Suggesting Citations for Wikidata Claims based on Wikipedia's External References

Paolo Curotto and Aidan Hogan

DCC, Universidad de Chile & IMFD
{pcurotto,ahogan}@dcc.uchile.cl

Abstract. Given a Wikidata claim, we explore automated methods for locating references that support that claim. Our goal is to assist human editors in referencing claims, and thus increase the ratio of referenced claims in Wikidata. As an initial approach, we mine links from the references section of English Wikipedia articles, download and index their content, and use standard relevance-based measures to find supporting documents. We consider various forms of search phrasings, as well as different scopes of search. We evaluate our methods in terms of the coverage of reference documents collected from Wikipedia. We also develop a gold standard of sample items for evaluating the relevance of suggestions. Our results in general reveal that the coverage of Wikipedia reference documents for claims is quite low, but where a reference document is available, we can often suggest it within the first few results.

Keywords: Wikidata, Wikipedia, citations, references

1 Introduction

Wikidata [15] is a collaboratively-edited knowledge graph. Much like its sibling project Wikipedia, Wikidata is continuously extended and curated by a large community of volunteers. Unlike Wikipedia, Wikidata manages structured statements about items. Items include people, places, proteins, papers, printers, planets, political parties, and many more besides. A statement consists of an item, a property, and a value. For example, a statement might claim that the album Pulse (*item*) has the performer (*property*) Pink Floyd (*value*). Values may be items, datatypes (numbers, booleans, dates, times, etc.), or special terms indicating an unknown value or that no such value exists. Statements can also have qualifiers that scope the validity of the claim, or provide additional details; this may state, for example, a time period in which a claim was true, the previous or next item with that value for that property; etc.

   As per Wikipedia, Wikidata does not aim to be a primary source of knowledge, but rather a secondary source of knowledge: statements in Wikidata should

be interpreted as claims held true according to a specific, external and authoritative source.[1] Thus it is important that statements be independently verifiable, meaning that parties other than the editor that added the statement should be able to verify the validity of that statement. Some statements are considered self-verifiable. The cases listed by Wikidata editor guides include:[2]

- *Common human knowledge*: statements that are obvious to most and can be considered self-evident; for example, that Paris is an instance of city, that Paris is the capital of France, that city is a subclass of urban area, etc.
- *The value is an external source*: statements that point to an external source, such as identifiers associated with the item in external catalogues.
- *The value refers to an external source*: statements that point to an Wikidata item that itself can verify the statement, such as an album stating its artist, a book stating its author, etc.

In the case of statements not falling into one of these three categories, the onus is on the editor that adds (modifies or restores) a statement to establish verifiability by adding a reference for the statement based on an *authoritative source*. Authoritative sources include books, publications, news media, laws, other popular media, reputable websites, etc. Questionable sources, sponsored sources, self-published sources, etc., may be rejected as non-authoritative sources.

At the time of writing (August 2020), Wikidata describes 1.124 billion statements about 88 million items and has over 23 thousand active users. Of these statements, 771 million are referenced to external sources (68.56%), 68 million are referenced to Wikipedia (6.02%), leaving 286 million without reference (25.42%).[3] Of these items, 71 million (80.34%) have at least one referenced statement. While collecting 771 million referenced statements is an impressive achievement, more can be done to improve the coverage of references [5]. Valid but unreferenced statements run the risk of being removed; conversely leaving them in the knowledge graph runs the risk of hosting invalid statements, which may in turn cause adverse effects for applications that use Wikidata.[4] Furthermore, Wikidata does not currently offer its editors much assistance in finding references for a claim; a tool to automatically suggest references would help make the most of these volunteers' time and effort. Finally, the aforementioned statistics count statements with some reference, but Piscopo et al. [10] estimate that only 61% of Wikidata's references can be considered authoritative and relevant.

In summary, we see a need for research on methods to (semi-)automatically find authoritative references for statements in Wikidata. Herein we describe our work on an initial such method based on searching over reference documents scraped from Wikipedia. This approach seems initially quite natural: Wikipedia

---

[1] See https://www.wikidata.org/wiki/Wikidata:Verifiability

[2] See https://www.wikidata.org/wiki/Help:Sources/Items_not_needing_sources

[3] See https://wikidata-todo.toolforge.org/stats.php

[4] As an anecdotal example of the latter, we refer to Siri reporting the death of Stan Lee, apparently based on an invalid statement added to Wikidata: https://io9.gizmodo.com/siri-erroneously-told-people-stan-lee-was-dead-1827322243.

articles are linked with Wikidata items; Wikipedia references follow similar principles of verifiability and authority as for Wikidata; Wikipedia is an older project and thus one might expect more extensive reference lists to have developed over time; a considerable number of Wikidata statements already reference a Wikipedia article that should itself cite an authoritative source; the factual nature of Wikipedia means that one could expect overlap in terms of the claims made about the same entities/items on both sites; Wikidata guides suggest to search Wikipedia for sources; etc. Our results, however, show that Wikipedia's references are quite limited in terms of coverage for Wikidata claims.

2   Related Works

A number of works have analysed referencing in Wikipedia. In a study of the quality of Wikipedia articles, Warncke-Wang et al. [16] found that the number of references was a key signal for predicting the quality of articles as manually labelled through Wikipedia's peer review process. Lewoniewski et al. [6] analyse the differences and overlap between Wikipedia references across seven different language versions; of the languages studied, they found that over half (25.5 million) of the total (41.2 million) references came from English Wikipedia. Kousha and Thelwall [4] analyse whether or not Wikipedia citations predict the impact of academic publications, finding that few indexed articles are cited. Redi et al. [13] construct a taxonomy of reasons why claims should be cited in Wikipedia, and then develop a machine learning model to predict which claims require citation and for which reason. More recently, Piccardi et al. [9] found low user engagement with external citations in English Wikipedia, with about 1-in-300 page visits resulting in a click-through to a reference on the article.

With respect to references on Wikidata, WikiCite is a Wikimedia initiative to develop and expand the citation data available through Wikidata.[5] As part of the WikiCite initiative, Nielsen et al. [8] discuss how Wikipedia references provide limited data about the source being referred to, contrasting this with Wikidata, which contains structured data about books, articles, authors, publishers, identifiers, etc.; they provide statistics on such data, and build a scientometric application called Scholia on top of them. Piscopo et al. [10,12] have provided in-depth studies comparing external references on both Wikipedia and Wikidata, finding that there is low overlap between both in terms of the references used and the domains of those references [12]; they further estimate that 61% of Wikidata's external references are considered relevant and authoritative [10]. Lemus-Rojas and Pintscher [5] identify the "citation gap" as a problematic issue, suggesting that librarians are well-positioned to help address this gap, as they have already done for Wikipedia. Piscopo and Simperl [11] discuss the importance of references to various dimensions of Wikidata quality.

Regarding datasets, Delpeuch [3] and more recently Singh et al. [14] have published metadata for citations extracted from English Wikipedia associated

---

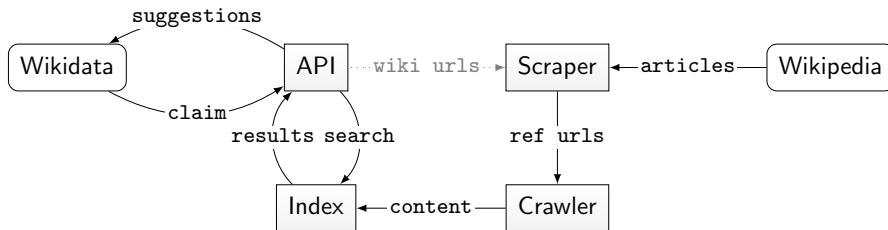[5] See `https://meta.wikimedia.org/wiki/WikiCite`

Fig. 1. Proposed architecture for suggesting references

with external identifiers (e.g., DOIs). Chou et al. [1] also recently published a dataset of English Wikipedia articles annotated with the aforementioned model of Redi et al. [13]. However these datasets do not provide the textual content of the external references, rather focusing on meta-data extracted from Wikipedia.

3   Proposed Approach & Research Questions

We propose to scrape external reference URLs from English Wikipedia, and to download and index their content. Thereafter, given a Wikidata claim for which an editor requires suggestions of potential references, using the labels and aliases of the items involved, we will convert the claim to a search using English terms and apply the search over the inverted indexes of the content of the external documents, using standard relevance-measures to prioritise documents. Finally, to assist the editor, we will return not only the document itself, but also a snippet of text from the document that contains the relevant keywords.

We present the high-level architecture in Figure 1. The API provides an interface that accepts a claim from Wikidata (along with associated metadata) and returns suggestions of potential references. In order to provide these references, a Scraper collects and parses the URLs of external references from articles on English Wikipedia. These URLs are passed to a Crawler that downloads the URLs and saves their content into an Index. The API can then formulate a search for the claim over the Index, which returns relevant documents as results that are returned as suggestions. We consider the option of both an offline and online mode. In the offline mode, the Scraper and Crawler process all of (English) Wikipedia, generating the Index over the full corpus that can be searched at runtime. We also consider an online/lazy mode, where the Scraper rather accepts a list of relevant Wikipedia article URLs from the API, which are passed to the Crawler, which in turn populates the Index at runtime before the search is performed. The offline mode has the benefit of less latency, but *a priori* it is not clear that performing such a crawl of all external references is feasible; also the Index would require periodic updates. The online mode is easier to keep up-to-date, where the Index rather acts as a cache, but is associated with slower runtime responses as the Crawler operates while the editor is waiting for suggestions.

Our initial goal is to study the feasibility of this overall approach, and to establish baseline methods and datasets for further research. Within this proposal a number of initial research questions arise:

RQ1 *Offline vs. online indexing*: Is it feasible to scrape, download and index the content of all of English Wikipedia's external references offline? Or would it be better to scrape, download and index the content online/lazily for the external references of the Wikipedia articles relevant to the claim at hand?

RQ2 *Coverage*: How many external references can we source from the article corresponding to each Wikidata item? Can we build a corpus with good coverage of Wikidata items in general?

RQ3 *Search phrasing*: How best should we phrase the search? Should we use only primary labels, or also aliases? What connectives should we use?

RQ4 *Relevance*: Are traditional IR measures sufficient to generate good suggestions? Should we search only in references for Wikipedia articles corresponding to the item(s) involved in the claim, or across the entire corpus?

RQ5 *Suggestion Quality*: How often can we generate good suggestions of references for claims? Are the rankings of suggestions suitable? Can we also suggest relevant text snippets from with the documents to support the claim?

In this initial work our goal is to gain insights regarding these research questions, rather than seeking definitive answers.

4 Scraping, Crawling & Indexing

We first explore the offline approach. We start with a dump of Wikidata, from which we extracted the mapping to Wikipedia articles. These articles were then retrieved from a 2018 HTML corpus of Wikipedia [7]. A custom scraper extracts the external reference URLs from the articles. We used Apache Nutch[6] for crawling, which uses Apache Solr[7] as an underlying index. To avoid Denial of Service (DoS) attacks, we configured Nutch to wait 5 seconds between requests to the same website. Nutch indexes the content, title host and URL of the successfully retrieved webpages in Solr; we enrich this index with the `Q` codes of the Wikidata items corresponding to the Wikipedia article of the external reference.

Results: A total of 32,329,989 raw external reference URLs were extracted from 5,461,401 articles. Removing repeated and ill-formed URLs yielded 23,036,318 well-formed, unique URLs. Loading the URLs into the crawler, a filter was applied to remove URLs with extensions referring to file-types – images, videos, etc. – that we cannot currently process. This yielded 17,781,974 crawlable URLs. Crawling was disabled; in other words we set Nutch to download the content of the URLs, rather than to recursively follow further URLs. The download was run from August 2019 to December 2019, in which time 2,475,461 URLs were

---

[6] http://nutch.apache.org/
[7] https://lucene.apache.org/solr/

Table 1. Top 10 domains in terms of raw URLs vs. indexed URLs

№	Raw URLs	Indexed URLs
1	archive.org	bbc.co.uk
2	doi.org	nytimes.com
3	nih.gov	archive.org
4	nytimes.com	billboard.com
5	bbc.co.uk	newspapers.com
6	webcitation.org	thegazette.co.uk
7	allmusic.com	sports-reference.com
8	youtube.com	reuters.com
9	theguardian.com	baseball-reference.com
10	archive.is	bbc.com

successfully downloaded and indexed in Nutch. We let Nutch decide the order of the URLs to be accessed; no configuration was made in this matter.[8] Though not all URLs were processed at this point, progress in the crawler had slowed to a halt. An issue we did not anticipate was that of redirects: Nutch does not provide a clean mechanism to retrace redirects, though from the logs it is possible to retrace the URLs accessed and, in most cases, recover the original URLs. This was important to match the original Wikipedia articles and external URLs with the redirected content location of the indexed document. In total, we could recover links for 2,058,896 documents (83%) from their original Wikipedia article.

In Table 1 we present the top 10 domains for raw URLs extracted from Wikipedia and indexed (redirected) URLs. We see that the indexed URLs tend to refer to media sources. Notably (for example) `doi.org` is primarily a redirection service, and hence we do not see this domain appearing in the indexed URLs, which follow the redirects. Regarding *Coverage*, we managed to associate 3,899,953 (Q-identified) items with at least one indexed external reference. Of these, 1,136,477 items (29.1%) had more than one reference indexed.

Complete sample: Given the incompleteness of the crawl for the full reference corpus, we decided to also develop a complete crawl for a subset of Wikidata items. Based on some initial samples, which were largely composed of items without English Wikipedia articles, we decided to split our sample into five groups based on the `Q` identifiers: A: `Q1`–`Q10000`; B: `Q10001`–`Q100000`; C: `Q100001`–`Q1000000`; D: `Q1000001`–`Q10000000`; E: `Q10000001`–`Q100000000`. This sampling is based on the idea that Wikidata ids were defined chronologically, and that the most important entities (countries, major cities, recent presidents, etc.) would fall into the earlier groups, with later groups being populated by successively more obscure items. From each group we sample 1,000 items and then apply the same process as before; in this case we run the download to completion.

---

[8] By default, Nutch partitions URLs by host and then randomly selects URLs within each partition.

Table 2. Crawl for selected sample of five groups

	A	B	C	D	E	Total
Raw URLs	40,666	12,763	7,111	4,917	5,365	70,822
Indexed URLs	22,268	6,945	3,682	2,399	2,945	37,983

Table 2 indicates the number of raw URLs extracted from Wikipedia, and the number of URLs indexed. Corresponding to the design of each group, in general we see more references available for earlier groups; for example, group A contains many countries, whose articles in Wikipedia contain potentially hundreds of references. The difference between raw URLs and indexed URLs refers to duplicate or malformed URLs, filtered URLs, and URLs that returned 4xx or 5xx errors. For the 5,000 Wikidata items, we found 74 (1.4%) that used some reference also found in the indexed URLs from Wikipedia. On the other hand, of the 37,983 indexed URLs, only 163 (0.43%) were found to be used on one of the Wikidata items as a reference URL. We checked for exact URL matches, which may lead to under-reporting overlap, but these results offer strong support for the results of Piscopo et al. [12] indicating a low overlap in references between Wikipedia and Wikidata. This does not necessarily imply, however, that claims for Wikidata do not have support in the content of the references from Wikipedia.

It is worth noting that the download of references for some of the most popular items took tens of minutes to complete, which suggests that the online mode will often be too slow for interactive runtimes.

5 Search & Recommendation

We assume an inverted index of content of potential external references and now turn to the question of how to search with the documents. We assume that the API receives a claim as exemplified in Table 3, with the item IDs/terms, labels and aliases in English. Note that 1982 refers to a date value, where we use the lexical form as the label. There is no clear individual way to construct the search. Using just the labels may run the risk of missing some potentially relevant documents with alias terms. On the other hand using alias terms may introduce noise and return irrelevant documents. We experiment with four options:

1. construct a query for any of the three labels;
2. construct a query for any of the three labels or any property alias;
3. construct a query for any of the three labels or any alias;
4. construct a query for at least one label or alias for each of the three elements.

We provide examples of the searches for each of the four options in Table 4. While it may perhaps seem quite broad to use the OR connective, initial experience suggested that using AND, particularly on property labels (without aliases), meant that few documents were returned as the search was too specific. Furthermore,
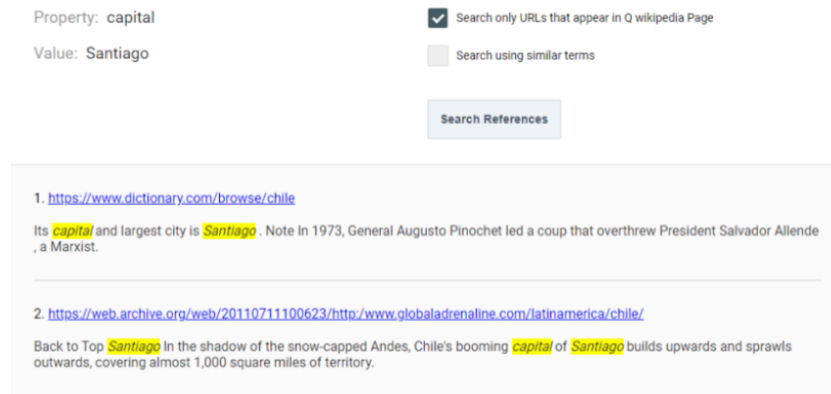
Fig. 2. Suggestions generated for "Chile capital Santiago" in a prototype user interface.

Table 3. Example of a Wikidata claim with IDs, labels and aliases

Ids/Terms:	Q232141	P571	1842
Labels (en):	University of Chile	inception	—
Aliases (en):	UChile	date founded	—
	Universidad de Chile	date created	—
	—	incorporated	—
	—	...	—

Solr uses the BM25F relevance metric (based on TF–IDF), which will rank documents with more occurrences of more terms more highly.

We consider searching only over the references of the article associated with the subject item (similar to the online option) to boost relevance,[9] and searching over all documents collected for the offline corpus to boost recall.

As a further feature, Solr allows for returning a snippet of each document determined to be a highly relevant part of the document for the search. The typical application of this feature is for building results lists, where the user can preview the most relevant part of the text, which also fits our use-case of letting editors preview snippets of text from different documents that might support a given claim. We illustrate this feature in Figure 2 for an example claim.

Held-out evaluation: As an initial test of the different search options, for our set of 5000 items, we can use the 163 URLs that appear on a Wikidata claim

---

[9] Another alternative would be to further include documents for the value item. We discarded this option in order to simplify experiments, observing that the value of a claim is often much more general than the subject item; for example, considering the claim that Neil Young was born in Canada, it would not make sense to search within the external references for Canada.

Table 4. Example searches for the four options considered based on Table 3

Option 1:	`"university of chile" OR "inception" OR "1842"`
Option 2:	`"university of chile" OR "inception" OR "date founded"` `OR ... OR "1842"`
Option 3:	`"university of chile" OR "la u de chile" OR ...` `OR "inception" OR "date founded" OR ... OR "1842"`
Option 4:	`("university of chile" OR "la u de chile" OR ...)` `AND ("inception" OR "date founded" OR ...) AND "1842"`

Table 5. Recall@3 for the four search options and a random baseline

	Option 1	*Option 2*	*Option 3*	*Option 4*	*Random*
R@3	0.72	0.64	0.66	0.57	0.37

and appear in our index. We take the Wikidata claim that they appear on, and measure the recall of the 163 URLs in the top 3 suggestions for each option searching with the external references of the article associated with the subject item. We also consider a baseline that selects 3 random external references for the article of the subject item. The results are shown in Table 5, where we see that the best results are offered by search option 1, which retrieves the known external reference as a top-3 suggestion in 72% of the cases. It is important to note that any result returned may be correct as we only know a subset of the correct references, so the recall should be interpreted as a lower bound.

Gold standard evaluation: Given the aforementioned limitations of the held-out experiments, we opted to manually label a subset of claims, where we choose 5 items from each of the five groups A–E, which we then labelled. The labelling indicates which external reference in the Wikipedia article associated with the chosen (subject) item supports which claim on that item. We first tried a random sampling of 5 items from each group but labelling became infeasible as items with hundreds of associated external references and claims were found, where manually pairing them off was considered too complex; furthermore, in the later groups, some items had only one reference associated. Instead we choose to sample items with a number of associated references close to the mean for that group. We show some statistics for the gold standard in Table 6, where we indicate the average number of claims in Wikidata per item, the average number of references indexed from the corresponding Wikipedia articles per item, and average percentage of claims per item supported by at least one reference from the corresponding Wikipedia article. The All column considers the statistics across all groups. It is worth noting that given the low numbers of references for groups C–E, the results for searches become somewhat trivial; for this reason we will include random baselines. The searches of our gold standard are then formed by the claims for which at least one supporting reference is found.

Table 6. High-level statistics from the gold standard

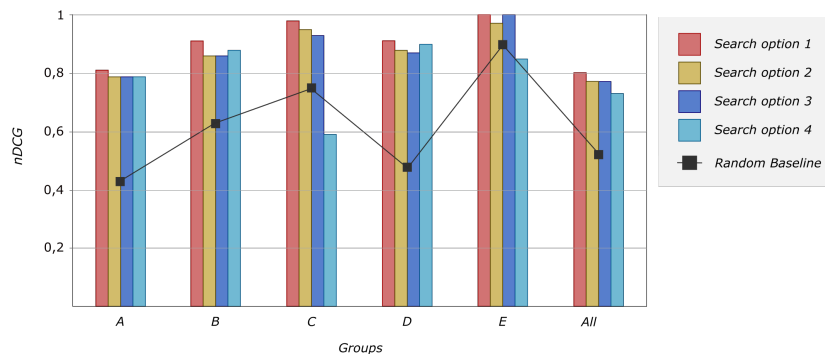|	A	B	C	D	E	All
Average claims	48	18	17	13	8	21
Average indexed references	23	7	4	2	3	7
Average claims supported	42%	27%	26%	52%	31%	37%

Fig. 3. nDCG for search methods on the gold standard

In Figure 3 we present the normalised Discounted Cumulative Gain (nDCG) metric for the different search options with respect to the different groups. We also include the random baseline for comparison. Intuitively speaking, a score of 1 indicates the best possible ordering possible, ranking all supporting references above all non-supporting references. The results are divided by group. We see that for groups A and B, the search methods perform much better than the random baseline. The best results are given for group E, but this is largely due to the trivial nature of the task when given few references, as noted by the high performance of the random baseline. In general there is not much difference of note between the different search options, though we can perhaps indicate that *Option 1* performs (slightly) best and *Option 4* performs worst.

Given that the nDCG measure is somewhat difficult to interpret, in Figure 4 we present the Any@k measure: noting that in order to establish verifiability, in general one reference is sufficient, we look at the percentage of claims/searches for which at least one suggestion in the top-k was relevant. We believe that this gives a more direct measure of how the reference suggestions perform in practice. We see that considering the top-3 results, *Options 1–3* succeed in finding supporting references in close to 88% of cases, increasing to 90% for top-4 results.

Snippets: For the 25 gold standard items, we manually evaluated the text snippets that Solr selects to indicate why a document is relevant, where we found that only 9% of these snippets were sufficient to support a claim by themselves, although they were often useful to help understand more about the content of
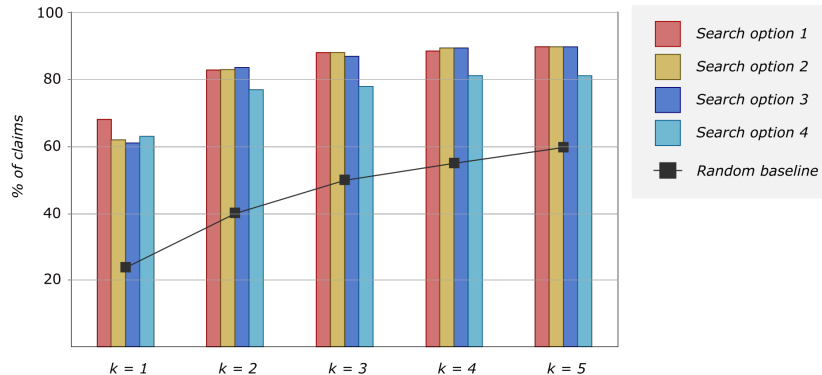
Fig. 4. Any-at-k for search methods on the gold standard

the webpage without visiting it. In particular, we found that reference documents often support claims in a more implicit way, requiring a more general understanding of different parts of the text, rather than just one part.

Global results: Finally we used our 25 gold standards to run searches over the full corpus of 2.5 million references using the search options previously outlined. The results were largely negative: the best results were obtained using *Option 1*, which yielded Any@5 values of 19%. Manually revising the results, we found that most of the documents returned by Solr were irrelevant to the topic at hand, due to the broader corpus being used. It may, however, be possible to better fine-tune the queries to return better results.

6 Conclusions

We now briefly summarise our insights regarding research questions **RQ1–5**.

RQ1 *Offline vs. online indexing*: Online indexing was slow, with references for well-known entities taking up to 20 minutes to download and index. However, achieving a complete corpus by offline indexing is very time consuming.

RQ2 *Coverage*: Similar to the results of Piscopo et al. [12], we find low overlap between references in Wikipedia and Wikidata; in terms of our goal standard developed for a small sample of 25 items, we estimate that about 37% of claims had supporting references in their corresponding Wikipedia articles.

RQ3 *Search phrasing*: The best results were given by using an or connective on primary labels, though including aliases gave similar results.

RQ4 *Relevance*: BM25F gave good results when searching for claims within the references of the corresponding Wikipedia article, but poor results for the given search phrasing options when considering the full corpus.

RQ5 *Suggestion Quality*: When a claim has a supporting reference in the corresponding Wikipedia article for the subject item, the proposed method will

find at least one such supporting reference in the top-5 results around 90% of the time; however, the generated snippets rarely suffice to support the claim, meaning the editor will often have to visit and revise the documents.

When claims are supported by references in the corresponding Wikipedia article, traditional Information Retrieval methods appear sufficient to give good recommendations. The more general issue we encountered in this initial research is that few Wikidata claims have relevant references in the corresponding Wikipedia article. This suggests two possible future directions:

- *Offline*: Given that some Wikidata items do not have an associated Wikipedia article, that many Wikipedia articles have few references, etc., it would be interesting to develop a broader corpus with more documents from the Web, perhaps from the Common Crawl. In order to ensure that the documents are authoritative, this corpus might only include content from web-sites with a threshold number of references detected in Wikipedia. A challenge will be to ensure the relevance of search results, where the connection between the Wikidata items and the indexed documents would be lost; however, this challenge could be addressed with more advanced relevance measures based on the fields of the documents, comparing the similarity of each document's content to relevant Wikipedia articles, amongst other such techniques.
- *Online*: We have found that our online option is too slow due to the need to crawl references at runtime. Another option similar to the online option – in terms of obviating the need for a local index of documents – would be to use the existing infrastructure of major search engines to search the Web at runtime, filtering for sites that are considered authoritative. A major benefit of such an approach is that the (costly) retrieval, indexing and refreshing of content could be delegated to the search engine. The downside of such an approach would be the issues of respecting rate-limits for the search API, plus the inability to pre-process the content for the specific task.

In summary, a method for automatically suggesting references for Wikidata claims would help human editors to be more productive, and would help to make better use of their (often volunteered) time. As a result, the coverage of references on Wikidata would increase, and its quality as a secondary source of knowledge would improve. While this paper does not provide a definitive solution, we have gained some important insights into the strengths and limitations of basing suggestions on Wikipedia's references. We further provide online material to facilitate future research, including the retrieved content of a large subset of documents found in the reference sections of English Wikipedia.

Material online. Available on Zenodo [2].

References

1. Chou, A.J., Gonçalves, G., Walton, S., Redi, M.: Citation Detective: a Public Dataset to Improve and Quantify Wikipedia Citation Quality at Scale. In: Wiki-Workshop. pp. 1–5 (2020)
2. Curotto, P., Hogan, A.: External References of English Wikipedia (ref- wiki-en) (Aug 2020), `https://doi.org/10.5281/zenodo.4001139`
3. Delpeuch, A.: Structured citations in the English Wikipedia (Jun 2016), `https://doi.org/10.5281/zenodo.55004`
4. Kousha, K., Thelwall, M.: Are Wikipedia citations important evidence of the impact of scholarly articles and books? J. Assoc. Inf. Sci. Technol. **68**(3), 762–779 (2017). https://doi.org/10.1002/asi.23694, `https://doi.org/10.1002/asi.23694`
5. Lemus-Rojas, M., Pintscher, L.: Wikidata and Libraries: Facilitating Open Knowledge. In: Leveraging Wikipedia: Connecting Communities of Knowledge. pp. 143–158. ALA Editions (2018)
6. Lewoniewski, W., Wecel, K., Abramowicz, W.: Analysis of References Across Wikipedia Languages. In: Information and Software Technologies (ICIST). pp. 561–573. Springer (2017)
7. Luzuriaga, J., Hogan, A., Muñoz, E., Rosales, H.: Wikitables (Oct 2019), `https://doi.org/10.5281/zenodo.3483254`
8. Nielsen, F.Å., Mietchen, D., Willighagen, E.L.: Scholia, Scientometrics and Wikidata. In: ESWC Satellite Events. pp. 237–259. Springer (2017)
9. Piccardi, T., Redi, M., Colavizza, G., West, R.: Quantifying Engagement with Citations on Wikipedia. In: The Web Conference (WWW). pp. 2365–2376. ACM / IW3C2 (2020)
10. Piscopo, A., Kaffee, L., Phethean, C., Simperl, E.: Provenance Information in a Collaborative Knowledge Graph: An Evaluation of Wikidata External References. In: International Semantic Web Conference (ISWC). pp. 542–558. Springer (2017)
11. Piscopo, A., Simperl, E.: What we talk about when we talk about Wikidata quality: a literature survey. In: International Symposium on Open Collaboration (OpenSym). pp. 17:1–17:11. ACM (2019)
12. Piscopo, A., Vougiouklis, P., Kaffee, L., Phethean, C., Hare, J.S., Simperl, E.: What do Wikidata and Wikipedia Have in Common?: An Analysis of their Use of External References. In: International Symposium on Open Collaboration (OpenSym). pp. 1:1–1:10. ACM (2017)
13. Redi, M., Fetahu, B., Morgan, J.T., Taraborelli, D.: Citation needed: A taxonomy and algorithmic assessment of Wikipedia's verifiability. In: The Web Conference (WWW). pp. 1567–1578. ACM (2019)
14. Singh, H., West, R., Colavizza, G.: Wikipedia Citations: A comprehensive dataset of citations with identifiers extracted from English Wikipedia. CoRR **abs/2007.07022** (2020)
15. Vrandečić, D., Krötzsch, M.: Wikidata: A Free Collaborative Knowledgebase. Comm. ACM **57**, 78–85 (2014)
16. Warncke-Wang, M., Cosley, D., Riedl, J.: Tell me more: an actionable quality model for Wikipedia. In: International Symposium on Open Collaboration (OpenSym). pp. 8:1–8:10. ACM (2013)