KDO formulated in OWL

```

<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
  <!ENTITY ns "http://rdfs.org/sioc/ns#" >
  <!ENTITY foaf "http://xmlns.com/foaf/0.1/" >
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY kdo "http://kdo.render-project.eu/kdo#" >
  <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY HumanEmotions "http://www.semanticweb.org/ontologies/2009/2/HumanEmotions.owl#" >
]>

<rdf:RDF xmlns="http://kdo.render-project.eu/kdo#"
  xml:base="http://kdo.render-project.eu/kdo"
  xmlns:HumanEmotions="http://www.semanticweb.org/ontologies/2009/2/HumanEmotions.owl#"
  xmlns:kdo="http://kdo.render-project.eu/kdo#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:ns="http://rdfs.org/sioc/ns#"
  xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <owl:Ontology rdf:about="">
    <rdfs:comment
      >The Knowledge Diversity Ontology.</rdfs:comment>
    <owl:imports rdf:resource="http://rdfs.org/sioc/ns"/>
    <owl:imports rdf:resource="http://www.semanticweb.org/ontologies/2009/2/HumanEmotions.owl"/>
  </owl:Ontology>

  <!--
  //////////////////////////////////////
  //
  // Object Properties
  //
  //////////////////////////////////////
  -->

```

```
<!-- http://kdo.render-project.eu/kdo#hasAgentOpinion -->
```

```
<owl:ObjectProperty rdf:about="#hasAgentOpinion">  
  <rdfs:range rdf:resource="#OpinionStatement"/>  
  <rdfs:domain rdf:resource="&ns;Post"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#hasBias -->
```

```
<owl:ObjectProperty rdf:about="#hasBias">  
  <rdfs:range rdf:resource="#Bias"/>  
  <rdfs:domain rdf:resource="&ns;Post"/>  
  <rdfs:domain rdf:resource="&ns;Space"/>  
  <rdfs:domain rdf:resource="&foaf;Agent"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#hasEmotion -->
```

```
<owl:ObjectProperty rdf:about="#hasEmotion">  
  <rdfs:domain rdf:resource="#Opinion"/>  
  <rdfs:range rdf:resource="&HumanEmotions;Emotion"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#hasOpinion -->
```

```
<owl:ObjectProperty rdf:about="#hasOpinion">  
  <rdfs:range rdf:resource="#Opinion"/>  
  <rdfs:domain rdf:resource="#OpinionStatement"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#hasPolarity -->
```

```
<owl:ObjectProperty rdf:about="#hasPolarity">  
  <rdfs:range rdf:resource="#Polarity"/>  
  <rdfs:domain rdf:resource="#Sentiment"/>  
</owl:ObjectProperty>
```



```
<!-- http://kdo.render-project.eu/kdo#hasPublicationGeoLocation -->
```

```
<owl:ObjectProperty rdf:about="#hasPublicationGeoLocation">  
  <rdfs:domain rdf:resource="#Post"/>  
  <rdfs:range rdf:resource="#owl;Thing"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#hasSentiment -->
```

```
<owl:ObjectProperty rdf:about="#hasSentiment">  
  <rdfs:domain rdf:resource="#Opinion"/>  
  <rdfs:range rdf:resource="#Sentiment"/>  
  <rdfs:domain rdf:resource="#Post"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#isOpinionHeldBy -->
```

```
<owl:ObjectProperty rdf:about="#isOpinionHeldBy">  
  <rdfs:domain rdf:resource="#OpinionStatement"/>  
  <rdfs:range rdf:resource="#foaf;Agent"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#mentions -->
```

```
<owl:ObjectProperty rdf:about="#mentions">  
  <rdfs:domain rdf:resource="#Opinion"/>  
  <rdfs:domain rdf:resource="#Post"/>  
  <rdfs:range rdf:resource="#owl;Thing"/>  
</owl:ObjectProperty>
```

```
<!-- http://kdo.render-project.eu/kdo#relatedOpinion -->
```

```
<owl:ObjectProperty rdf:about="#relatedOpinion">  
  <rdfs:domain rdf:resource="#Bias"/>  
  <rdfs:range rdf:resource="#Opinion"/>  
</owl:ObjectProperty>
```

```

<!-- http://kdo.render-project.eu/kdo#relatedTo -->

<owl:ObjectProperty rdf:about="#relatedTo">
  <rdfs:domain rdf:resource="#Bias"/>
  <rdfs:range rdf:resource="#Bias"/>
</owl:ObjectProperty>

<!--
////////////////////////////////////
//
// Data properties
//
////////////////////////////////////
-->

<!-- http://kdo.render-project.eu/kdo#hasOpinionExpression -->

<owl:DatatypeProperty rdf:about="#hasOpinionExpression">
  <rdfs:domain rdf:resource="#Opinion"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<!--
////////////////////////////////////
//
// Classes
//
////////////////////////////////////
-->

<!-- http://kdo.render-project.eu/kdo#Bias -->

<owl:Class rdf:about="#Bias">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</owl:Class>

```

```
<!-- http://kdo.render-project.eu/kdo#Opinion -->

<owl:Class rdf:about="#Opinion"/>

<!-- http://kdo.render-project.eu/kdo#OpinionStatement -->

<owl:Class rdf:about="#OpinionStatement"/>

<!-- http://kdo.render-project.eu/kdo#Polarity -->

<owl:Class rdf:about="#Polarity">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</owl:Class>

<!-- http://kdo.render-project.eu/kdo#Sentiment -->

<owl:Class rdf:about="#Sentiment"/>

<!-- http://rdfs.org/sioc/ns#Post -->

<owl:Class rdf:about="&ns;Post"/>

<!-- http://rdfs.org/sioc/ns#Space -->

<owl:Class rdf:about="&ns;Space"/>

<!-- http://www.semanticweb.org/ontologies/2009/2/HumanEmotions.owl#Emotion -->

<owl:Class rdf:about="&HumanEmotions;Emotion"/>

<!-- http://www.w3.org/2002/07/owl#Thing -->

<owl:Class rdf:about="&owl;Thing"/>
```

```
<!-- http://xmlns.com/foaf/0.1/Agent -->

<owl:Class rdf:about="&foaf;Agent"/>

<!--
////////////////////////////////////
//
// Individuals
//
////////////////////////////////////
-->

<!-- http://kdo.render-project.eu/kdo#negativePolarity -->

<owl:Thing rdf:about="#negativePolarity">
  <rdf:type rdf:resource="#Polarity"/>
</owl:Thing>

<!-- http://kdo.render-project.eu/kdo#neutralPolarity -->

<Polarity rdf:about="#neutralPolarity">
  <rdf:type rdf:resource="&owl;Thing"/>
</Polarity>

<!-- http://kdo.render-project.eu/kdo#positivePolarity -->

<Polarity rdf:about="#positivePolarity">
  <rdf:type rdf:resource="&owl;Thing"/>
</Polarity>
</rdf:RDF>

<!-- Generated by the OWL API (version 2.2.1.1138) http://owlapi.sourceforge.net -->
```