How Contentious Terms About People and Cultures are Used in Linked Open Data

Andrei Nesterov
nesterov@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Laura Hollink
hollink@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Jacco van Ossenbruggen
jacco.van.ossenbruggen@vu.nl
VU University Amsterdam
Amsterdam, The Netherlands

ABSTRACT

Web resources in linked open data (LOD) are comprehensible to humans through literal textual values attached to them, such as labels, notes, or comments. Word choices in literals may not always be neutral. When outdated and culturally stereotyping terminology is used in literals, they may appear as offensive to users in interfaces and propagate stereotypes to algorithms trained on them. We study how frequently and in which literals contentious terms about people and cultures occur in LOD and whether there are attempts to mark the usage of such terms. For our analysis, we reuse English and Dutch terms from a knowledge graph that provides opinions of experts from the cultural heritage domain about terms' contentiousness. We inspect occurrences of these terms in four widely used datasets: Wikidata, The Getty Art & Architecture Thesaurus, Princeton WordNet, and Open Dutch WordNet. Some terms are ambiguous and contentious only in particular senses. Applying word sense disambiguation, we generate a set of literals relevant to our analysis. We found that outdated, derogatory, stereotyping terms frequently appear in descriptive and labelling literals, such as preferred labels that are usually displayed in interfaces and used for indexing. In some cases, LOD contributors mark contentious terms with words and phrases in literals (implicit markers) or properties linked to resources (explicit markers). However, such marking is rare and non-consistent in all datasets. Our quantitative and qualitative insights could be helpful in developing more systematic approaches to address the propagation of stereotypes via LOD.

KEYWORDS

linked open data, literals, contentious terms, stereotypes

Disclaimer. In this paper, contentious words and phrases presented in *"quotation marks and italicised"* can be derogatory and offensive. They are provided solely as illustration of the research and do not reflect the opinions of the authors or their organisations.

1 INTRODUCTION

Developers of knowledge graphs and linked open data (LOD) describe machine-readable web resources with literal values to make them also human-readable. Labels, notes, comments, and other literals contain information in natural language, which can be presented to users in interfaces. Some data providers state in their guidelines that developers and contributors should be aware of their word choices when describing resources to stay impartial. For example, Wikidata recommends to its contributors to "keep descriptions neutral by avoiding opinionated or biased terms".[1] The guidelines of The Getty Art & Architecture Thesaurus (AAT), a

controlled vocabulary for the cultural heritage domain, instruct its editors to "be objective" and avoid language that expresses cultural biases or "may be considered offensive by groups of people".[2] Our study demonstrates, however, that outdated and derogatory words, colonial categories, racial slurs, and other contentious terms (for example *"coolie"*, *"colored"*, *"hottentot"*, *"mongoloid"*, *"mulatto"*, *"negro"*, *"transvestite"*) are still being used to describe resources about (historically) marginalised people and cultures in widespread linked open datasets. If such stereotyping language is left unexamined, there is a risk of further propagation of stereotypes in user interfaces [2] and ML algorithms that use literals as training data [16].

We collect empirical evidence on how contentious terms about people and cultures occur in LOD to illustrate the extent of their usage in literals and provide insights into the existing practices of handling potentially stereotyping language. For our study, we adopted a list of English and Dutch terms from a knowledge graph of contentious terminology [21], which describes terms related to historically marginalised peoples (*"gypsy"*, *"eskimo"*, *"indigenous"*), colonial territories (*"Batavia"*), traditions (*"headhunter"*), and other categories. This knowledge graph is based on the publication "Words Matter" [18] produced by professionals from the cultural heritage domain, who recommend whether certain terms referring to groups of people and cultures should be avoided, marked and explained, or replaced with appropriate synonyms in cultural heritage datasets. This expert knowledge allows us to circumvent judging ourselves whether terms are contentious or not. We operationalize the definition "contentious terms" to refer to the terms we extract from the knowledge graph, in which these terms were explicitly defined as contentious.

Our analysis focusses on four datasets with English and Dutch literals: Wikidata, AAT, and the lexical databases Princeton WordNet and Open Dutch WordNet. These datasets encompass a variety of domains and applications, have different structure and level of curation, which can enrich our insights.

We answer two research questions. **RQ1**: In which literals and how often are contentious terms used in LOD datasets? **RQ2**: Do contentious terms in literals have any markers of their contentiousness and if so, what are these markers and how are they given: implicitly (in text of literals next to contentious terms) or explicitly (via specific properties)?

From the four LOD datasets, we extract occurrences of contentious terms in the literals used for labelling (such as "preferred label") as well as descriptive literals (for example, "notes"). Some contentious terms are ambiguous. This leads to retrieving literals in which the found terms are used in a non-contentious sense. For

---

[1]https://www.wikidata.org/w/index.php?title=Help:Description&oldid=1889469468

[2]https://www.getty.edu/research/tools/vocabularies/guidelines/aat_3_4_scope_note. html. See paragraph 3.4.1.5.11. Last accessed on 13.11.2023

example, the ambiguous term *"primitive"* is used in such literals as "primitive data type" and *"primitive society"*. According to the reused knowledge graph, only the latter literal mentions the term in the contentious sense. We apply a word sense disambiguation approach that enables us to gather a subset of resources with more relevant literals mentioning terms in their contentious senses. Additionally, we analyse a smaller, highly reliable set of resources, which were collected manually to ensure that contentious terms in their literals are the closets in meaning and scope to the terms from the knowledge graph.

To identify implicit and explicit markers of contentiousness, we first inspect literals and properties of the manually selected resources. Then, we find similar markers automatically in all the resources we extracted from the four datasets as well as the resources with disambiguated literals.

The contributions of this paper are the following:

(1) Our quantitative findings prove that most of the investigated English and Dutch contentious terms are still being used frequently in LOD datasets. They occur in descriptive property values (for example notes, descriptions, definitions) as well as in labels (preferred or alternative). There are implicit and explicit markers of contentiousness found in all datasets, however, they are rarely used.

(2) Our qualitative results reveal problematic cases when literals, including preferred labels, communicate stereotypes about people and cultures with outdated or derogatory terms and slurs. We describe different ways of how contentious terms are currently being marked in LOD, however, the usage of these markers is not consistent.

Our findings, including the datasets of resources with disambiguated literals and contentiousness markers, are meant as a starting point for researchers, data curators, and developers to design more systematic and reusable approaches of addressing potentially stereotyping terminology on the Web.

2 RELATED WORK

Cultural Bias in Knowledge Graphs. Different types of bias have been investigated in knowledge graphs [11, 15]. Bias may spread to applications powered by knowledge graphs, such as search engines or question-answering systems [5, 28], subsequently influencing end-users' perception of represented information [11, 26]. Knowledge graph embeddings capture bias [8] and propagate it to algorithms, for example, recommender systems [27]. Existing empirical research on bias in knowledge graphs focusses mainly on data imbalances in representation of gender [14, 28], citizenship and race in Wikidata [26]. In our research, we analyse literals that may contain bias expressed with particular contentious terms. This type of bias differs from the data imbalances bias. It is "bias in a cultural context" [11] that propagates prejudices and stereotypes against people and cultures. This bias may come from, but is not limited to, "socio-cultural" or "political and religious" factors [13].

There are different stages on which cultural bias can infiltrate knowledge systems, for example during the creation of ontologies or data [16]. We focus on the stage of literals creation and gather empirical evidence on whether, and if so, how LOD developers address cultural contentiousness.

Table 1: N literals in the analysed LOD-datasets by properties

Dataset	Property	# literals	
		EN	**NL**
Wikidata	skos:prefLabel	86,397,295	62,964,028
	skos:altLabel	7,324,345	1,357,679
	schema:description	83,565,644	78,592,681
AAT	xl:prefLabel\xl:literalForm	49,892	43,788
	xl:prefLabel\rdfs:comment	130	42
	xl:altLabel\xl:literalForm	114,119	28,992
	xl:altLabel\rdfs:comment	342	5
	skos:scopeNote\rdf:value	42,851	34,280
PWN	ontolex:writtenRep	207,272	N/A
	wn:definition ("Definition")	117,791	
	wn:definition ("Examples")	32,990	
ODWN	Lemma writtenForm	N/A	90,897
	Sense definition		50,021
	SenseExamples		31,600
	Synset Definition gloss		32,098

The Role of Literals in Linked Open Data. Literals with natural language have numerous purposes in LOD. They make data more accessible to users providing human-readable information, allowing indexing [10] and search in natural language [4]. Besides the user-experience purposes, literals serve as background information in approaches of entity linking [19] and ontology alignment [3], including approaches with knowledge graph embeddings [9].

Prior research into natural language literals primarily relates to the quality of LOD. A taxonomy to analyse the quality of literals in linked data is presented in [1], in which the quality of "textual strings" is viewed mainly as compliance to the standards of language tagging. The influence of labelling practices on the accessibility of LOD to applications and users is studied in [12] and [7] across several metrics: such as completeness (if all entities carry human-readable information), efficiency of querying triples with literals, unambiguity of labelling properties, and multilinguality of language tags. For the purpose of calculating such metrics, those studies provide the distribution of properties used for labelling in different datasets (including Wikidata in [12]).

Terms used in a knowledge graph inform about its scope, as it is demonstrated in the study, in which terms (labels) of resources are extracted and visualised to profile cultural heritage knowledge graphs [6]. In our work, besides labels, we also extract longer descriptive literals to further investigate how contentious terms occur in knowledge graphs.

3 DATA

As a starting point of our research, we reuse a knowledge graph of contentious terminology [21] based on expert knowledge from the Words Matter publication [18], which we call the Words Matter KG. There are 75 English and 82 Dutch contentious terms described in the Words Matter KG, each term has a link to explanations and suggestions on its usage. In some cases, suggestions for contentious terms provide synonymous terms which are judged to be more appropriate. Apart from the terms, we also reuse links between them and closely related resources from external LOD-datasets provided in the Words Matter KG. Contentious terms are represented as

SKOS-XL labels and connected to 58 English and 63 Dutch entities in Wikidata, 37 English and 27 Dutch concepts in AAT, and 81 synsets in PWN via the property *skos:relatedMatch*. For example, the Wikidata entity Q3254959[3] is about *"race"* in the sense of human categorization by physical features, which is linked to the term *"race"* in the Words Matter KG, because it is related to the term's meaning. Such related resources were collected by human annotators.

We extend the list of terms with their inflected forms. For example, the canonical term *"aboriginal"* also has the plural form *"aboriginals"*. With canonical and inflected forms, the resulting list includes 154 English and 242 Dutch terms. We retrieve literals containing terms from this list in the following LOD-datasets:

Wikidata, one of the largest user-generated knowledge bases on the Web with 12,5 billion triples,[4] in English and Dutch;

The Getty Art & Architecture Thesaurus (AAT), a controlled vocabulary[5] that serves as a reference for cultural heritage institutions,[6] in English and Dutch;

Princeton WordNet (PWN) [17], a lexical database for English, which is frequently used in ML tasks, such as word sense disambiguation [23];

Open Dutch WordNet (ODWN) [24], a lexical database for Dutch, which is a part of the larger Open Multilingual WordNet[7] and other dictionaries, for example BabelNet.[8]

We extract literals from property values that give main names for resources (labels) as well as descriptive properties. Table 1 presents an overview of the selected properties in each dataset and overall counts of their literal values.[9] Wikidata uses redundant properties for defining labels: we use *skos:prefLabel* ignoring the *rdfs:label* and *schema:name* properties with identical values. Alternative labels and descriptions are defined using *skos:altLabel* and *schema:description*. Similar to Wikidata, AAT provides preferred and alternative labels for concepts but using SKOS-XL,[10] which represents labels as URIs with literal values given via *xl:literalForm*. In a few cases, there is additional textual information about labels contained in the values of *rdfs:comment*. We also search for terms in literals of the *skos:scopeNote* property (via the *rdf:value* path).

In PWN (version 3.1), we search contentious terms in lemmas, definitions, and examples of synsets. Lemmas represent canonical forms of words similar to dictionary entries, and synonymous lemmas are grouped in synsets. The RDF representation models lemmas with the property *ontolex:writtenRep* and synset definitions with the property *wn:definition*. Examples are not represented separately, but included in definition texts and separated with double quotes. In our analysis, we separate definitions from examples.

ODWN (version 1.3) is available only in XML format. Although, it is not released as LOD, we have included ODWN in the study,

because it is the Dutch language counterpart of PWN and it constitutes other WordNets and dictionaries. We provide the names of selected XML tags with literals in Table 1. Lemmas in ODWN are contained in the attribute value of the *Lemma writtenForm* tag, which is a child of *LexicalEntry*. Lexical entries have child elements *Senses*, which are connected to synsets. Unlike in PWN, there are two types of definitions in ODWN: definitions attached to senses (*Sense definition*) and definitions of synsets (the attribute value of *Definition gloss* with the parent *Definitions*). Examples are given for senses and explicitly separated from definitions with the tag *SenseExamples*; we used its both child tags *textualForm* and *canonicalForm*. For ODWN, we ignored the English literals.

Note that while English and Dutch are the two languages with the largest number of preferred labels and descriptions in Wikidata; and the largest number of preferred labels, alternative labels, and scope notes in AAT; there are significantly fewer Dutch alternative labels than English in both datasets (see Table 1).

4 APPROACH

We construct three sets of literals. In this section, we first explain how we reuse related resources from the Words Matter KG to create Set 1. Then, we describe the process of querying English and Dutch contentious terms in 4 datasets (Set 2). The literals from the query results required disambiguation. We explain and evaluate our disambiguation approach, which helped us to generate a subset of literals more relevant to our analysis, which we call Set 3. Figure 1 illustrates the relationships between the Words Matter KG and the three sets. Lastly, we explain how implicit and explicit markers of contentiousness were identified in each set.

4.1 Reusing Related Resources (Set 1)

From the Words Matter KG, we take URIs of the related resources in Wikidata, AAT, and PWN. Since links between contentious terms and ODWN were not yet available, we add 65 links between contentious terms and lexical entries in ODWN using the same guidelines as in [21]. By retrieving the literals of all related resources, we construct Set 1. The results from this set are highly reliable, because the manual identification of resources aimed at ensuring that contentious terms in their literals are closest in meaning and scope to the terms from the Words Matter KG.

Besides analysing Set 1, we also use it as background information of contentious terms during disambiguation to construct Set 3 (see Step 1 in Subsection 4.3).

4.2 Querying Literals in LOD (Set 2)

From each LOD dataset, we retrieve resources that have contentious terms in the literal values of the labelling and descriptive properties. In this subsection, we explain how we queried the literals in each dataset. The retrieval pipelines are packaged in an open license Python module which we call *LODlit*.[11] It serves as a tool to reproduce our results and can be reused for other tasks.

Wikidata. We use the MediaWiki Action API[12] to retrieve Wikidata entities. The API limits requests to 10K results. Because some

---

[3]https://www.wikidata.org/w/index.php?title=Q3254959&oldid=1923044725

[4]https://lod-cloud.net/dataset/wikidata. Last accessed on 30.08.2023

[5]https://www.getty.edu/research/tools/vocabularies/aat/about.html. Last accessed on 13.11.2023

[6]https://pro.europeana.eu/post/europeana-enriches-its-data-with-the-art-and-architecture-thesau. Last accessed on 13.11.2023

[7]https://omwn.org/. Last accessed on 13.11.2023

[8]https://babelnet.org/. Last accessed on 13.11.2023

[9]Wikidata counts: https://www.wikidata.org/w/index.php?title=User:Mr._Ibrahem/Language_statistics_for_items&direction=next&oldid=1816384208

[10]https://www.w3.org/TR/skos-reference/skos-xl.html. Last accessed on 13.11.2023

[11]https://github.com/mnn-001-p/LODlit

[12]https://www.mediawiki.org/wiki/API:Main_page. Last accessed on 13.11.2023
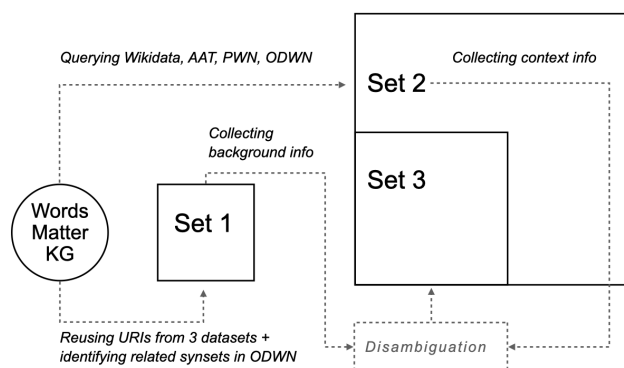
Figure 1: Set 3 with disambiguated literals is a subset of the largest Set 2 with all retrieved literals. Background information taken from the literals in Set 1 is used in disambiguation.

terms yield more search results, we use the top 10K ranked by incoming links. The search results were dominated by instances of a few over-represented categories likely irrelevant to our analysis. Therefore, we take three additional steps to filter these out and increase the relevance of the results. First, we exclude entities belonging to 10 selected categories. Excluded are, for example, instances and subclasses of "scholarly article" (Q13442814) and "taxon" (Q16521). Second, we exclude entities with the phrases "scholarly" or "scientific article" in their descriptions, because many scientific articles are not categorised as such with properties. Third, we filter out entities for which the query term appears as a person's name. For example, the terms "black" and "page" appear in the names Jack Black and Brian S. Page. For this purpose, we filter out entities if the value of P31 ("instance of") is Q5 ("human") AND if the preferred label contains a capitalised query term. For the complete list of excluded categories and the breakdown of search results before and after filtering, we refer to the research documentation.[11] Wikidata resources were retrieved on 31.01.2023.

AAT. The Getty Research Institute provides a SPARQL endpoint, which we use to construct English and Dutch subgraphs of the AAT.[13] Then, we query the subgraphs for concepts that have a contentious term in their literal values.

PWN. From Princeton WordNet 3.1, we retrieve all synsets with contentious terms using Python's NLTK package.

ODWN. We query Open Dutch WordNet using the dedicated Python module.[14] Because not all properties could be queried with the native module functions, some changes were made, which we documented and archived.[15]

Aggregating Search Results. We count the number of occurrences (hits) of each contentious term per dataset, language, and property. For example, the query "slave" has 21 hits in English literals of AAT appearing 2 times in preferred labels, 17 times in alternative labels, and 2 times in scope notes. The plural form "slaves" has

19 hits in English AAT. For the analysis, we group terms by their canonical form, so the term "slave" has 40 hits in total. A resource may have multiple values of one property that mention the same contentious term. For example, a resource has two alternative labels *"slave owner"* and *"slave master"*. We count these as separate hits. If a single literal mentions the same contentious term multiple times, we count this as one hit (for example, the difinition from PWN: *"relating to or involving slaves or appropriate for slaves or servants"*).

4.3 Disambiguating Literals (Set 3)

Some contentious terms appear in LOD datasets in more than one sense. This may lead to a large number of irrelevant results. For example, the term "colored" appears in more than 10 thousand literals we retrieved. Many of these are not in the contentious sense "non-white people" but rather in the sense of "having colour, colourful". To construct a set with more relevant literals mentioning contentious terms, we perform word sense disambiguation (WSD).

Before designing our disambiguation process, we tested off-the-shelve entity linking tools, such as Babelfy [20] and Falcon [25], which link concepts mentioned in an input text to concepts from knowledge bases. However, in our case, these tools did not resolve the challenges associated with our datasets' size (limits of the Babelfy API), the language and varying length of literals (Falcon is designed for short English texts).

To disambiguate English and Dutch literals en masse, we set up a WSD process based on the approach of calculating similarity between the word vectors of a target term's context and its sense inventory [22]. Instead of a sense inventory, we use terms' background information specifying their contentious senses.

Our WSD process includes four steps. On Step 1, we experimented with an alternative source of background information – descriptions of contentious terms provided in the Words Matter KG. We also experimented with common token overlap between background and context information of a term instead of calculating cosine similarity scores between word vectors. The experiments showed that the steps we describe below produced better results.[16]

Step 1. Collecting Background Information. For each contentious term, we collect background information from the literals of its related resources in Set 1. For Wikidata resources, if a contentious term occurs in an entity's labels, we extend the background information with the literals of the properties "instance of" (P31) and "subclass of" (P279). For ODWN resources, if a contentious term was found in the *Synset Definition*, we use only that definition and the *Lemma writtenForm* of the synset as background information, but not *Sense definitions* or *Sense examples* of that lemma, because they are specific to lemmas. The background information text is then converted to a bag-of-words (BoWs) by tokenising, lemmatising, and removing non-word characters, digits, and stop words in English and Dutch. Tokens of fewer than 3 characters are discarded.

Step 2. Collecting Context Information. For each retrieved resource in Set 2, we collect context information from the literals of labelling and descriptive properties. Context information is converted into a BoW in the same way as background information.

Step 3. Using Pre-trained Word Vectors. For every token in the BoWs of background and context information, we infer word vectors from the pre-trained open-license language models by Spacy.[17] We use the *en_core_web_lg* and *nl_core_news_lg* models for English and Dutch tokens respectively. Subsequently, each BoW is represented as an average of all token vectors it contains.

Step 4. Ranking by Cosine Similarity Scores. We calculate cosine similarity (CS) between the context vector and the corresponding background vector for each resource in Set 2. The higher the CS, the more likely the resource mentions the contentious term in the sense intended in the Words Matter KG. We filter out resources that are likely not relevant setting a CS threshold to 0.5. To construct Set 3 with disambiguated literals, we rank all resources by CS and take maximum 10 resources with the highest scores (top-10) per term's canonical form. Some groups of terms had a small number of resources, which led to fewer than 10 entities per canonical form added to the set.

Evaluation. We annotated a sample of the resulting set. Two annotators, who are co-authors of this paper, checked whether the sense of the contentious term mentioned in the resource is similar to the sense of the same term in the provided background information. We used a stratified sample to ensure that both common and rare terms were included in the evaluation. We divided (canonical) terms into quartiles based on the number of search results they yielded. For example, rare terms, such as *"coolie"* or *"half-blood"*, are in the first quartile, while terms with a lot of results, such as *"black"*, are in the fourth quartile. Then, we drew 10 random resources from each quartile, which resulted in 40 resources per dataset per language. Thus, each annotator checked 240 resources for the four datasets (two of which are in two languages).

Inter-rater agreement was high with a Krippendorff's α of 0.80. Annotators deemed 72% of the resources in the sample relevant. In addition, as a form of common sense evaluation, we checked whether Set 3 included the closely related resources from Set 1. Intuitively, these resources should be included, because they would have a high CS score. For Dutch, all related resources were included. For English, all resources from PWN were included; from AAT, all but 1 were included; and for Wikidata, all but 8 were included.

4.4 Identifying Markers of Contentiousness

Some resources contain information about contentiousness of terms used in their literals. We call such information "markers of contentiousness", which can be implicit and explicit. Implicit markers are words and phrases found in literals alongside contentious terms in the same or different property values of a resource. As an example, the synset *"fagot.n.01"* from PWN with the term *"queer"* in lemmas is defined as "offensive term for a homosexual man". The word "offensive" is an implicit marker. Properties and relations between resources indicating that a resource uses contentious terms are explicit markers. For example, the property P31 ("instance of") in Wikidata links the entity Q1135775 (*"redneck"*) to another entity Q545779 ("pejorative"). The principal difference between implicit and explicit markers is that the latter have machine readable URIs or tags opposed to implicit markers requiring textual search.

Table 2: N hits of contentious terms in 3 result sets

Dataset	Properties	Set 2		Set 3		Set 1	
		EN	NL	EN	NL	EN	NL
Wikidata	skos:prefLabel	84,261	7,327	364	321	44	51
	skos:altLabel	18,126	1,518	306	132	67	53
	schema:description	87,058	33,885	210	137	4	0
AAT	xl:prefLabel\xl:literalForm	526	295	58	55	23	25
	xl:prefLabel\rdfs:comment	6	0	1	0	0	0
	xl:altLabel\xl:literalForm	1,428	94	155	14	60	11
	xl:altLabel\rdfs:comment	6	0	1	0	0	0
	skos:scopeNote\rdf:value	5,742	3,568	342	279	10	6
PWN	ontolex:writtenRep	230	N/A	124	N/A	90	N/A
	wn:definition ("Definition")	5,148		244		11	
	wn:definition ("Examples")	578		126		36	
ODWN	Lemma writtenForm	N/A	147	N/A	86	N/A	66
	Sense definition		219		70		2
	SenseExamples		433		171		38
	Synset Definition gloss		190		62		0

We identify implicit and explicit markers of contentiousness in all sets. First, we manually collect both the words associated with implicit markers (for example, "offensive", "derogatory", "outdated"), properties of explicit markers present in Set 1, and additional markers we found while querying datasets. Second, we automatically collect markers in Sets 2 and 3: for implicit markers we use string matching with regular expressions, for explicit – SPARQL (in Wikidata and AAT) and custom Python functions (in PWN and ODWN).

Besides the markers of contentiousness, we search for words and phrases that were suggested to be used instead of contentious terms in the Words Matter KG. We find suggestions in Sets 1 and 3 using fuzzy string matching. The usage of these suggestions together with contentious terms in the same or different property values can be also indicative of terms' contentiousness.

5 RESULTS

For each of the 3 sets, we report how frequently contentious terms occur in all datasets' property values (Table 2). We answer why some datasets, properties, and terms stand out with a large number of hits, check whether synonyms of contentious terms are also used in literals, and provide examples of potentially problematic situations. Separately, we report on which implicit and explicit markers of contentiousness we identified, how frequently they are used, and what they indicate (Table 3). For Set 2 with all query results, we present figures with the most frequent contentious terms in Appendix C. Additionally, there are interactive figures with distribution of hits by properties for each contentious term in the online appendix to this paper.[18]

5.1 Set 1: Closely Related Resources

This set includes only resources closest in meaning to contentious terms. There are 176 and 155 resources with English and Dutch literals, respectively, in which we found 345 English and 252 Dutch hits of contentious terms ("Set 1" in Table 2). 6 English and 5 Dutch terms did not have related resources in any of the datasets. AAT had the least related resources leaving 34 English and 55 Dutch terms without links to corresponding concepts, which was unexpected, since the terms originate from the cultural heritage domain. The terms without related resources are listed in Appendix A.

---

[17]https://spacy.io/models. Last accessed on 13.11.2023

[18]https://github.com/mnn-001-p/LODlit

Primarily, contentious terms in both languages are found in preferred and alternative labels of Wikidata and AAT as well as in lemmas of PWN and ODWN. This is expected since closely related resources mention contentious terms in labels.

Resources with contentious terms in preferred labels often denote (groups of) people, for example *"coolie"* (Q548135), *"mulatto"* (Q191923), *"transvestite"* (Q112918934), *"pygmy"* (Q171927) in Wikidata, *"homosexuals (people)"* (300435115), *"Indiaans"* (*"Indians"* in Dutch, referring to Native American people) (300017437) in AAT, *"mongool"* (*"mongoloid"* in Dutch, referring to people with Down syndrome) (synset eng-30-10197525-n) in ODWN. Such preferred labels are displayed online as standard names of resources, which creates risks of propagation of stereotypes. In some cases, the resources are described as *terms*, for example Wikidata's *"Colored"* (Q5149038) with the description "Term used in the United States to describe black people"[19] or the entity Q235155 with the Dutch preferred label *"blanken"* (*"whites"*) and the definition "term voor mensen met zichtbare Europese oorsprong" ("term for people with visible European origin").[20] Even in these dictionary-like resources, the terms are not always defined as contentious.

As alternative labels, contentious terms are used frequently in Wikidata and AAT. First, they are used as synonyms of preferred labels, as in the Wikidata entity "Mumbai" (Q1156) with the alternative label *"Bombay"*, which is the city's outdated name with colonial connotations. In this case, contentious alternative labels might be used for discoverability purposes, so that users can find a concept even if they query outdated terms. Not all resources, however, contain information about the contentiousness of alternative labels and usually present them as interchangeable synonyms. Second, an alternative label can be a spelling variation or a word form of a preferred label (which can be both contentious), for example, the AAT concept 300386060 with the Dutch preferred label *"hermafrodieten"* (*"hermaphrodites"*) and the alternative label *"hermafrodiet"* (*"hermaphrodite"*). In Dutch literals of AAT, there are more preferred labels with contentious terms than alternative.

We checked the pairs of preferred and alternative labels in Wikidata and AAT: whether there were suggestions from the Words Matter KG in alternative labels when contentious terms were used as preferred and vice versa. In the first case, there is only 1 resource from Wikidata, which is *"Berbers"* (Q45315) with alternative labels *"Berber"*, "Amazigh", "Imazighen".[21] The last two terms are suggested to be used instead of *"Berber"* in the Words Matter KG because more people now refer to themselves as Amazigh (while "Berber" remains an identity category for some people). Since the date of retrieval of this entity, its labels were edited several times, and the latest edit placed the term "Amazigh" as preferred label.[22] The second case has 6 resources from Wikidata and 2 from AAT. For example, the Wikidata entity "enslaved person" (Q12773225) had the alternative label *"slave"*,[23] which was also edited since the retrieval, but the latest edit placed the contentious term *"slave"* as preferred label[24] (in contrast to the example with "Amazigh"). These edits

illustrate the ongoing discussions about the terms' usage among the Wikidata contributors.

PWN and ODWN, besides lemmas, also frequently mention contentious terms in definitions and examples of synsets. This is because in both datasets definitions often repeat the term they define and examples illustrate how a term can be used in speech. In these cases, contentious terms are also often left unexplained in definitions and examples, which reinforces stereotypes. For example, PWN defines the synset *"mentally_retarded.n.01"* as *"people collectively who are mentally retarded"*. And ODWN provides an example for the term *"zigeuner"* (*"gypsy"*) in the synset "eng-30-10154186-n": *"er uitzien als een zigeuner"* (*"looking like a gypsy"*).

5.2 Set 2: All Retrieved Resources

Querying terms in four datasets resulted in more than 203,000 English and 47,000 Dutch hits. In Wikidata, we found hits for each query term. In other datasets, some terms were absent. Appendix B lists terms without hits.

Because of term ambiguity, Set 2 contains literals with terms in contentious meanings, but also other meanings irrelevant to our analysis. Among the top-10 most frequent terms in all datasets (Appendix C), we found ambiguous as well as non-ambiguous terms. Such terms as "black" and "white", ambiguous in both languages, are frequent across all datasets often denoting colour attributes not related to contentious categories of people's skin colour. An example of a non-ambiguous term is "ethnic group", which occurs in more than 4,000 English and 3,000 Dutch[25] literals of Wikidata entities describing peoples. There is a separate entity *"ethnic group"* (Q41710) in Wikidata used for categorisation of groups of people. In AAT, *"ethnic group"* appears in English and Dutch scope notes of more than 400 concepts. The Words Matter KG explains that the term *"ethnic group"* should be used with caution, because it is usually associated with minority groups and often racialized.

Hits in Set 2 appear in all datasets' property values, except *rdfs:comment* used for Dutch labels in AAT ("Set 2" in Table 2). The majority of the hits is contained in descriptive properties: English and Dutch *skos:scopeNote* of AAT, *schema:description* of Dutch Wikidata, definitions and examples of PWN and ODWN. There are fewer hits in Dutch literals of Wikidata and AAT. This can be due to fewer Dutch literals overall, as Table 1 shows.

Wikidata stands out with a large number of hits in English preferred labels (84,261), where in Dutch, there are 11 times fewer hits. We found two large categories (values of the "instance of" property) of entities only with English preferred labels without Dutch translation. These are "Wikimedia category" (Q4167836) with 18,100 hits and "collection" (Q2668072) with 10,400 hits. The first is used to categorise articles in Wikipedia and the second represents archival records. Entities of these two categories frequently mention the terms "colored" and "descent" in their preferred labels (Figure 2 in Appendix C). The term *"colored"* is used in preferred labels of the "collection" entities (8,552 hits out of 9,665) representing records about soldiers from the American Civil War period, for example, soldiers from the *"1st US Colored Infantry"*. These entities

---

[19]https://www.wikidata.org/w/index.php?title=Q5149038&oldid=1745131299 In later revisions, information about the term's offensiveness was added to the description.

[20]https://www.wikidata.org/w/index.php?title=Q235155&oldid=2005925948

[21]https://www.wikidata.org/w/index.php?title=Q45315&oldid=1822618270

[22]https://www.wikidata.org/w/index.php?title=Q45315&oldid=1945487562

[23]https://www.wikidata.org/w/index.php?title=Q12773225&oldid=1802607889

[24]https://www.wikidata.org/w/index.php?title=Q12773225&oldid=1950372555

[25]There are more than 9,000 entities with Dutch descriptions mentioning the term *"etnische groep"*, but 5,600 of them have missing preferred labels, so it is not possible to determine exactly what these entities denote.

were contributed to Wikidata by the U.S. National Archives. In the "Wikimedia category" entities, the term *"descent"* in preferred labels (7,618 hits out of 9,555) denotes people's background, such as *"Category:Canadian people of Chinese descent"* (Q7031448). Other frequent categories of entities without Dutch preferred labels include titles of literary work, art, and digital media products.

5.3 Set 3: Resources with Disambiguated Literals

Disambiguation of all search results produced a subset with 2,307 unique resources. Almost half of them (1,109) are from Wikidata. As our evaluation confirmed, 72% of 240 annotated resources mention terms in a contentious sense. If we extrapolate this proportion to the whole Set 3, it would result in more than 1,600 unique resources with contentious terms in their literals.

Similar to Set 2, disambiguated literals with contentious terms belong to *skos:scopeNote* of AAT, definitions and examples of PWN and ODWN. In Wikidata, contentious terms primarily occur in preferred labels: 364 English and 321 Dutch hits ("Set 3" in Table 2).

As we did in Set 1, we searched for synonyms of contentious terms in the pairs of preferred and alternative labels of Wikidata and AAT. In both datasets, there are no other resources (except already included in Set 1) when a contentious term in a preferred label has a suggested synonym from the Words Matter KG in alternative labels. Only two more resources in Set 3 use suggested terms as preferred when a contentious term is an alternative. We can conclude that these patterns are rare.

Among the disambiguated literals, we found more resources with contentious preferred labels that are potentially problematic (according to the Words Matter KG): for example, *"Eskimo"* (Q131242), *"Hottentot"* (Q1631241), *"Half-breed"* (Q17144151), *"Mongoloid"* (Q207912), *"gekleurde"* (*"colored"*, about a person) (Q2072081), *"zwarten"* (*"blacks"*, about people) (Q817393) from Wikidata; *"Negro spirituals"* (300393224), *"male homosexuals"* (300435114), *"Pygmee"* (*"Pygmy"*) (300016430), *"dwergen"* (*"dwarfs"*) (300236748) from AAT.

In AAT, there are significantly more hits in English alternative labels (155) than preferred (58). This is due to variations of the same label given as alternatives, for example, *"Canadian Eskimo"* and *"Eskimo, Canadian"* (300017455).

PWN and ODWN primarily use contentious terms in definitions and examples of synsets, even when the synsets are not associated with contentious terms. For example, PWN provides an example sentence *"The gypsies roamed the woods"* for the verb "roam" (synset "roll.v.12") and the Dutch lemma "soulmuziek" ("soul music") in ODWN is defined as *"muziek onstaan bij negers"* (*"music originating with negroes"*). In total, there are 370 and 233 potentially problematic definitions and examples in PWN and ODWN, respectively.

5.4 Markers of Contentiousness

Implicit Markers. We identified 31 resources with implicit markers in Set 1 out of 261 unique resources (12%). Most of these resources come from Wikidata (12); no implicit markers were found in ODWN. Implicit markers were given via descriptive properties: *schema:description* in Wikidata, synset definitions in PWN, *skos:scopeNote* and *rdfs:comments* in AAT. Querying literals of these properties, we retrieved 74 and 20 resources more with implicit

markers from Sets 3 and 2, respectively. Most implicit markers are only in English. In Dutch, they occurred in 19 resources of all sets.

Depending on what implicit markers indicate, we categorised them by 6 types (Table 3) . In almost a third of the cases, implicit markers indicate that a term carries offensive meaning. For example, PWN defines the term *"coolie"* as "(ethnic slur) an offensive name for an unskilled Asian laborer" in the synset *coolie.n.01.* In some resources, AAT contained implicit markers in long texts of scope notes with explanations and suggestions on terms' usage (Type #6). For example, the concept *"Eskimo (culture or style)"* (300017447) has a scope note "For names of specific native peoples of the present, use descriptors such as "Chugach," "Inuit," or "Katladlit."". These suggestions are similar to those given in the Words Matter KG.

Explicit Markers. 36 resources in Set 1 (14%) contained properties, relations to other resources, and tags explicitly indicating contentiousness. Furthermore, we retrieved 121 resources in Set 2 and 28 resources in Set 3 by querying the identified properties and tags used for explicit marking.

Comparable to implicit markers, explicit markers also come in different types (Table 3). In Wikidata, several entities are used to mark contentiousness. For example, the entity *"mulatto"* (Q191923) is an "instance of" (P31) "historical race concept" (Q2042898). The Wikidata property "Wikidata usage instructions" (P2559) connects entities to information on how contentious terms should be used. AAT flags some preferred and alternative labels with the properties *gvp:termKind* and *gvp:historicFlag*. The values of *gvp:termKind* are, for example, "AvoidUse", "JargonOrSlang", or "Pejorative". And *gvp:historicFlag* has only two values "historic" and "currentAndHistoric". The synsets of PWN are connected via *usage_domain* to other synsets, for example, *disparagement.n.01* or *slang.n.02*. The ODWN tag *Pragmatics* has several attributes, including "connotation" and "chronology" with values indicating offensiveness ("pejorative", "offensive") or that a term is "oldfashioned".

By querying explicit markers outside of the retrieved resources, we found that the same markers were used for purposes other than signalling contentiousness. For example, "Wikidata usage instructions" (P2559) explains technical details on how qualifiers should be used. In AAT, the property *gvp:termKind* also has values not related to contentiousness, such as "Misspelling" or "Abbreviation". And the value "historic" of *gvp:historicFlag* also marks historical spelling of terms. The *usage_domain* in PWN indicates, besides offensive lemmas, comparative and plural forms. Such weak semantics of these properties complicate the identification of resources with contentious terms in their literals.

Resources Without Markers. In Set 3, we checked whether contentious terms in preferred labels of Wikidata and AAT resources and in lemmas of PWN and ODWN synsets had any markers. Resources are rarely marked either implicitly or explicitly in all datasets. Out of 657 Wikidata resources with disambiguated preferred labels, only 21 were marked. In AAT, this proportion is 91/14. For example, the Wikidata entity *"Pygmy people"* (Q171927) does not have any markers, while in AAT, the concept *"Pygmy (African culture or style)"* (300016430) has a scope note with an implicit marker stating that "Use of *"Pygmy"* for a culture is considered pejorative". At the same time, this AAT concept has the Dutch preferred label *"Pygmee"* (*"Pygmy"*), which is not marked. This

Table 3: Types of implicit and explicit markers of contentiousness collected from LOD datasets

Type (a marker indicates:)	Implicit	Explicit
#1: offensiveness	EN: "offensive", "pejorative", "derogatory", "slur", "disparaging term", "denigrating"; NL: "denigrerend" ("derogatory"), "scheldwoord" ("swear word"), "negatieve bijklank" ("negative connotation")	"pejorative" (Q545779), "slur" (Q22116852) in Wikidata; "Pejorative" in AAT (*gvp:termKind*); "disparagement.n.01" in PWN; "pejorative" and "offensive' in ODWN (*Pragmatics connotation*)
#2: historical usage	EN: "obsolete", "historical usage", "archaic", "antiquated", "used formerly", "older term"; NL: "historische" ("historic"), "verouderde benaming" ("obsolete designation")	"historical race concept" (Q2042898), "historical profession" (Q16335296) in Wikidata; the property *gvp:historicFlag* and "Deprecated" (*gvp:termKind*) in AAT; "oldfashioned" (*Pragmatics chronology*) in ODWN;
#3: informal speech	EN: "informal", "slang", "colloquial"; NL: "informele term" ("informal term")	"slang.n.02", "colloquialism.n.01" in PWN; "JargonOrSlang" (*gvp:termKind*) in AAT
#4: (self-)identity categories	EN: "term of self-reference", "identity term", "self-identifying term"; NL: "zichzelf aanduiden" and "noemen zichzelf" ("call themselves")	"reappropriation" (Q1520214) in Wikidata
#5: stereotypes incl. racism	EN: "stereotypical", "based on stereotypes", "ethnic slur", "racialized classification"; NL: "racistische" ("racist")	"ethnic slur" (Q1371427) in Wikidata; "ethnic_slur.n.0" in PWN
#6: usage suggestions	EN: "use instead [alternative]", "now prefer [alternative]", "consider", "use specific terms"; NL: "gebruik een specifiekere term" ("use a more specific term")	"AvoidUse" (*gvp:termKind*) in AAT; literals of "Wikidata usage instructions" (P2559) in Wikidata;

illustrates such cases where even if a contentious term marked as "pejorative", it still can be used as a preferred label and shown to users in interfaces that, for example, ignore scope notes.

In PWN, there are 117 disambiguated lemmas with contentious terms, and 11 of them were marked. ODWN marked 10 lexical entries out of 79. The definition of the PWN synset "black.n.05" marks the terms *"Negro"* and *"Negroid"* as "archaic and pejorative today", however the synset "colored.s.02" lists the term *"negro"* as a synonym of *"colored"* and *"coloured"* without markers. The tag "Pragmatics connotation" in ODWN marks the lemma "flikker"(*"faggot"*) and its synonyms as "pejorative" and "offensive" (synset "eng-30-10076033-n"), but the synset "eng-30-10182913-n" with the same lemma (with a different id) does not contain any markers. It results in the situation when the words from this synset are being used in interfaces of online dictionaries,[26] which display offensive terms *"bruinwerker"* (literally *"brown worker"*), *"flikker"* (*"faggot"*), and *"geïnverteerde"* (*"inverted"*) as regular synonyms of *"homosexuality"*.

6 CONCLUSION

Contentious terms that express stereotypes about people and cultures and which are potentially offensive to users appear on a large scale in four widely used linked open datasets, our study showed. We extracted and analysed the literals mentioning contentious terms from Wikidata, The Getty Art & Architecture Thesaurus (AAT) common in the cultural sector, and two lexical databases Princeton WordNet (PWN) and Open Dutch WordNet (ODWN). The terms we studied originate from a knowledge graph, which models judgements of experts from the cultural heritage domain about terms' cultural sensitivities.

From the four datasets, we retrieved more than 203,000 English and 47,000 Dutch literals. Because of term ambiguity, there were literals with terms in non-contentious senses. Applying word sense disambiguation, we constructed a subset with a large proportion of relevant literals, which still contains more than 2,000 resources.

In the set with disambiguated literals, Wikidata has the largest number of resources using both English and Dutch contentious terms as preferred labels. We found preferred labels of Wikidata resources, which refer to people in an outdated and derogatory fashion without (more appropriate) synonyms in alternative labels or explanations about the terms' usage in descriptions. In other datasets, contentious terms are primarily used in descriptive literals: scope notes of AAT, definitions and examples of PWN and ODWN.

Our analysis revealed various attempts of the LOD contributors marking resources with contentious terminology: using special words and phrases in literals (implicit markers) or properties (explicit markers). Implicit markers are not frequent and appear mostly in English literals of descriptive properties. There are language independent properties used for explicit marking in each dataset, however, some of them have weak semantics: they are used for multiple purposes other than signalling contentiousness. In all datasets, we found marking rare and not systematic.

The insights we gathered can serve as a starting point towards compiling more informative guidelines for those who describe data and designing approaches to detect and prevent the propagation of stereotypes on the Web more systematically.

REFERENCES

[1] Wouter Beek, Filip Ilievski, Jeremy Debattista, Stefan Schlobach, and Jan Wielemaker. 2018. Literally better: Analyzing and improving the quality of literals. *Semantic Web* 9, 1 (2018), 131–150. https://doi.org/10.3233/SW-170288

[2] Camiel J. Beukeboom and Christian Burgers. 2020. How Stereotypes Become Shared Knowledge: An Integrative Review on the Role of Biased Language Use in Communication about Categorized Individuals. *Review of Communication Research* 7 (2020), 1–37. https://rcommunicationr.org/index.php/rcr/article/view/51

[3] Michelle Cheatham and Pascal Hitzler. 2013. String Similarity Metrics for Ontology Alignment. In *The Semantic Web – ISWC 2013*, Harith Alani, Lalana Kagal, Achille Fokoue, Paul Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha Noy, Chris Welty, and Krzysztof Janowicz (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 294–309. https://doi.org/10.1007/978-3-642-41338-4_19

[4] Danica Damljanovic, Milan Agatonovic, and Hamish Cunningham. 2012. FREyA: An Interactive Way of Querying Linked Data Using Natural Language. In *The Semantic Web: ESWC 2011 Workshops*, Raúl García-Castro, Dieter Fensel, and Grigoris Antoniou (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 125–138. https://doi.org/10.1007/978-3-642-25953-1_11

---

[26] https://babelnet.org/synset?id=bn%3A00037547n&orig=homoseksuele&lang=NL. Last accessed on 13.11.2023

[5] Gianluca Demartini. 2019. Implicit Bias in Crowdsourced Knowledge Graphs. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 624–630. https://doi.org/10.1145/3308560.3317307

[6] Lamine Diop, Béatrice Markhoff, and Arnaud Soulet. 2023. TTProfiler: Types and Terms Profile Building for Online Cultural Heritage Knowledge Graphs. *J. Comput. Cult. Herit.* 16, 3, Article 61 (aug 2023), 22 pages. https://doi.org/10.1145/3588443

[7] Basil Ell, Denny Vrandečić, and Elena Simperl. 2011. Labels in the Web of Data. In *The Semantic Web – ISWC 2011*, Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Noy, and Eva Blomqvist (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 162–176.

[8] Joseph Fisher, Dave Palfrey, Christos Christodoulopoulos, and Arpit Mittal. 2020. Measuring Social Bias in Knowledge Graph Embeddings. https://www.amazon.science/publications/measuring-social-bias-in-knowledge-graph-embeddings

[9] Genet Asefa Gesese, Russa Biswas, and Harald Sack. 2019. A Comprehensive Survey of Knowledge Graph Embeddings with Literals: Techniques and Applications. In *Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019)* (Portoroz, Slovenia) *(CEUR Workshop Proceedings, Vol. 2377)*, Mehwish Alam, Davide Buscaldi, Michael Cochez, Francesco Osborne, Diego Reforgiato Recupero, and Harald Sack (Eds.). CEUR, Portoroz, Slovenia, 31–40. https://ceur-ws.org/Vol-2377/#paper4

[10] Filip Ilievski, Wouter Beek, Marieke van Erp, Laurens Rietveld, and Stefan Schlobach. 2015. LOTUS: Linked Open Text UnleaShed. In *Proceedings of the 6th International Workshop on Consuming Linked Data co-located with 14th International Semantic Web Conference (ISWC) 2015, Bethlehem, Pennsylvania, USA, October 12th, 2015 (CEUR Workshop Proceedings, Vol. 1426)*, Olaf Hartig, Juan F. Sequeda, and Aidan Hogan (Eds.). CEUR-WS.org, Bethlehem, Pennsylvania, US, 12 pages. http://ceur-ws.org/Vol-1426/paper-06.pdf

[11] Krzysztof Janowicz, Bo Yan, Blake Regalia, Rui Zhu, and Gengchen Mai. 2018. Debiasing Knowledge Graphs: Why Female Presidents are not like Female Popes. In *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), October 8th - to - 12th, 2018 (CEUR Workshop Proceedings, Vol. 2180)*, Marieke van Erp, Medha Atre, Vanessa López, Kavitha Srinivas, and Carolina Fortuna (Eds.). CEUR-WS.org, Monterey, USA, 5 pages. http://ceur-ws.org/Vol-2180/ISWC_2018_Outrageous_Ideas_paper_17.pdf

[12] Lucie-Aimée Kaffee and Elena Simperl. 2018. The Human Face of the Web of Data: A Cross-sectional Study of Labels. *Procedia Computer Science* 137 (2018), 66–77. https://doi.org/10.1016/j.procs.2018.09.007 Proceedings of the 14th International Conference on Semantic Systems 10th – 13th of September 2018 Vienna, Austria.

[13] C. Maria Keet. 2021. An Exploration into Cognitive Bias in Ontologies. In *Proceedings of the Joint Ontology Workshops 2021* (2021-09-11) *(CEUR Workshop Proceedings, Vol. 2969)*, Emilio M. Sanfilippo, Oliver Kutz, Nicolas Troquard, Torsten Hahmann, Claudio Masolo, Robert Hoehndorf, Randi Vita, Maria M. Hedblom, Oliver Kutz, and Guendalina Righetti (Eds.). CEUR, Bolzano, Italy, 17 pages. http://ceur-ws.org/Vol-2969/#paper38-CAOS

[14] Maximilian Klein, Harsh Gupta, Vivek Rai, Piotr Konieczny, and Haiyi Zhu. 2016. Monitoring the Gender Gap with Wikidata Human Gender Indicators. In *Proceedings of the 12th International Symposium on Open Collaboration* (Berlin, Germany) *(OpenSym '16)*. Association for Computing Machinery, New York, NY, USA, Article 16, 9 pages. https://doi.org/10.1145/2957792.2957798

[15] Konstantinos Ilias Kotis, Konstantina Zachila, and Evaggelos Paparidis. 2021. Machine Learning Meets the Semantic Web. *Artificial Intelligence Advances* 3 (2021), 63–70. Issue 1.

[16] Angelie Kraft and Ricardo Usbeck. 2022. The Lifecycle of "Facts": A Survey of Social Bias in Knowledge Graphs. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online only, 639–652. https://aclanthology.org/2022.aacl-main.49

[17] George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38, 11 (nov 1995), 39–41. https://doi.org/10.1145/219717.219748

[18] Wayne Modest and Robin Lelijveld. 2018. Words Matter: an unfinished guide to word choices in the cultural sector. https://www.tropenmuseum.nl/en/about-tropenmuseum/words-matter-publication

[19] Cedric Möller, Jens Lehmann, and Ricardo Usbeck. 2022. Survey on English Entity Linking on Wikidata: Datasets and approaches. *Semantic Web* 13, 6 (2022), 925–966. https://doi.org/10.3233/SW-212865

[20] Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics* 2 (2014), 231–244. https://doi.org/10.1162/tacl_a_00179

[21] Andrei Nesterov, Laura Hollink, Marieke van Erp, and Jacco van Ossenbruggen. 2023. A Knowledge Graph of Contentious Terminology for Inclusive Representation of Cultural Heritage. In *The Semantic Web*, Catia Pesquita, Ernesto Jimenez-Ruiz, Jamie McCusker, Daniel Faria, Mauro Dragoni, Anastasia Dimou, Raphael Troncy, and Sven Hertling (Eds.). Springer Nature Switzerland, Cham, 502–519. https://doi.org/10.1007/978-3-031-33455-9_30

[22] Korawit Orkphol and Wu Yang. 2019. Word Sense Disambiguation Using Cosine Similarity Collaborates with Word2vec and WordNet. *Future Internet* 11, 5 (2019), 16 pages. https://doi.org/10.3390/fi11050114

[23] Tommaso Pasini and Jose Camacho-Collados. 2020. A Short Survey on Sense-Annotated Corpora. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 5759–5765. https://aclanthology.org/2020.lrec-1.706

[24] Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen. 2016. Open Dutch WordNet. In *Proceedings of the 8th Global WordNet Conference (GWC)*. Global Wordnet Association, Bucharest, Romania, 302–310. https://aclanthology.org/2016.gwc-1.43

[25] Ahmad Sakor, Kuldeep Singh, Anery Patel, and Maria-Esther Vidal. 2020. Falcon 2.0: An Entity and Relation Linking Tool over Wikidata. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 3141–3144. https://doi.org/10.1145/3340531.3412777

[26] Zaina Shaik, Filip Ilievski, and Fred Morstatter. 2021. Analyzing Race and Citizenship Bias in Wikidata. In *2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, Denver, CO, USA, 665–666. https://doi.org/10.1109/MASS52906.2021.00099

[27] Michael Matthias Voit and Heiko Paulheim. 2021. Bias in Knowledge Graphs - An Empirical Study with Movie Recommendation and Different Language Editions of DBpedia. In *3rd Conference on Language, Data and Knowledge (LDK 2021) (Open Access Series in Informatics (OASIcs), Vol. 93)*, Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 14:1–14:13. https://doi.org/10.4230/OASIcs.LDK.2021.14

[28] Charles Chuankai Zhang and Loren Terveen. 2021. Quantifying the Gap: A Case Study of Wikidata Gender Disparities. In *17th International Symposium on Open Collaboration* (Online, Spain) *(OpenSym 2021)*. Association for Computing Machinery, New York, NY, USA, Article 6, 12 pages. https://doi.org/10.1145/3479986.3479992

A TERMS WITHOUT RELATED RESOURCES

Wikidata EN (15): *"baboo"*, *"bush negro"*, *"developing nations"*, *"discover"*, *"dwarf"*, *"exotic"*, *"full blood"*, *"half-blood"*, *"headhunter"*, *"homo"*, *"low-income countries"*, *"native"*, *"oriental"*, *"roots"*, *"traditional"*.

Wikidata NL (13): *"achterlijk"*, *"exotisch"*, *"halfbloed"*, *"inboorling"*, *"knecht"*, *"koppensneller"*, *"ontdekken"*, *"oriëntaals"*, *"primitief"*, *"roots"*, *"traditioneel"*, *"trans"*, *"volbloed"*.

AAT EN (34): *"allochtoon"*, *"baboo"*, *"barbarian"*, *"batavia"*, *"bombay"*, *"burma"*, *"bush negro"*, *"calcutta"*, *"colored"*, *"coolie"*, *"descent"*, *"discover"*, *"exotic"*, *"footmen"*, *"full blood"*, *"half-blood"*, *"half-breed"*, *"handicap"*, *"headhunter"*, *"homo"*, *"kaffir"*, *"lilliputian"*, *"low-income countries"*, *"madras"*, *"mestizo"*, *"mongoloid"*, *"mulatto"*, *"page"*, *"roots"*, *"second world"*, *"southern rhodesia"*, *"trans"*, *"western"*, *"white"*.

AAT NL (55): *"afkomst"*, *"allochtoon"*, *"baboe"*, *"barbaar"*, *"batavia"*, *"birma"*, *"blank"*, *"bombay"*, *"boslandcreool"*, *"bosneger"*, *"calcutta"*, *"derde wereld"*, *"eerste wereld"*, *"exotisch"*, *"gay"*, *"gekleurd"*, *"halfbloed"*, *"handicap"*, *"homo"*, *"homoseksueel"*, *"hottentot"*, *"inboorling"*, *"inlander"*, *"islamiet"*, *"jap"*, *"jappenkamp"*, *"kaffer"*, *"knecht"*, *"koelie"*, *"koppensneller"*, *"lagelonenland"*, *"lilliputter"*, *"madras"*, *"marron"*, *"medicijnman"*, *"mesties"*, *"mohammedaan"*, *"mulat"*, *"neger"*, *"ontdekken"*, *"oriëntaals"*, *"page"*, *"politionele actie"*, *"primitief"*, *"queer"*, *"roots"*, *"traditioneel"*, *"trans"*, *"travestiet"*, *"tweede wereld"*, *"volbloed"*, *"westers"*, *"wit"*, *"zuid-rhodesië"*, *"zwart"*.

PWN (20): *"allochtoon"*, *"baboo"*, *"batavia"*, *"bush negro"*, *"developing nations"*, *"ethnic groups"*, *"first world"*, *"full blood"*, *"half-blood"*, *"hottentot"*, *"indo"*, *"low-income countries"*, *"maroon"*, *"medicine man"*, *"métis"*, *"roots"*, *"second world"*, *"southern rhodesia"*, *"third world"*, *"trans"*.

ODWN (25): *"batavia"*, *"boslandcreool"*, *"derde wereld"*, *"eerste wereld"*, *"etnische groep"*, *"exotisch"*, *"gekleurd"*, *"indisch"*, *"indo"*, *"inheems"*, *"lagelonenland"*, *"marron"*, *"métis"*, *"oriëntaals"*, *"politionele actie"*, *"primitivisme"*, *"queer"*, *"ras"*, *"roots"*, *"traditioneel"*, *"tweede wereld"*, *"volbloed"*, *"westers"*, *"wit"*, *"zuid-rhodesië"*.

All datasets EN (6): *"baboo"*, *"bush negro"*, *"full blood"*, *"half-blood"*, *"low-income countries"*, *"roots"*.

All datasets NL (5): *"exotisch"*, *"oriëntaals"*, *"roots"*, *"traditioneel"*, *"volbloed"*.

B TERMS WITHOUT SEARCH HITS

AAT EN (11): *"allochtoon"*, *"baboo"*, *"bush negro"*, *"coolie"*, *"full blood"*, *"half-blood"*, *"half-breed"*, *"headhunter"*, *"lilliputian"*, *"low-income countries"*, *"southern rhodesia"*.

AAT NL (26): *"allochtoon"*, *"baboe"*, *"barbaar"*, *"boslandcreool"*, *"bosneger"*, *"eerste wereld"*, *"halfbloed"*, *"inboorling"*, *"inlander"*, *"jap"*, *"jappenkamp"*, *"kaffer"*, *"koelie"*, *"koppensneller"*, *"lagelonenland"*, *"lilliputter"*, *"marron"*, *"medicijnman"*, *"mesties"*, *"neger"*, *"politionele actie"*, *"queer"*, *"roots"*, *"tweede wereld"*, *"volbloed"*, *"zuid-rhodesië"*.

PWN (7): *"allochtoon"*, *"batavia"*, *"bush negro"*, *"full blood"*, *"half-blood"*, *"low-income countries"*, *"southern rhodesia"*.

ODWN (11): *"boslandcreool"*, *"eerste wereld"*, *"lagelonenland"*, *"marron"*, *"métis"*, *"oriëntaals"*, *"primitivisme"*, *"queer"*, *"roots"*, *"tweede wereld"*, *"zuid-rhodesië"*.

C THE MOST FREQUENT TERMS IN SET 1

The stacked bar charts illustrate the proportion of the property values, in which contentious terms occur. The total number of hits per term is shown on the right.
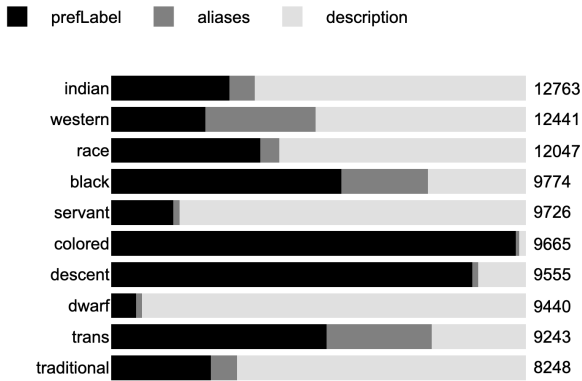
prefLabel · **aliases** · **description**

Term	N hits
indian	12763
western	12441
race	12047
black	9774
servant	9726
colored	9665
descent	9555
dwarf	9440
trans	9243
traditional	8248

Figure 2: Top-10 EN contentious terms by N hits in Wikidata

prefLabel · **aliases** · **description**

Term	N hits
wit	11601
etnische groep	9249
indiaan	9223
homo	3234
zwart	1524
moor	723
ras	722
traditioneel	714
gay	575
indisch	574
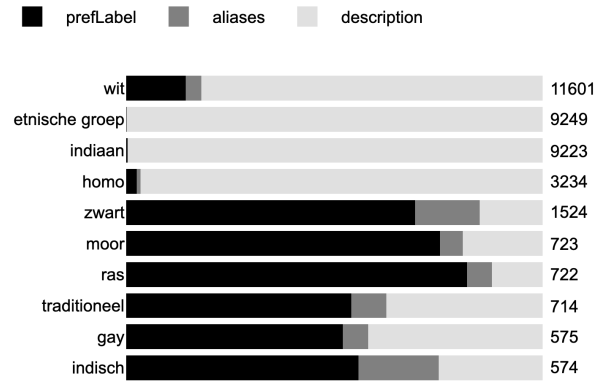
Figure 5: Top-10 NL contentious terms by N hits in Wikidata

prefLabel · **altLabel** · **scopeNote** · **comment (pref)** · **comment (alt)**

Term	N hits
white	1249
native	1230
black	992
western	689
indian	679
ethnic groups	414
traditional	347
colored	316
indigenous	279
tribe	216
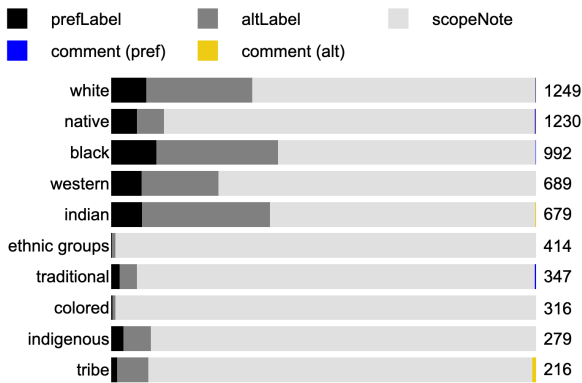
Figure 3: Top-10 EN contentious terms by N hits in AAT

prefLabel · **altLabel** · **scopeNote** · **comment (pref)** · **comment (alt)**

Term	N hits
inheems	750
wit	730
zwart	495
etnische groep	434
traditioneel	313
stam	243
gekleurd	187
westers	137
bediende	97
indiaan	76
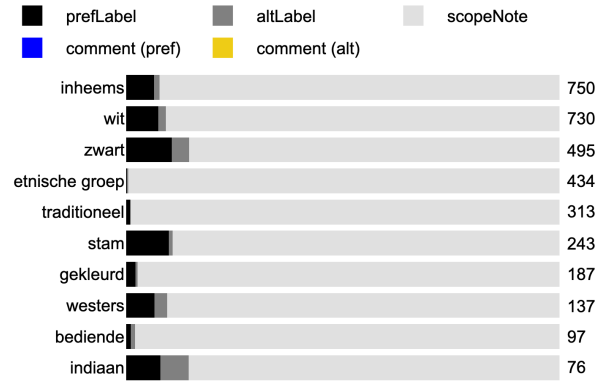
Figure 6: Top-10 NL contentious terms by N hits in AAT

synset lemmas · **definitions** · **examples**

Term	N hits
white	1601
western	744
black	722
native	634
colored	346
indian	319
race	235
roots	207
primitive	137
page	113

Figure 4: Top-10 EN contentious terms by N hits in PWN

LE form · **synset def** · **sense def** · **sense examples**

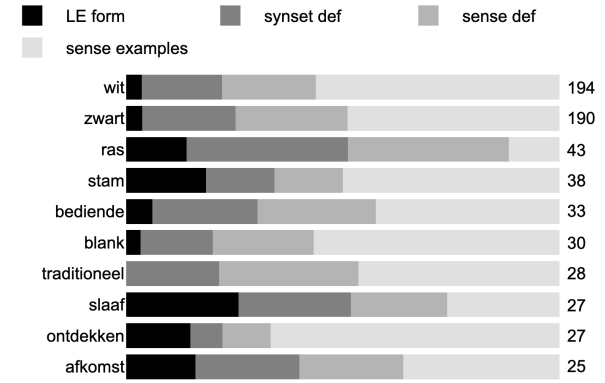| Term	N hits
wit	194
zwart	190
ras	43
stam	38
bediende	33
blank	30
traditioneel	28
slaaf	27
ontdekken	27
afkomst	25

Figure 7: Top-10 NL contentious terms by N hits in ODWN