

# WDProp: Web Application to Analyse Multilingual Aspects of Wikidata Properties

John Samuel

john.samuel@cpe.fr

CPE Lyon, LIRIS, UMR-CNRS 5205, Université de Lyon  
Lyon, France

## ABSTRACT

Compared to Wikipedia, Wikidata is a single domain website with the possibility to view information in multiple languages. Translation plays a significant role in Wikidata. Unlike Wikidata items, Wikidata properties are influenced less by translation bots and require a meaningful amount of human effort. The study of Wikidata property creation and translation is, therefore, very essential. Since the inception of Wikipedia, several research works have focused on the information flow among different language Wikipedias. The attention has now shifted to the way information on Wikidata is created and translated. The focus of this article is the Wikidata properties. WDProp is a web application created to understand and obtain an integrated view on the various multilingual aspects of Wikidata properties, from their proposition to their use on multiple domains.

## CCS CONCEPTS

• **Software and its engineering** → **Software libraries and repositories**; • **Information systems** → **Collaborative and social computing systems and tools**; **Open source software**; **Data mining**.

## KEYWORDS

Wikidata, Wikidata Properties, Collaboration, Multilingual Development

## ACM Reference Format:

John Samuel. 2021. WDProp: Web Application to Analyse Multilingual Aspects of Wikidata Properties. In *17th International Symposium on Open Collaboration (OpenSym 2021)*, September 15–17, 2021, Online, Spain. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3479986.3479996>

## 1 INTRODUCTION

Wikidata [17], a Wikimedia project created in 2012 is a free, linked, open, collaborative, and multilingual knowledge base. There are significant differences between Wikidata and Wikipedia. Unlike

Wikipedia, which contains the majority of information in unstructured form (e.g., textual description), Wikidata is a structured knowledge base. Wikipedia uses multiple subdomains for different languages, each of which is managed by the associated language community. Wikidata, on the other hand, is a single domain website. All the information related to a single topic in multiple languages is referred to by a single URL. Logged-in users may be able to view the information in their local language by configuring the local language in their settings. Users can also make use of *uselang = lc*, in the URL to view the information in any other language *lc*, where *lc* is the language code. This is indeed a major change for Wikipedia users used to multiple websites for looking up information in different languages. Such an approach of building a central store for storing facts (or statements) related to multiple domains is challenging. This approach of collaborative multilingual and multi-domain ontology development on Wikidata has been the focus of many recent works [6, 8, 11].

Nevertheless, it is important to state that not all communication on Wikidata exists in a multilingual way. There exist specific pages where contributors need to use monolingual text for expressing their opinions, like in discussion and User:Talk pages. Whether a truly multilingual experience can be obtained on Wikidata is an open and relevant research question. This article focus on a small, but key aspect of Wikidata: properties to comprehend the multilingual aspects around them. Wikidata properties play a major role in describing knowledge across domains. Unlike Wikidata items, they cannot be simply created but requires discussion and voting. Depending on the need, the properties are proposed by the Wikidata contributors. Such propositions are discussed by the community members, where some may point to the availability of existing properties and others agreeing to the need for the creation of the property. Every property proposition undergoes voting, and if a property has achieved a sufficient number of supporting votes, the property may be created and made available for use.

A newly created property may not be available in all the languages and may have one or more translations based on the initial translations proposed by the property proposer. This makes it difficult for users who do not speak or comprehend any of the available translations. The role played by bilingual or multilingual speakers is therefore very important. It is also difficult to use the properties, immediately after their creation, if their usage is not properly described and documented. For this purpose, Wikidata WikiProjects for different domains have been created, which curates the properties and documents the usage for describing Wikidata items belonging to the given domain. The problem of finding relevant properties has also garnered some attention in the research community. There have been several works on predicting properties

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*OpenSym 2021, September 15–17, 2021, Online, Spain*

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8500-8/21/09...\$15.00

<https://doi.org/10.1145/3479986.3479996>

[3, 18] for describing domain knowledge, including a recent work on their prediction from Wikipedia (e.g., Wiki2Prop [10]). It is important to state here that such property suggestions or predictions may require some validation by the contributors and hence their translations cannot be overlooked.

Translation of Wikidata properties is an important topic, and its analysis in real-time is equally important, given the highly evolving, open, and collaborative nature of Wikidata. Wikidata adds support for new languages and datatypes from time to time. There is no easy way to analyze the multilingual translation of Wikidata properties. There are some static pages/tools to track the datatypes and the properties, but they are not linked to languages and the property translations. In this article, we discuss the development of WDPProp as a solution to these problems. WDPProp bridges this gap between properties and their multilingual translations, giving the users multiple possible ways to navigate and analyze properties and their translations. WDPProp also considers WikiProjects for contributors who are interested in a subset of curated information and the associated translation statistics. Section 2 gives a brief introduction of Wikidata, particularly focusing on properties. Some works on Wikidata properties are presented in section 3. Section 4 presents the development of WDPProp and the associated results. Starting from some coarser analysis, it shows the results related to some fine-granular analysis. Finally, section 5 concludes the article and presents the future course of action.

## 2 WIKIDATA PROPERTIES

Wikidata item pages are used to describe different classes like museum, person, archive, website, conference, scientific event<sup>1</sup> etc. and their instances like OpenSym 2019<sup>2</sup>, Wikidata<sup>3</sup> etc. In addition to Wikidata item pages, there are property pages (e.g., instance of<sup>4</sup>, country (P17)<sup>5</sup>, official website (P856)<sup>6</sup>, etc.), entity schema pages (Wikimedia project<sup>7</sup>) and the recently introduced lexicographical pages (e.g., lexeme (L315)<sup>8</sup>). Properties are used to describe items (e.g., the country of OpenSym 2019 was described using P17). Entity schemas are used to validate the Wikidata properties and items. It is worth mentioning here that all these pages have an identifier of the form [QPLE][0-9][0-9]\*. Unlike Wikipedia, the contributors may not know the topic of the page from the URL, without looking at the page content. For every such page, there is a discussion page (e.g., Wikidata (Q2013) talk page<sup>9</sup>), where contributors discuss the ways to improve the associated page.

It is difficult to remember all the identifiers of properties, items, etc., especially for usage by humans. Properties, items, and entity schemas need to be translated, i.e., these pages must have labels, descriptions, and aliases in any of the supported Wikidata languages. A property, item, and entity schema can have one label, one description, and multiple aliases in any supported language. Such translations are useful for finding the relevant items through the

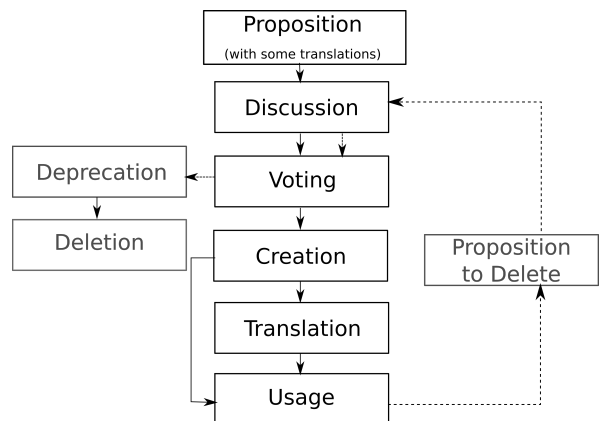
search interface. Thus contributors can find the relevant identifier by searching their local names.

Properties have datatypes. For example, to specify the official website of OpenSym, we may need a property with the datatype URL. Property P856 (official website) is one such example belonging to the datatype URL. There are several other datatypes like external identifiers, media, geographical coordinates, etc. Properties are not just limited to Wikidata, there are properties in other external data stores like DBPedia, VIAF, OCLC, ISBN, etc. Properties could be matched to properties of other external data stores through a special property called equivalent property<sup>10</sup>. Items can also be linked to entities in other external data stores through the properties belonging to the datatype: external identifier.

### 2.1 Property proposal and creation

Unlike Wikidata items, which require that any item created to follow the notability guidelines<sup>11</sup>, the creation of property items is a much longer process. Figure 1 shows the process of Wikidata property creation and its possible deletion.

A contributor looking to describe an item may not find a relevant property to describe an item. In this case, they may propose a new property on the property proposal page<sup>12</sup>. A selected number of topics have been created for allowing the contributors to discuss and find contributors belonging to a particular domain. Some examples include authority control, person, place, transportation, etc.



**Figure 1: Property proposition, creation and possible deletion.** Dark boxes and arrows shows the steps related to property creation, translation and usage. Light gray boxes and dotted arrows shows the steps related to property deletion. Some steps are shared by property creation and deletion, like discussion and voting.

An example property discussion is highlighted in Figure 2. This is the property proposal discussion of P3966 (programming paradigm)<sup>13</sup>. Figure 2a shows the information proposed by the property

<sup>1</sup><https://www.wikidata.org/wiki/Q52260246>

<sup>2</sup><https://www.wikidata.org/wiki/Q56259215>

<sup>3</sup><https://www.wikidata.org/wiki/Q2013>

<sup>4</sup><https://www.wikidata.org/wiki/Property:P31>

<sup>5</sup><https://www.wikidata.org/wiki/Property:P17>

<sup>6</sup><https://www.wikidata.org/wiki/Property:P856>

<sup>7</sup><https://www.wikidata.org/wiki/EntitySchema:E2>

<sup>8</sup><https://www.wikidata.org/wiki/Lexeme:L315>

<sup>9</sup><https://www.wikidata.org/wiki/Talk:Q2013>

<sup>10</sup><https://www.wikidata.org/wiki/Property:P1628>

<sup>11</sup><https://www.wikidata.org/wiki/Wikidata:Notability>

<sup>12</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal](https://www.wikidata.org/wiki/Wikidata:Property_proposal)

<sup>13</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/programming\\_paradigm](https://www.wikidata.org/wiki/Wikidata:Property_proposal/programming_paradigm)

proposer. Figure 2b shows the voting and in Figure 2b, we see the source code of the property proposition.

Readers may refer to the French<sup>14</sup> or the Spanish<sup>15</sup> rendering of this property discussion page. We see that some of the information like the French or the Spanish translations are available for some information like Support, Oppose, Comment. These template pages support, oppose, comment have already been translated in these languages.

As can be seen, the property proposer (or/and some other contributors) translated the property label and description in some languages. For this purpose, TranslateThis template has been used. At the time of property creation, these labels and descriptions made available, as seen in Figure 1. Properties may be translated into other languages after creation. It should be stressed that the labels, descriptions, and aliases may be modified during the lifespan of a property. Some of these modifications may be possible vandalism and require to be detected[16]. Finally, the lifespan of a property may be limited. Sometimes, contributors may propose that a property should be deleted<sup>16</sup>. This may be because of the need to change the datatype or because a new and better property has been proposed with which the concerned property can be merged.

## 2.2 Property curation: WikiProjects and Property Classes

Another major problem is the difficulty in finding relevant property items for describing an item. Take for example, the following three properties:

- country (P17) used for specifying the country of the item.
- country of origin (P495) used for specifying the country of origin of the item, when the item is a creative work, food, etc.
- country of citizenship (P27) for specifying the nationality of a person (item).

A new contributor may not be aware of all such properties, and often may end up using P17 (country) for specifying the country of origin of a creative work. For helping the contributors, there are Wikidata WikiProjects<sup>17</sup> that curate the properties relevant to a particular domain. For example, there are WikiProjects related to books<sup>18</sup>, museums<sup>19</sup>, etc. Another possible way is to curate them using the property: properties for this type (P1963), that can be used to find the properties for a given type (a class etc.)

## 2.3 Property translation and evolution

As seen in Figure 1, Wikidata properties constantly evolve. Properties are created, modified, translated and even deleted. New languages, datatypes, and WikiProjects are also created from time to time. To improve the language coverage and coverage of topics from across the world, the translation of properties cannot be overlooked. Property translation and evolution is the focus of this article.

## 3 STATE OF THE ART

Wikidata provides a SPARQL endpoint and Mediawiki API for accessing the structured data as well as the commit history. A number of tools have been created during the past few years<sup>20</sup>, which help contributors to describe, analyse and visualize Wikidata items. Tools like PropBrowser<sup>21</sup>, SQID property browser<sup>22</sup> focus on Wikidata properties. Some commonly used tools on Wikidata by contributors for generating lists are Listeria<sup>23</sup> and InteGraality<sup>24</sup>. This has been used by certain contributors to maintain a list of datatypes<sup>25</sup>. However, these tools generate the lists in a periodic manner (like weekly updates etc.)

Collaborative [11] and multilingual aspects [6] of Wikidata has been the focus of many recent works. Many of these works make use of periodic Wikidata datadumps, focusing on the global coarser analysis. But the contributors focus on some properties or a subset of curated properties for their works. Certain visualizations like the visualization of deletion discussions considered by Notabilia [15] or information flow across different language Wikipedias [1, 4] give interesting insights. While analysing and visualizing the translations of Wikidata, one may observe the multilingual nature of the contributors and also the role played by the bots[5]. Some recent work have focuses on the stability of property labels[16]. It must be noted that the property labels are meant for human consumption and undesired or possible vandalism on the properties may produce unexpected semantics of information. Hence it is important to obtain real-time and fine-granular analysis of Wikidata properties.

Translation of properties may aid to extracting structured information from Wikipedia articles to populate infoboxes [9] and Wikidata<sup>26 27</sup>. Finally, it is equally important to point the role played by Wikidata WikiProjects [7, 8] in curation of properties. The focus of WDProp proposed in this article is on Wikidata properties and the multilingual aspects around the properties. The goal is to navigate and explore properties by languages, datatypes, WikiProjects, etc.

## 4 WDPROP: FROM COARSER ANALYSIS TO FINE-TUNED ANALYSIS

One approach for analysis of Wikidata items and properties is to download Wikidata datadump<sup>28</sup> with the commit history for the study. Another approach is to make use of SPARQL endpoints and Wikidata Mediawiki API, especially when the focus is on a small subset of information. *Multilingual Wikidata Property Translation Flow Dataset* [2] was created using the latter approach. Datadumps, complete or selective are quite useful for coarser analysis and may provide some useful insights [6, 11].

Table 1 shows the details of *Multilingual Wikidata Property Translation Flow Dataset*. This dataset contains the translation flow of

<sup>14</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/programming\\_paradigm?uselang=fr](https://www.wikidata.org/wiki/Wikidata:Property_proposal/programming_paradigm?uselang=fr)

<sup>15</sup>[https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/programming\\_paradigm?uselang=es](https://www.wikidata.org/wiki/Wikidata:Property_proposal/programming_paradigm?uselang=es)

<sup>16</sup>[https://www.wikidata.org/wiki/Wikidata:Properties\\_for\\_deletion](https://www.wikidata.org/wiki/Wikidata:Properties_for_deletion)

<sup>17</sup><https://www.wikidata.org/wiki/Wikidata:WikiProject>

<sup>18</sup>[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Books](https://www.wikidata.org/wiki/Wikidata:WikiProject_Books)

<sup>19</sup>[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Museums](https://www.wikidata.org/wiki/Wikidata:WikiProject_Museums)

<sup>20</sup><https://hay.toolforge.org/directory/>

<sup>21</sup><http://tools.wmflabs.org/hay/propbrowse>

<sup>22</sup><https://sqid.toolforge.org/#/browse?type=properties>

<sup>23</sup><https://www.wikidata.org/wiki/Wikidata:Listeria>

<sup>24</sup><https://integrality.toolforge.org/>

<sup>25</sup>[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Properties/Reports/Datatypes](https://www.wikidata.org/wiki/Wikidata:WikiProject_Properties/Reports/Datatypes)

<sup>26</sup><https://pltools.toolforge.org/harvesttemplates/>

<sup>27</sup><https://petscan.wmflabs.org/>

<sup>28</sup>[https://www.wikidata.org/wiki/Wikidata:Database\\_download](https://www.wikidata.org/wiki/Wikidata:Database_download)



**(b) Property discussion: Votes**

```

en = {{TranslateThis | anchor = en
| en = programming paradigm
| pt-br = paradigma de programação
| ru = парадигма программирования
| sr = програмска парадигма
| uk = парадигма програмування
<!-- [[xx = property names in some other languages -->
}} ==
{{Property proposal
|status          = 3966
|description     = {{TranslateThis
| en = the programming paradigm in which a programming language is classified
| pt-br = um paradigma de programação no qual uma linguagem de programação está classificada
| ru = парадигма программирования, в соответствии с которой классифицируется язык программирования
| sr = програмска парадигма у којој је програмски језик класификован
| uk = парадигма програмування до якої належить мова програмування
<!-- [[xx = property names in some other languages -->
}}
|subject item    = Q188267
|infobox parameter = "paradigm" in [[en:Template:Infobox programming language]]
|datatype        = item
|domain          = {{Q|9143}}
|allowed values  = <!-- type of linked items (Q template or text), list or range of allowed values, string
|source          = <!-- external reference, Wikipedia list article, etc. -->
|example         =
= {{Q|15777}} = {{Q|275596}}. {{Q|223335}}

```

### (c) Property discussion template

Figure 2: Property proposition and discussion of P3966 (programming paradigm)

labels, descriptions, and aliases of Wikidata properties collected on July 7, 2019. The dataset consists of four columns: timestamp of the translation, the identifier of the property translated, the language of translation, and the type of translation (label, description, or alias). The 6347 properties considered in the study showed that a mean of 21.44 property labels in different languages was available. The minimum and the maximum number of property labels are 1 and 154 respectively. The distribution of the count of properties and the number of languages is shown in Figure 3.

Measure	Value
count	6347
mean	21.444935
std	20.897675
min	1.000000
max	154.000000

Table 1: Details of *Multilingual Wikidata Property Translation Flow Dataset* [2]. The dataset contains the details of every type of translation (labels, description and aliases) on 6347 properties.

Further analysis of these property labels was made on the count of combined occurrences of languages on different properties and

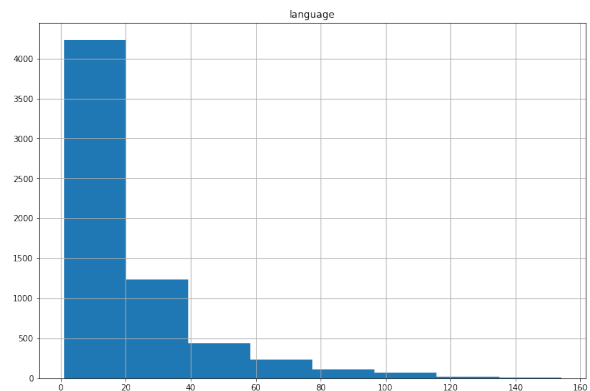


Figure 3: Distribution of labels in different languages. The X-axis corresponds to the number of languages and the Y-axis corresponds to the number of properties. Very less number of properties have label translations in more than 100 languages.

the results of the top seven combinations are shown in Table 2.

Count	Language Combinations	Count	Language Combinations	Count	Language Combinations
6347	en	6259	ar, uk	6257	en, nl, ar, uk
6338	ar	5923	fr, ar	5919	en, nl, fr, ar
6306	nl	5905	fr, uk	5903	en, nl, fr, uk
6263	uk	4819	ar, ca	5901	en, ar, fr, uk
5928	fr	4795	ca, uk	5899	nl, fr, uk, ar
4823	ca	4530	fr, ca	4818	en, nl, ar, ca
4019	de	4015	ar, de	4794	en, nl, ca, uk

**Table 2: Total number of properties with the label translations in the language combinations (Top 7)**

For example, considering a combination of four languages, English (en), Dutch (nl), Arabic (ar) and Ukrainian (uk) property labels are available on 6257 properties.

The above dataset can also be used to study the property translation flow for a given property as shown in Figure 4. It took a period between 2013 and 2019 for the translation of the labels (violet dots) of the property P17 (country) to be available in multiple languages. The first descriptions (light blue dots) were subsequently translated and the first aliases (red dots) were added after a long time. Check the distances between the violet, blue and red dots for some languages. By focusing on modifications (i.e, changes made after the first translations), a modified form of this visualization can be used to detect possible vandalisms.

However, the above analysis is limited to a snapshot of data at a given time. Considering the rapidly evolving nature of Wikidata, there is a need to have some real-time view of Wikidata properties. WDProp was developed to understand and obtain an integrated view on the different aspects of properties, their creation, and translation. The focus here is on obtaining real-time statistics, on a much fine-granular basis, i.e., instead of focusing on the global analysis of properties (as seen above), the users can focus on a single property or a combination of curated properties. WDProp provides a visual interface to such information.

Property translations can be analyzed in different ways. There are two major translations: property labels and property descriptions. Properties may also have some aliases in certain languages. Whether all the aliases in one language need to be available in other languages is an open question. For a given language, the language community may wish to ensure the translation of all the property labels and descriptions. For some generic properties like country (P17), the goal of the Wikidata community members is to ensure the translation in all the supported languages.

WDProp separates property translations in the following manner:

- translated and untranslated property labels
- translated and untranslated property descriptions
- translated and untranslated property aliases

A Wikidata contributor may find one or more of the above pieces of information relevant, some of whom may wish to obtain the latest translation statistics, and some others looking for properties not yet translated.

## 4.1 Development

WDProp was developed using basic web technologies like HTML, Javascript, and CSS. It makes use of the SPARQL queries and Wikidata MediaWiki API to obtain the latest information from the Wikidata servers. It, therefore, does not require any installation since it contains a collection of HTML, Javascript, and CSS files. This simple approach ensures that it can be tested on any modern browser, including desktop and mobile devices. It is also available on Toolforge<sup>29</sup>, for users wishing to use links that can be shared with others.

## 4.2 Results

WDProp can be used to analyze the collaborative approach to the development of multilingual ontology [12] and to obtain the real-time information related to multilingual aspects of Wikidata properties, especially their translation [13]. It provides the following information and features:

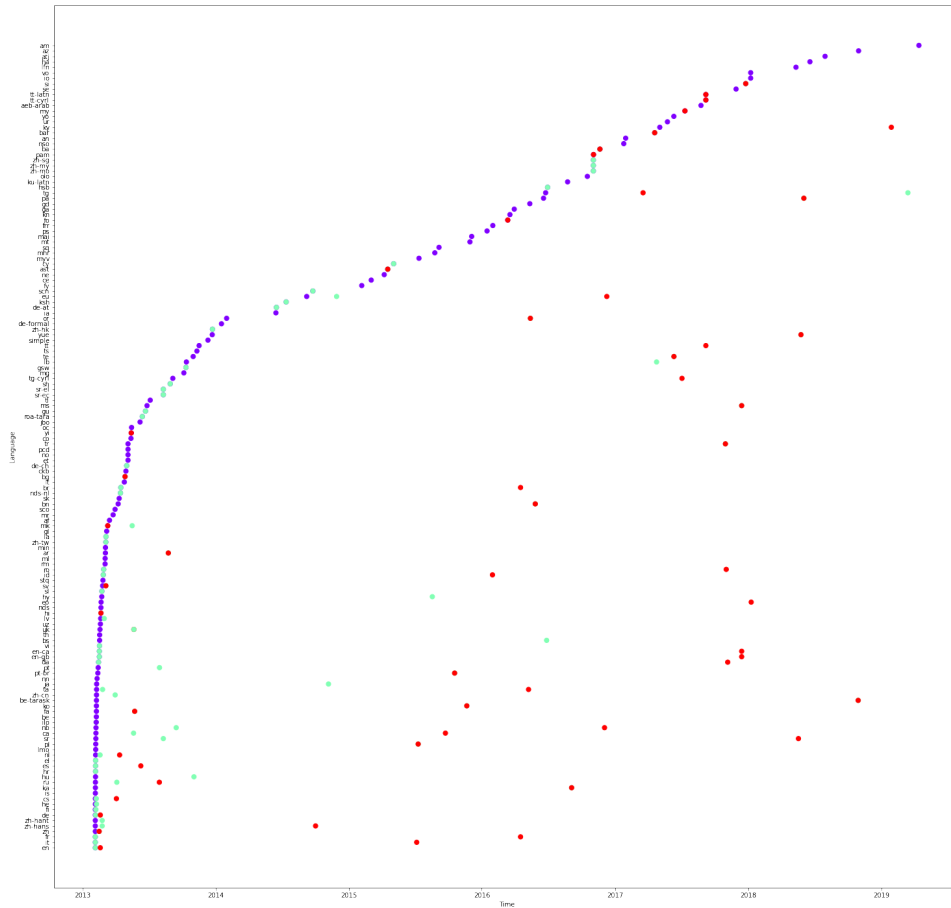
- **Bookmarkable links:** This feature ensures that the links are bookmarkable and use Wikidata supported codes and identifiers for obtaining the information related to languages, properties, datatypes, etc.
- **List of supported languages:** see Figure 5
- **Translation statistics of labels, descriptions and aliases of Wikidata properties:** see Figure 6
- **List of properties:** see Figure 8
- **Compare translation statistics among different languages:** see Figure 7
- **Use of references and equivalent properties:** see Figure 9.
- **Translation statistics in a given language:** see Figure 10 and Figure 11.
- **Translation statistics of property discussion templates:** see Figure 12
- **Navigation of properties by datatypes:** see Figure 13
- **Navigation of properties by classes:** see Figure 14
- **Navigation of properties by Wikidata WikiProjects:** see Figure 17
- **Visualize path of translation:** see Figure 18

In this article, the URLs used for the above screenshots are given as footnotes.

Figure 5 shows the list of supported languages<sup>30</sup> on Wikidata. New languages are incubated on Wikidata. Hence the number of

<sup>29</sup><https://wdprop.toolforge.org/>

<sup>30</sup><https://wdprop.toolforge.org/languages.html>



**Figure 4: Timeline of translation of labels (violet dots), descriptions (blue dots) and aliases (red dots) of property P17 (country). The Y-axis shows the languages and the X-axis shows the period between the year 2013 and 2019.**

supported languages on Wikidata has evolved over the years. This list is useful to know the number of supported languages at any given time.

Property translation statistics<sup>31</sup> is given in Figure 6. It shows the number of labels, descriptions, and aliases in each of the supporting

languages seen above. It is also possible to obtain the number of labels, descriptions, and aliases missing translation<sup>32</sup>.

Comparison of property translation statistics is given in Figure 7. It shows the property translation statistics English(en), French(fr) and Spanish(es)<sup>33</sup>. The users can choose a selection of languages

<sup>31</sup><https://wdprop.toolforge.org/translated.html>

<sup>32</sup><https://wdprop.toolforge.org/untranslated.html>

<sup>33</sup><https://wdprop.toolforge.org/compare.html?languages=en,%20fr,%20es>

## Compare translations among different languages

Given below is the list of supported languages on Wikidata. Click on any language code below to see the current translation statistics of properties

Total 319 languages

[illegible]

**Figure 5: Wikidata supported languages**

en, es, fr

[Submit Query](#)

URL: [compare.html?languages=en.es.fr](http://compare.html?languages=en.es.fr)

### Count of translated labels

### Count of translated labels

en (8905)   fr (8378)   es (4738)

### Count of translated descriptions

en (8819) fr (6806) es (3127)

### Count of available aliases

en (13024) fr (8577) es (6248)

**Figure 7: Comparison of property translation statistics on WDPop: English(en), French (fr) and Spanish (es)**

As discussed above, new properties are regularly proposed and voted by the community members. The properties that have been created have an associated identifier. In Figure 8, 100 properties are shown including the option to view the deleted properties<sup>34</sup>. It is interesting to observe here that many of the initial properties have been deleted.

Some properties have equivalent properties in other knowledge bases. Take, for example, Property (P17) is also available on DBpedia<sup>35</sup>. Wikidata tracks such equivalent properties, which can be considered as a quality metric for property relevance. Additionally, like Wikidata items, Wikidata properties may also have references to the different statements. However, not all properties have references. In Figure 9, the usage of equivalent properties and references<sup>36</sup> are shown.

It is also possible to obtain the translation statistics for a given language. In Figure 10, current information of property translations in the Afar (aa) language<sup>37</sup> can be seen.

In Figure 11, current information of property translations in English (en) language<sup>38</sup> can be seen. When compared to Figure 10, we observe the differences in property translation in the two languages. Similar observations can be made between the English language and the languages with few speakers (or few contributors on Wikidata).

Figure 12 shows the current translation statistics of the key templates used for property discussions<sup>39</sup>. The numbers are lower than the total number of supported languages on Wikidata, for all four templates.

Wikidata supports several datatypes: Math, URLs, external identifiers, etc. New datatypes like `wikibase:WikibaseLexeme` and

**Figure 6: WDPProp Translation Statistics of Properties:  
Count of available translations**

and compare the translation statistics of labels, descriptions and aliases.

<sup>34</sup><https://wdprop.toolforge.org/properties.html>

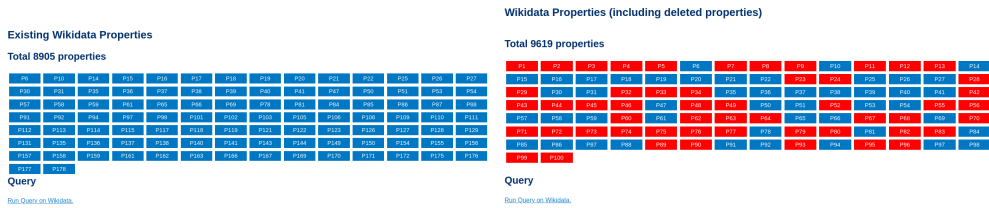
<sup>35</sup><http://dbpedia.org/ontology/country>

<sup>36</sup><https://wdprop.toolforge.org/provenance.html>

<sup>37</sup><https://wdprop.toolforge.org/language.html?language=aa>

<sup>38</sup><https://wdprop.toolforge.org/language.html?language=en>

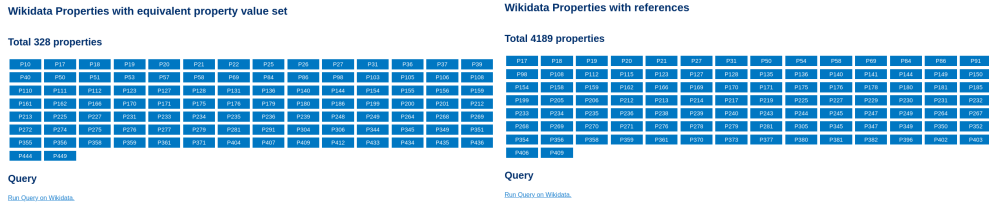
<sup>39</sup><https://wdprop.toolforge.org/templates/translated.html>



(a) Properties

(b) Properties including deleted properties

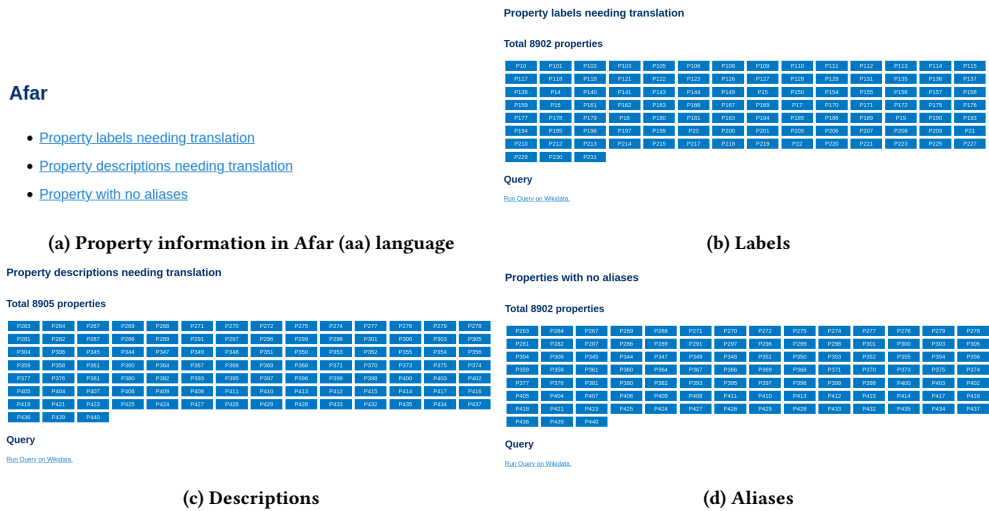
Figure 8: List of properties on WDPop: limited to view of 100 properties



(a) Properties with equivalent properties

(b) Properties with references

Figure 9: Equivalent properties and usage of references on Wikidata properties



(c) Descriptions

(d) Aliases

Figure 10: Property information in the Afar (aa) language

wikibase:WikibaseSense were recently created for representing lexicographical data. Figure 13 shows the list of available datatypes<sup>40</sup> and the properties belonging to datatype: wikibase:Math<sup>41</sup>.

Some contributors have created property classes to help to describe Wikidata items belonging to different domains, especially for new contributors who often find it difficult to find the relevant

properties to describe a given entity. Figure 14 shows a snapshot of the available property classes<sup>42</sup> and highlights lighthouse (Q39715).

Figure 15 shows the details of two such curated Wikidata items: lighthouse (Q39715)<sup>43</sup> and Wikidata property related to lighthouses (Q28739677)<sup>44</sup> and shows a screenshot of some of the properties proposed by them.

<sup>40</sup><https://wdprop.toolforge.org/datatypes.html>

<sup>41</sup><https://wdprop.toolforge.org/datatype.html?datatype=wikibase:Math>

<sup>42</sup><https://wdprop.toolforge.org/classes.html>

<sup>43</sup><https://wdprop.toolforge.org/class.html?class=Q39715>

<sup>44</sup><https://wdprop.toolforge.org/class.html?class=Q28739677>





Figure 11: Property information in the English (en) language

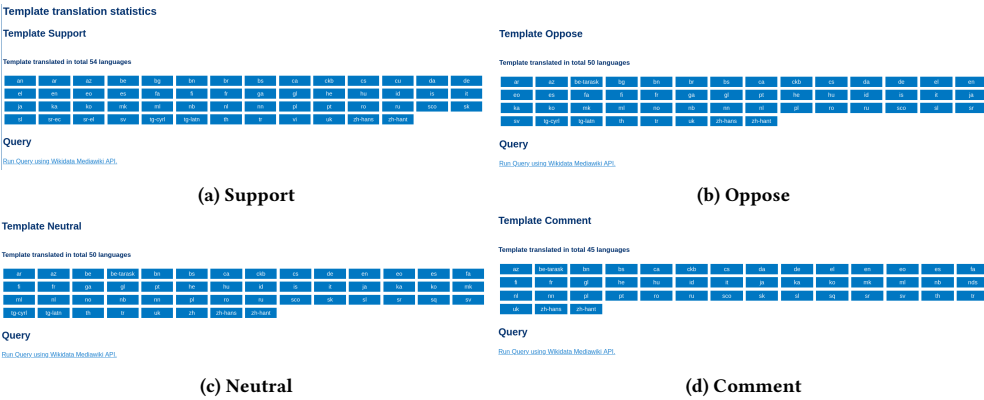


Figure 12: Translation Statistics of templates used for property discussion: Support, Oppose, Neutral and Comment

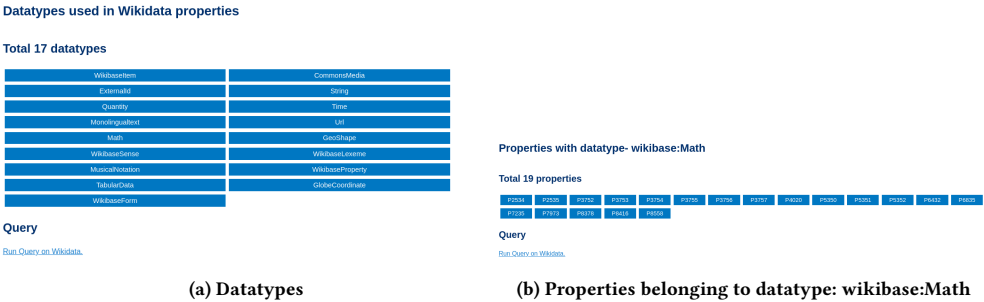


Figure 13: Datatypes

**Classes**

Total 1555 classes

Item	Class label
Q1281401	Q1281401
Q10262930	Q10262930
Q11430407	Q11430407
Q56903505	Q56903505
Q56243029	Q56243029
Q67202678	Q67202678
Q100216147	Q100216147
Q101112804	Q101112804
Q104457865	Q104457865
Q105834419	Q105834419
Q106435073	Q106435073
Q106518893	Q106518893
Q106772873	Q106772873
Q1261115	7" single

(a) Property classes

Q2034719	legal case
Q269371	legislation
Q269662	legislative act
Q269722	lens mount
Q334692	letter
Q344669	lexical item
Q1739958	light characteristic
Q11421350	light novel series
Q1142078	light source
Q39715	<b>lighthouse</b>
Q2690110	liquid capacity
Q1213063	list
Q1060380	list of Pokémon species
Q395034	literary character
Q7725634	literary work
Q395089	lithostratigraphic unit
Q209039	live album

(b) Highlighted property class: lighthouse (Q39715)

Figure 14: Property classes: curation of properties for a particular class

**Properties for the Class: lighthouse**

Total 15 properties

Property	Property label
P5025	coordinate location
P17	country
P3923	focal height
P2048	height
P1436	heritage designation
P18	image
P571	inception
P1306	light characteristic of lighthouse
P3922	light sector
P2929	lighthouse range
P206	located in or next to body of water
P1431	located in the administrative territorial entity
P706	located on terrain feature
P1427	owned by
P3994	racon signal

(a) Properties curated on property class: lighthouse (Q39715)

**Properties for the Class: Wikidata property related to lighthouses**

Total 18 properties

Property	Property label
P2386	ARLHS lighthouse ID
P2967	Admiralty number
P2929	Canadian Coastguard Lighthouse ID
P2929	Danish List of Lights and Fog signals ID
P5149	Finnish List of Lights ID
P5011	Heritage Lighthouse of Canada ID
P3985	Italian Navy Lighthouses and Beacons ID
P3961	MarineTraffic Lighthouse ID
P2968	NGA lighthouse ID
P2929	Norwegian List of Lights ID
P2929	Online List of Lights ID
P5522	Portuguese lighthouse ID
P3723	USCG lighthouse ID
P5593	Vitaskrå ID
P2929	focal height

(b) Properties curated on property class: Wikidata property related to lighthouses (Q28739677)

Figure 15: Properties curated on two property classes

The curators may be interested to know the translation statistics of all the properties considered in a given class. Figure 16 shows the statistics of property translations of the two property classes.

The translation statistics of the properties curated and used on the WikiProjects can also be seen. Take, for example, Figure 17 shows the translation statistics of labels of the 37 properties used on WikiProject Museums<sup>45</sup>.

In Figure 18, the translation path of a Wikidata property, P856 (official website)<sup>46</sup> is shown. This is an extension to the work previously done in [13], where a tabular column was used to highlight the translation of property labels, descriptions, and aliases. On clicking any property page (e.g., P17<sup>47</sup>), it is possible to obtain the current translation statistics in different languages, links to the provenance information (number of references on statements describing the property, number of equivalent properties, usage of references in the statements using the property), translation path, and visualization of the translation path.

### 4.3 Discussion

Wikidata is highly evolving, and new properties are being regularly proposed and translated. The WDProp interface has also undergone several changes during the last couple of years to incorporate these

changes. Take, for example, WDProp initially displayed all the properties on the properties page [14] (less than 5000 properties at that time). But this number is currently around 9,000 properties. So now, the interface shows only the first 100 properties obtained from the SPARQL query. However, the user now has the SPARQL query link to obtain the complete list. The SPARQL query links and links to Wikidata Mediawiki API calls have now been integrated into all the pages on WDProp. This feature gives the researchers and practitioners the possibility to not only verify the visualization results but also modify these SPARQL/API queries based on their specific requirements.

Some commonly used properties also face vandalism. The commit history of some initial properties has therefore become long, which means obtaining the complete translation path as shown in Figure 18 takes a significant amount of time. Also, the number of property classes and WikiProjects is increasing, which means the current display of these lists may require some modifications in the future, especially the possibility to search and filter the desired information.

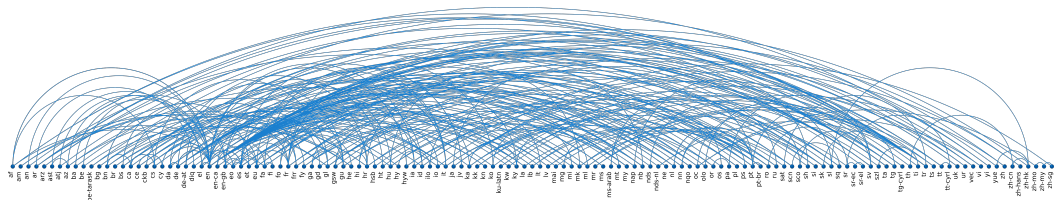
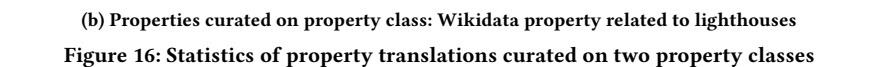
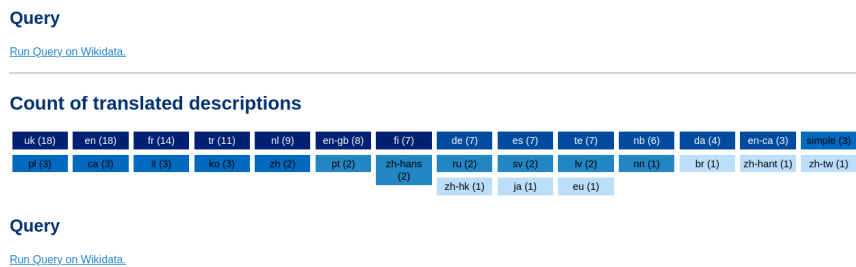
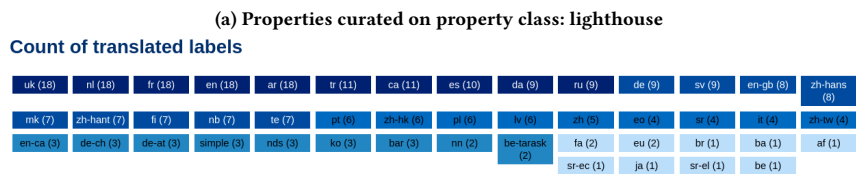
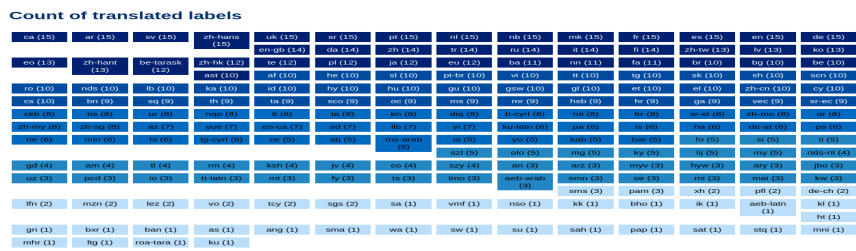
## 5 CONCLUSION AND FUTURE WORKS

With the growing usage of mobile devices and internet penetration across the world, web developers and content producers are looking for ways to ensure the availability of information in the local languages of the readers. Considering the number of languages, it is

<sup>45</sup><https://wdprop.toolforge.org/wikiproject.html?project=Wikidata:WikiProject%20Museums>

<sup>46</sup><https://wdprop.toolforge.org/pathviz.html?property=P856>

<sup>47</sup><https://wdprop.toolforge.org/property.html?property=P17>



properties and their translations and discussed the design and development of WDProp for this study. It also demonstrated how the web application can be used to perform granular analysis on Wikidata properties and their curation.

The future course of action includes options to download the data as CSV or JSON files and a multilingual user interface. Another possible direction is to explore how the property translation flow can be used to suggest the untranslated properties to the contributors of a given language. Some aspects of this work can also be integrated into Wikibase installations using SPARQL endpoints and Mediawiki API, but this requires additional experiments.

## ACKNOWLEDGMENTS

The author would like to thank the Wikidata and Wikimedia community members for their feedback during Wikimedia Hackathon 2018, WikidataCon 2019, and ContribuLing 2021 for the feedback, thanks to which the web application has been improved and has helped to provide a written version of this article.

## REFERENCES

- [1] Erik Borra, David Laniado, Esther Weltevrede, Michele Mauri, Giovanni Magni, Tommaso Venturini, Paolo Ciuccarelli, Richard Rogers, and Andreas Kaltenbrunner. 2015. A Platform for Visually Exploring the Development of Wikipedia Articles. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*. AAAI Press, 711–712. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10565>
- [2] Thibaut Chamard and John Samuel. 2019. *Multilingual Wikidata Property Translation Flow Dataset*. <https://doi.org/10.5281/zenodo.3271358>
- [3] Lars Christoph Gleim, Rafael Schimassek, Dominik Hüser, Maximilian Peters, Christoph Krämer, Michael Cochez, and Stefan Decker. 2020. SchemaTree: Maximum-Likelihood Property Recommendation for Wikidata. In *The Semantic Web - 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31-June 4, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12123)*, Andreas Harth, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Heiko Paulheim, Anisa Rula, Anna Lisa Gentile, Peter Haase, and Michael Cochez (Eds.). Springer, 179–195. [https://doi.org/10.1007/978-3-030-49461-2\\_11](https://doi.org/10.1007/978-3-030-49461-2_11)
- [4] Simon Gottschalk and Elena Demidova. 2017. MultiWiki: Interlingual Text Passage Alignment in Wikipedia. *TWEB* 11, 1 (2017), 6:1–6:30. <http://doi.acm.org/10.1145/3004296>
- [5] Lucie-Aimée Kaffee, Kemele M. Endris, and Elena Simperl. 2019. When humans and machines collaborate: cross-lingual label editing in wikidata. In *Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019, Skövde, Sweden, August 20-22, 2019*, Björn Lundell, Jonas Gamalielsson, Lorraine Morgan, and Gregorio Robles (Eds.). ACM, 16:1–16:9. <https://doi.org/10.1145/3306446.3340826>
- [6] Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. 2017. A Glimpse into Babel: An Analysis of Multilinguality in Wikidata. In *Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17)*. Association for Computing Machinery, Galway, Ireland, 1–5. <https://doi.org/10.1145/3125433.3125465>
- [7] Timothy Kanke. 2018. Preliminary Exploration of Knowledge Curation Activities in Wikidata WikiProjects. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. ACM, Fort Worth Texas USA, 349–350. <https://doi.org/10.1145/3197026.3203878>
- [8] Timothy Kanke. 2020. Knowledge curation work in Wikidata WikiProject discussions. *Library Hi Tech* 39, 1 (Jan. 2020), 64–79. <https://doi.org/10.1108/LHT-04-2019-0087>
- [9] Dustin Lange, Christoph Böhm, and Felix Naumann. 2010. Extracting structured information from Wikipedia articles to populate infoboxes. In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*. ACM Press, Toronto, ON, Canada, 1661. <https://doi.org/10.1145/1871437.1871698>
- [10] Michael Luggen, Julien Audiffren, Djelle Eddine Difallah, and Philippe Cudré-Mauroux. 2021. Wiki2Prop: A Multimodal Approach for Predicting Wikidata Properties from Wikipedia. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 2357–2366. <https://doi.org/10.1145/3442381.3450082>
- [11] Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. 2015. Peer-production system or collaborative ontology engineering effort: what is Wikidata?. In *Proceedings of the 11th International Symposium on Open Collaboration (OpenSym '15)*. Association for Computing Machinery, San Francisco, California, 1–10. <https://doi.org/10.1145/2788993.2789836>
- [12] John Samuel. 2017. Collaborative Approach to Developing a Multilingual Ontology: A Case Study of Wikidata. In *Metadata and Semantic Research (Communications in Computer and Information Science)*, Emmanouel Garoufallou, Sirje Virkus, Rania Siatra, and Damiana Koutsomihia (Eds.). Springer International Publishing, Cham, 167–172. [https://doi.org/10.1007/978-3-319-70863-8\\_16](https://doi.org/10.1007/978-3-319-70863-8_16)
- [13] John Samuel. 2018. Analyzing and Visualizing Translation Patterns of Wikidata Properties. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction (Lecture Notes in Computer Science)*, Patrice Bellot, Chiraz Trabelsi, Josiane Mothe, Fionn Murtagh, Jian Yun Nie, Laure Soulier, Eric SanJuan, Linda Cappellato, and Nicola Ferro (Eds.). Springer International Publishing, Cham, 128–134. [https://doi.org/10.1007/978-3-319-98932-7\\_12](https://doi.org/10.1007/978-3-319-98932-7_12)
- [14] John Samuel. 2018. Towards Understanding and Improving Multilingual Collaborative Ontology Development in Wikidata. In *Wikidata Workshop 2018*. [http://wikiworkshop.org/2018/papers/wikiworkshop2018\\_paper\\_12.pdf](http://wikiworkshop.org/2018/papers/wikiworkshop2018_paper_12.pdf)
- [15] Moritz Stefaner, Dario Taraborelli, and Giovanni L Ciampaglia. 2011. Notabilia—Visualizing Deletion Discussions on Wikipedia. <http://well-formed-data.net/archives/570/notabilia-visualizing-deletion-discussions-on-wikipedia>
- [16] Thomas Pellissier Tanon and Lucie-Aimée Kaffee. 2018. Property Label Stability in Wikidata: Evolution and Convergence of Schemas in Collaborative Knowledge Bases. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon, France, April 23-27, 2018*. ACM, 1801–1803. <http://doi.acm.org/10.1145/3184558.3191643>
- [17] Denny Vrandečić. 2012. Wikidata: a new platform for collaborative data collection. In *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*. ACM Press, Lyon, France, 1063. <https://doi.org/10.1145/2187980.2188242>
- [18] Eva Zangerle, Wolfgang Gassler, Martin Pichl, Stefan Steinhauser, and Günther Specht. 2016. An Empirical Evaluation of Property Recommender Systems for Wikidata and Collaborative Knowledge Bases. In *Proceedings of the 12th International Symposium on Open Collaboration (OpenSym '16)*. Association for Computing Machinery, Berlin, Germany, 1–8. <https://doi.org/10.1145/2957792.2957804>