O-Dang! The Ontology of Dangerous Speech Messages

Marco A. Stranisci[*], **Simona Frenda**[*◇], **Mirko Lai**[*], **Oscar Araque**[⋆], **Alessandra T. Cignarella**[*],
Valerio Basile[*], **Viviana Patti**[*], **Cristina Bosco**[*]

[*] Department of Computer Science - University of Turin, Turin, Italy
[◇] PRHLT Research Center - Universitat Politècnica de València, Valencia, Spain
[⋆] Intelligent Systems Group - ETSI Telecomunicación - Universidad Politécnica de Madrid, Madrid, Spain
{marcoantonio.stranisci simona.frenda, mirko.lai, alessandrateresa.cignarella,
valerio.basile, viviana.patti, cristina.bosco}@unito.it
o.araque@upm.es

Abstract

Inside the NLP community there is a considerable amount of language resources created, annotated and released every day with the aim of studying specific linguistic phenomena. Despite a variety of attempts in order to organize such resources has been carried on, a lack of systematic methods and of possible interoperability between resources are still present. Furthermore, when storing linguistic information, still nowadays, the most common practice is the concept of "gold standard", which is in contrast with recent trends in NLP that aim at stressing the importance of different *subjectivities* and points of view when training machine learning and deep learning methods. In this paper we present O-Dang!: The Ontology of Dangerous Speech Messages, a systematic and interoperable Knowledge Graph (KG) for the collection of linguistic annotated data. O-Dang! is designed to gather and organize Italian datasets into a structured KG, according to the principles shared within the Linguistic Linked Open Data community. The ontology has also been designed to account a *perspectivist* approach, since it provides a model for encoding both gold standard and single-annotator labels in the KG. The paper is structured as follows. In Section 1. the motivations of our work are outlined. Section 2. describes the O-Dang! Ontology, that provides a common semantic model for the integration of datasets in the KG. The Ontology Population stage with information about corpora, users, and annotations is presented in Section 3.. Finally, in Section 4. an analysis of offensiveness across corpora is provided as a first case study for the resource.

Keywords: Knowledge Graph, LLOD, Hate Speech, Misogyny, Irony, Sarcasm, NLP, Annotations, Subjectivity, Perspectivism.

1. Introduction and Motivation

In this day and age, in almost every research field – as well as in Computational Linguistics – it is considered an enormous wealth to have the presence of manually annotated data sets in order to implement Machine Learning and Deep Learning pipelines. In the last 15 years there has been a very extensive effort within many research groups that deal with Natural Language Processing (NLP) for the creation, development and maintenance of corpora of linguistic data annotated with regard to various phenomena.

Nowadays, there are thousands of data sets that model similar phenomena in many different languages, and it often happens that each research group models a phenomenon on the basis of their own annotation scheme, usually not shared with other researchers, who are involved in studying similar phenomena on different languages. Another frequent case in the modeling of linguistic phenomena is to develop new annotations adding further layers of information on top of pre-existing ones to train, for instance, models based on multitask learning.

The research idea we would like to present in this paper stems from the need to provide a more structured organization to the myriad of linguistic resources and datasets developed in the NLP field, and to guarantee interoperability and dialogue between similar resources.

Among the many projects that already devoted their efforts in creating a bridge between sentiment and emotion analysis and linguistic data, we mostly referred to EUROSENTIMENT (Sánchez Rada et al., 2014) developing a common language resource representation model based on established Linked Data formats such as Onyx (Sánchez-Rada and Iglesias, 2016) and Marl (Westerki and Sánchez-Rada, 2013).

In this paper, we describe the creation of a Linguistic Linked Open Data (LLOD) resource, focused on collecting dangerous messages that indirectly contribute to the spread of discriminatory contents, thus called **The Ontology of Dangerous Speech** (O-Dang!).

Dangerous Speech has been defined by Benesch (2012) as a speech that "has a reasonable *chance* of catalyzing or amplifying violence by one group against another, given the circumstances in which it was made or disseminated". This *chance* materializes when the circumstances in which the speech takes place consist of: 1) a powerful speaker or source with a high degree of influence, 2) an audience that believe to be subject to a threat, 3) a social and historical context propitious for the violence, 4) the means of dissemination (such as social media), 5) the content of the speech that aims at the process of dehumanization, guilt attribution, threat construction, destruction of alternatives, creation of a new semantics of the violence conceived as admirable, linked to praiseworthy qualities and based on specific biased references that justify it (Leader Maynard and Benesch, 2016). Dangerous speech, therefore, is a type of speech that aims at contributing to create a climate of violence and intolerance against protected groups of people, such as women, immigrants, religious minorities, and others.

As some scholars highlighted, there are various rhetorical and pragmatic devices that play a part in the expression of dangerous utterances. For instance, Grimminger and Klinger (2021) and Frenda et al. (2019) reflected on the use
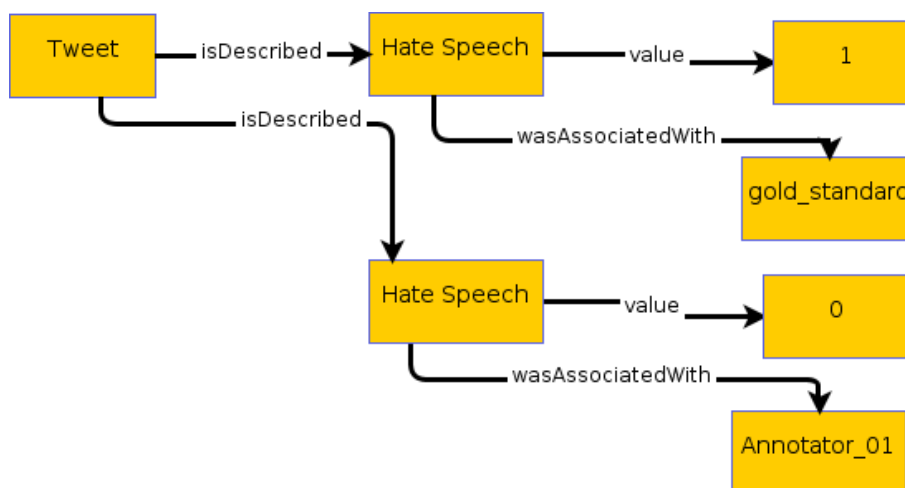
Figure 1: A snapshot of the O-Dang Semantic Model in which gold standard and un-aggregated annotations are encoded.

of offensive and toxic communication in tweets expressing a stance towards specific political candidates (such as Biden and Trump) or sensible social issues involving a particular target such as women (like feminist movements or abortion). Others focused more on the use of the ironic language to lessen the negative tones of the hateful messages, making their automatic recognition challenging (Nobata et al., 2016; Frenda et al., 2022). The employment of these kinds of devices actually lets speakers or users to be less explicit in their claims, limiting, thus, their exposure.

From this perspective, we designed an ontology for storing existing Italian corpora in a Knowledge Graph (KG) that is interoperable and that takes into account general characteristics of the various NLP datasets annotated for various dimensions of hate such as Hate Speech (HS), misogyny, stereotypes, and offensive and aggressive language. The semantic model is general enough to populate the KG with other corpora focused on orthogonal phenomena to hate, such as stance or ironic language, realizing a tool that is open to collaborative effort of the scientific community.

In this sense we are inspired by the work of Bender and Friedman (2018) in which the authors propose data statements as a design solution and professional practice for natural language processing technologists to be followed when creating a linguistic resource and making it available to other researchers.

Furthermore, our work follows the directives of the *Perspectivist Data Manifesto*[1] (Basile, 2020), and that is, we do not limit ourselves to consider the data of written texts and the gold standard labels, but – where possible – we try to store in the KG the labels of the different annotations in un-aggregated form for emphasizing the importance of the different perspectives and points of view of individual subjectivities of human annotators. Finally, our work is also inspired by Lewandowska-Tomaszczyk et al. (2021) which is focused on aligning several phenomena correlated to discrimination in a unique semantic model.

The contributions to be found in this article are the following:

- **The O-Dang! Ontology**[2], a semantic model aimed at describing and linking a variety of datasets containing Dangerous Speech and orthogonal phenomena;

- A KG containing 11 existing Italian NLP data sets on Dangerous Speech and parallel phenomena. The KG serves as a first case study for providing interoperability between corpora annotatet fo Dangerous Speech;

- **an Entity Linking pipeline** for recovering the specific targets of Dangerous Speech and abusive language;

- **un-aggregated annotations** of the datasets developed by our research group in the past years;

The resulting KG will be available through endpoint SPARQL, allowing several applications, among which:

- exporting personalized portion of the KG for the study of specific phenomena across corpora and the configuration of different training sets.

- querying all Dangerous Speech referred to specific persons and groups

- filtering gold standard and un-aggregated annotation

- querying the communication interaction among users and messages

2. The O-Dang! Ontology

The O-Dang! Ontology provides a general encoding for the harmonization of different datasets in a unique resource. The model relies on existing authoritative resources, such as Dolce (Gangemi et al., 2002), Prov-O (Lebo et al., 2013), and FRBR (Tillett, 2005)[5], and represents three aspects

---

[1] https://pdai.info/.

[2] https://github.com/marcostranisci/o-dang

[3] Implicit or explicit.

[4] Analogy, euphemism, context shift, false assertion, hyperbole, oxymoron-paradox, rhetorical question, other.

[5] Prefixes of existing ontologies reused in our model are the following: Dolce (dul), Prov-O (prov), FRBR (frbr). Properties of classes of O-Dang! are introduced by ':'
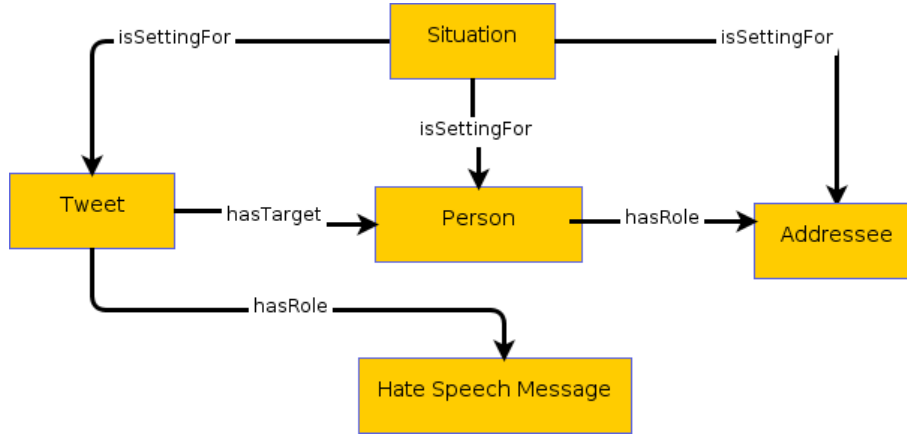
Figure 2: A portion of the O-Dang Semantic Model where a communicative situation with a participant is represented.

about data: (i) the encoding of the annotated text, (ii) the provenance of annotations, (iii) the conversational situation in which the annotated message is present.

A message is encoded as a FRBR:EXPRESSION embedded in one or more FRBR:MANIFESTATION and linked to one or more annotated corpora through the property **dul:isPartOf**. All annotation schemes are represented as subclasses of DUL:DESCRIPTION, since each scheme may be intended as a shared description of a concept between researchers and annotators. As it can be observed in Figure 1, a DUL:ISDESCRIBED as 'Hate Speech' with a specific value. Such a modeling enables the comparison of different schemes adopted for annotating same concepts (Poletto et al., 2021) (eg: binary, scalar). Finally, the **prov:wasAssociatedWith** property links all annotations to their annotators Figure 1 shows two types of annotator: gold_standard, namely a label to identify all aggregated annotations, and a set of individual annotators for researchers interested in querying un-aggregated data from the KG. It is important to notice that no socio-demographic information about annotators is provided within O-Dang!, but only anonymized ids of the type 'annotator_*n*'.

Messages annotated as expressing a given concept are also encoded within a conversational situation, that is a **dul:Situation** in which people, messages, and groups may hold a role. Such representation is focused on the interaction between messages, concept related to Dangerous Speech, and Agents, allowing to query all messages that express a given phenomenon, and have a specific category as a target. In Figure 2 the representation of an HS message may be observed. The Situation DUL:ISSETTINGFOR a Tweet with the role of Hate Speech Message, which is the result of the annotation process depicted in 1. The target of this message has also setting in the same situation with the role of Addressee. Below, an example of materialized triples encoding a HS message against Cécile Kyenge[6] is provided.

```
odang_situation_1342 a :Situation;
    isSettingFor :@ckyenge;
    :isSettingFor [
```

---
[6]She is an Italian politician and ex member of the European Parliament.

```
a :Tweet;
:hasRole HateSpeechMessage;
:hasText ''@ckyenge per fare
sentire    a casa voi africani
e musulmani e stranieri'';
:hasTarget :@ckyenge
];
@ckyenge a :Person;
:hasRole :Addressee;
:gender :female;
:citizenship :ITA;
:placeOfBirth :Kambove.
```

3. Datasets and Ontology Population

The O-Dang! KG includes $898,016$ triples about $62,193$ tweets and $21,972$ users. The Ontology Population stage was performed in two steps: the integration of different data sets in the KG, and a Entity Linking pipeline for the population of the ontology with socio-demographic information about users who are target of Dangerous Speech.

3.1. Dataset Integration

Table 1 shows the datasets that are already populating O-Dang!. As said in Section 1., these corpora are related to Dangerous Speech and parallel phenomena such as irony and stance. For each dataset, we provide the bibliographic reference, textual genres of data, the considered phenomena and the values used to label their presence, and finally the type of annotation ('aggregated' and 'un-aggregated') provided by authors. The un-aggregated annotations reveal the different perspectives or subjectivities on the perception of Dangerous Speech, as well as the difficulty of annotation of the phenomenon and, consequently, of ambiguous cases. For instance, the following news headline (Example 1) was annotated by annotator_1 as hateful and by annotator_2 as non-hateful.

(1) *Alessandria, straniero con ascia e martello aggredisce coppia in casa*
→Alexandria, a foreigner with ax and hammer attacks a couple at home

4 Beyond the clearness of the guidelines, the interpretation of

name	reference	genre	phenomena	annotation	size
IronITA 2018	(Cignarella et al., 2018)	tweets	irony (0/1), sarcasm (0/1), — for some data: type of irony[3] category of irony[4] PoS tags & UD	un-aggregated	4,849
AMI 2018	(Fersini et al., 2018)	tweets	misogyny (0/1), category (stereotype / dominance derailing / sexual harassment discredit), target classification (active/passive)	aggregated	5,000
HaSpeeDe 2018	(Sanguinetti et al., 2018)	facebook posts and tweets	hate speech (0/1)	aggregated	7,996
Hate Speech Corpus	(Sanguinetti et al., 2018)	tweets	hate speech (0/1), stereotype (0/1), aggressiveness (0/1), irony (0/1), intensity (0–>4)	un-aggregated	6,928
SardiStance 2020	(Cignarella et al., 2020)	tweets	stance (against/favor/none) irony (0/1)	un-aggregated	3,242
AMI 2020	(Fersini et al., 2020)	tweets	misogyny (0/1) aggressiveness (0/1)	aggregated	7,000
HaSpeeDe 2020	(Sanguinetti et al., 2020)	tweets and news headlines	hate speech (0/1), stereotype (0/1), aggressiveness (0/1), irony (0/1), sarcasm (0/1) — for some data: offensiveness (0/1), intensity (0–>4) nominal utterances	un-aggregated	8,602
Moral ConvITA	(Stranisci et al., 2021)	tweets	moral stance	un-aggregated	1,722
Populismo Penale	N/A	tweets	stance (against/favor/none)	un-aggregated	12,479
Silvia Romano Corpus	N/A	tweets	stance, abusive language	un-aggregated	4,913
Crowd-HS	N/A	tweets	hate speech (0-7)	un-aggregated	926

Table 1: Summary of the datasets which are already populating O-Dang!

the instances is subjective and relies on the backgrounds of the annotators (Akhtar et al., 2021).

3.2. Entity Linking pipeline

Information about addressees of dangerous messages from Twitter is provided in the KG through an Entity Linking pipeline. Names of each user who is mentioned in a reply have been retrieved through the Twitter API and then searched using Google KG. After a disambiguation process relying on exact string matching between the name provided in input and Google KG output, and on the Google KG score, all the corresponding Wikidata ID were retrieved. Finally, sociodemographic information about each user has been collected from Wikidata. The resulting number of users mapped within the KG is 344. For each, the following information are provided: date of birth, place of birth, country of citizenship, sex or gender, occupation, political party. Below, an example of user associated with such properties is shown.

```
odang_usr_7986 a :Person;
```

```
:hasID 322933929;
:gender female;
:birthYear 1985;
:countryOfCitizenship :ITA;
:placeOfBirth :Lugano;
:occupation :politician;
:politicalParty :DemocraticParty .
```

4. Lexical Analysis

To perform lexical analysis catching the offensiveness of the messages contained in the datasets that at the moment populate O-Dang!, we employed HurtLex (Bassignana et al., 2018). HurtLex is a multilingual lexicon of hateful words created from the Italian lexicon "Le Parole per Ferire" by Tullio de Mauro. The entries in the lexicon are categorized in 17 types of offenses (see Table 2) enclosed in two macro-categories: *conservative* (words with literally offensive sense) and *inclusive* (words with not literally offensive sense, but that could be used with negative connotation). In particular, we considered only the conservative version of the hurtful categories which have been mapped

within O-Dang! through OntoLex-Lemon (McCrae et al., 2017). Each *conservative* word in HurtLex is represented as the following:

```
:IT1241 a :LexicalEntry;
rdfs:label 'fannullone'/'loafer';
lexinfo:partOfSpeech :Noun;
:isDescribed :dmc.
:dmc a :Offensive;
rdfs:label 'moral defects' .
```

The idea is to exploit HurtLex as a means to cross-evaluate the offensiveness of the datasets in the KG (even those which are not annotated expressly as dangerous), and to provide a further description of them.

category	length	description
PS	254	Ethnic Slurs
RCI	36	Location and Demonyms
PA	167	Profession and Occupation
DDP	496	Physical Disabilities and Diversity
DDF	80	Cognitive Disabilities and Diversity
DMC	657	Moral Behavior and Defect
IS	161	Words Related to Social and Economic advantages
OR	144	Words Related to Plants
AN	775	Words Related to Animals
ASM	303	Words Related to Male Genitalia
ASF	191	Words Related to Female Genitalia
PR	138	Words Related to Prostitution
OM	145	Words Related to Homosexuality
QAS	536	Descriptive Words with Potential Negative Connotations
CDS	2042	Derogatory Words
RE	391	Felonies and Words Related to Crime and Immoral Behavior
SVP	424	Words Related to the Seven Deadly Sins of the Christian Tradition

Table 2: HurtLex Categories.

In this way, we have characterized the datasets with HurtLex using a straightforward approach. For each document, the words that are included in each HurtLex category and in the document are counted. This outputs a count for each HurtLex category that is related to a document. To aggregate these counts on a dataset, we average over all documents.

Table 3 shows the result of the described lexical analysis. Such characterization profiles the use of selected HurtLex categories across all datasets. One of the most interesting categories in these datasets, due to its prevalence, is CDS (derogatory words). It can be seen that it is specially relevant in the hate speech and stance datasets.

Continuing with this, the one of the highest metric for the CDS category is obtained in the HaSpeeDe 2018 dataset. Interestingly enough, when looking at this same metric aggregated by annotation class, we see a shift. The HaSeeDe

2018 describes a hate speech binary annotation. For the negative class, the CDS metric has a value of 0.1165 while for the positive class it reaches 0.1546. This observation gives further insight into the language of the data.

5. Conclusion and Future Work

In this paper, we presented O-Dang!, a KG of Italian data sets annotated for Dangerous Speech-related phenomena. The KG includes $62,193$ tweets and $258,704$ annotations both aggregated and un-aggregated. The underlying Semantic Model enables to perform comparative analysis between data sets and phenomena. A first exploratory analysis of offensiveness across corpora has also been provided.

Future work will be devoted to employ this resource to fully inform the systems of abusive language detection, gathering useful pragmatic, semantic and interactional patterns. Moreover, O-Dang! will be integrated with corpora containing different genres of texts in various languages. Finally, a more robust Entity Linking pipeline will be applied, in order to provide more information about Dangerous Speech targets, that may be used for building more explainable systems for abusive language detection.

6. Acknowledgements

7. Bibliographical References

Akhtar, S., Basile, V., and Patti, V. (2021). Whose opinions matter? Perspective-aware models to identify opinions of hate speech victims in abusive language detection. *CoRR*, abs/2106.15896.

Basile, V. (2020). It's the end of the gold standard as we know it. On the impact of pre-aggregation on the evaluation of highly subjective tasks. In Giuseppe Vizzari, et al., editors, *Proceedings of the AIxIA 2020 Discussion Papers Workshop co-located with the 19th International Conference of the Italian Association for Artificial Intelligence (AIxIA2020), Anywhere, November 27th, 2020*, volume 2776 of *CEUR Workshop Proceedings*, pages 31–40. CEUR-WS.

Bender, E. M. and Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Dataset	PS	DDP	DDF	DMC	ASM	ASF	QAS	CDS	SVP
IronITA 2018	0.0080	0.0217	0.0031	0.0431	0.0151	0.0103	0.0210	0.0693	0.0054
HaSpeeDe 2018	0.0228	0.0430	0.0015	0.0540	0.0313	0.0180	0.0350	0.1286	0.0088
Hate Speech Corpus	0.0172	0.0345	0.0013	0.0453	0.0240	0.0141	0.0282	0.1068	0.0079
SardiStance 2020	**0.0275**	0.0584	0.0050	**0.0617**	0.0413	0.0333	0.0333	0.1275	0.0110
HaSpeeDe 2020	0.0199	0.0396	0.0016	0.0548	0.0263	0.0149	0.0307	0.1128	0.0074
Moral ConvITA	0.0175	0.0420	0.0014	0.0467	0.0236	0.0175	0.0331	0.1176	0.0080
Populismo Penale	0.0142	0.0454	0.0015	0.0402	0.0156	0.0260	0.0226	**0.1478**	0.0064
Silvia Romano Corpus	0.0196	**0.0589**	0.0036	0.0429	**0.0429**	**0.0339**	**0.0405**	0.1024	**0.0107**
Crowd-HS	0.0151	0.0356	0.0011	0.0508	0.0140	0.0130	0.0313	0.1274	0.0097

Table 3: Characterization of the O-Dang! datasets using HurtLex.

Benesch, S. (2012). Dangerous speech: A proposal to prevent group violence. *Voices That Poison: Dangerous Speech Project proposal paper*.

Frenda, S., Noriko, K., Patti, V., Rosso, P., et al. (2019). Stance or insults? In *Proceedings of the Ninth International Workshop on Evaluating Information Access (EVIA 2019), a Satellite Workshop of the NTCIR-14 Conference*, pages 15–22. National Institute of Informatics.

Frenda, S., Cignarella, A. T., Basile, V., Bosco, C., Patti, V., and Rosso, P. (2022). The unbearable hurtfulness of sarcasm. *Expert Systems with Applications*, 193.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). Sweetening ontologies with DOLCE. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 166–181. Springer.

Grimminger, L. and Klinger, R. (2021). Hate towards the political opponent: A Twitter corpus study of the 2020 US elections on the basis of offensive speech and stance detection. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 171–180, Online, April. Association for Computational Linguistics.

Leader Maynard, J. and Benesch, S. (2016). Dangerous speech and dangerous ideology: An integrated model for monitoring and prevention. *Genocide Studies and Prevention: An International Journal*, 9(3).

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., and Zhao, J. (2013). Prov-o: The prov ontology.

Lewandowska-Tomaszczyk, B., Žitnik, S., Bączkowska, A., Liebeskind, C., Mitrović, J., and Valunaite Oleskeviciene, G. (2021). LOD-connected offensive language ontology and tagset enrichment. In *Proceedings of the Workshops and Tutorials held at LDK 2021 co-located with the 3rd Language, Data and Knowledge Conference (LDK 2021)*, pages 135–150. CEUR-WS.

McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). The ontolex-lemon model: development and applications. In *Proceedings of eLex 2017 conference*, pages 19–21.

Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., and Chang, Y. (2016). Abusive language detection in online user content. In Jacqueline Bourdeau, et al., editors, *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 145–153. ACM.

Poletto, F., Basile, V., Sanguinetti, M., Bosco, C., and Patti, V. (2021). Resources and benchmark corpora for hate speech detection: a systematic review. *Language Resources and Evaluation*, 55(2):477–523.

Sánchez Rada, J. F., Vulcu, G., Iglesias Fernandez, C. A., and Buitelaar, P. (2014). EUROSENTIMENT: Linked data sentiment analysis. In *Proceedings of the ISWC 2014 Posters & Demonstrations Track a track within the 13th International Semantic Web Conference (ISWC 2014)*, pages 145–148. CEUR-WS.

Tillett, B. (2005). What is FRBR? A conceptual model for the bibliographic universe. *The Australian Library Journal*, 54(1):24–30.

8. Language Resource References

Bassignana, E., Basile, V., and Patti, V. (2018). Hurtlex: A multilingual lexicon of words to hurt. In *5th Italian Conference on Computational Linguistics, CLiC-it 2018*, volume 2253, pages 1–6. CEUR-WS.

Cignarella, A. T., Frenda, S., Basile, V., Bosco, C., Patti, V., and Rosso, P. (2018). Overview of the EVALITA 2018 Task on Irony Detection in Italian Tweets (IronITA). In *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*. CEUR-WS.org.

Cignarella, A. T., Lai, M., Bosco, C., Patti, V., and Rosso, P. (2020). SardiStance @ EVALITA2020: Overview of the Stance Detection in Italian Tweets. In *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. CEUR-WS.org.

Fersini, E., Nozza, D., and Rosso, P. (2018). Overview of the EVALITA 2018 task on automatic misogyny identification (AMI). In Tommaso Caselli, et al., editors, *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, volume 2263 of *CEUR Workshop Proceedings*. CEUR-WS.

Fersini, E., Nozza, D., and Rosso, P. (2020). AMI @ EVALITA2020: Automatic misogyny identification. In Valerio Basile, et al., editors, *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA*

2020), Online event, December 17th, 2020, volume 2765 of *CEUR Workshop Proceedings*. CEUR-WS.

Sánchez-Rada, J. F. and Iglesias, C. A. (2016). Onyx: A linked data approach to emotion representation. *Information Processing & Management*, 52(1):99–114.

Sanguinetti, M., Poletto, F., Bosco, C., Patti, V., and Stranisci, M. (2018). An Italian Twitter corpus of hate speech against immigrants. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Sanguinetti, M., Comandini, G., Di Nuovo, E., Frenda, S., Stranisci, M. A., Bosco, C., Tommaso, C., Patti, V., and Irene, R. (2020). HaSpeeDe 2 @ EVALITA2020: Overview of the EVALITA 2020 Hate Speech Detection Task. In *EVALITA 2020 Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, pages 1–9. CEUR.

Stranisci, M., De Leonardis, M., Bosco, C., and Patti, V. (2021). The expression of moral values in the twitter debate: a corpus of conversations. *IJCoL. Italian Journal of Computational Linguistics*, 7(7-1, 2):113–132.

Westerki, A. and Sánchez-Rada, J. F. (2013). Marl ontology specification. V1.0, May.