

How to contribute to Scholia in 10 easy steps

Finn Årup Nielsen

DTU Compute
Technical University of Denmark

27 May 2020

Scholia

SCHOLIA Author Work Organization Location Event Project Award Topic Tools Help Search...

author

Denny Vrandečić (Q18618629)

Related: Egon Willighagen · Dietmar Schomburg · Patrick O. Brown · Christoph Steinbeck · Cowen · Ion Idriess · Arne Skerra · Joseph Schlessinger · Aled Edwards · Tyler

<https://orcid.org/0000-0002-9593-2294>

List of publications

Show 10 entries

Date	Work	Type	Pages	Venue	Authors
2019-09-23	Wikidata: A large-scale collaborative ontological medical database	scientific article	13	Journal of Biomedical Informatics	Helmi Hamdi, Mohamed Ali Hadj Taieb, Mohamed Ben Aouicha, Houcemeddine Turki, Thomas Shafee, Diptanshu Das, Denny Vrandečić
2019-01-01	Describing datasets in Wikidata	scientific article			Denny Vrandečić
2018-10-01	Capturing meaning: Toward an abstract Wikipedia	document	4		Denny Vrandečić
2017-10-01	Using WikiData as a multi-lingual multi-dialectal dictionary for Arabic dialects	scientific article	6	2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)	Helmi Hamdi, Houcemeddine Turki, Denny Vrandečić
2017-09-04	RDFIO: extending Semantic MediaWiki for interoperable biomedical data management.	scientific article	13	Journal of Biomedical Semantics	Samuel Lampa, Pekka Kohonen, Roland Grafström, Ola Spjuth, Denny Vrandečić, Egon Willighagen
2016-01-01	From Freebase to Wikidata: The Great Migration	scientific article	10	Proceedings of the 25th International Conference on World Wide Web	Thomas Pellissier Tanon, Thomas Steiner, Denny Vrandečić, Lydia Pintscher, Sebastian Schaffert

Scholia: online profiles extracted via SPARQL from Wikidata information: <https://tools.wmflabs.org/scholia/>

Profiles can be about authors, works, venues, organizations, events, chemical compounds, clinical trials, taxon, etc.

Here about Denny Vrandečić at <https://tools.wmflabs.org/scholia/author/Q18618629>: list of publications, topics, citations, etc.

1: Wikidata editing



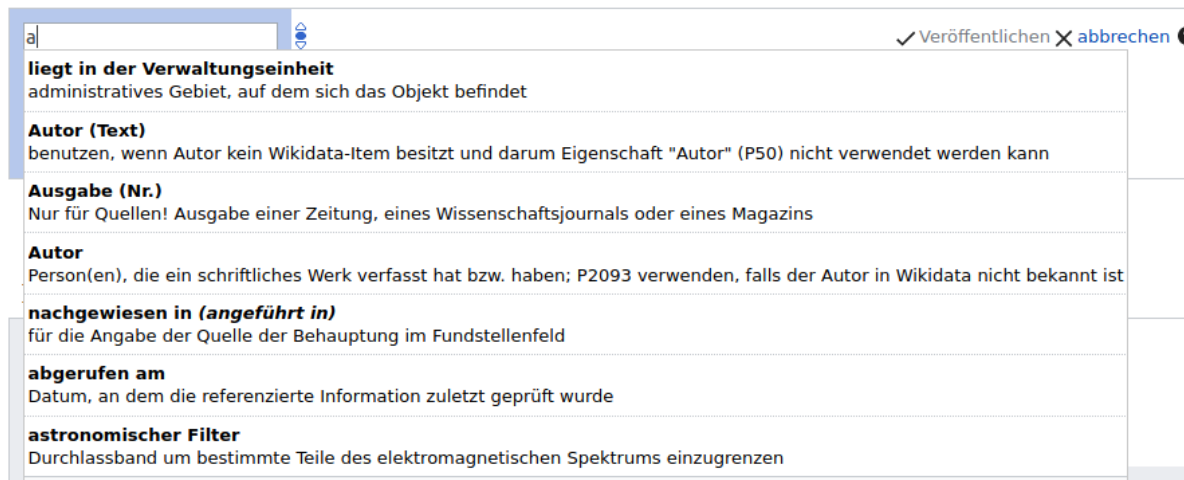
exakte Übereinstimmung <http://www.aifb.uni-karlsruhe.de/Personen/viewPersonOWL/id2097instance> [bearbeiten](#)

▼ 0 Fundstellen

[+ Fundstelle hinzufügen](#)

[+ Wert hinzufügen](#)

Most of Scholia's data comes from Wikidata (a bit is extracted from the English Wikipedia and English Wiktionary).



al [Veröffentlichen](#) [abbrechen](#) [?](#)

liegt in der Verwaltungseinheit
administratives Gebiet, auf dem sich das Objekt befindet

Autor (Text)
benutzen, wenn Autor kein Wikidata-Item besitzt und darum Eigenschaft "Autor" (P50) nicht verwendet werden kann

Ausgabe (Nr.)
Nur für Quellen! Ausgabe einer Zeitung, eines Wissenschaftsjournals oder eines Magazins

Autor
Person(en), die ein schriftliches Werk verfasst hat bzw. haben; P2093 verwenden, falls der Autor in Wikidata nicht bekannt ist

nachgewiesen in (angeführt in)
für die Angabe der Quelle der Behauptung im Fundstellenfeld

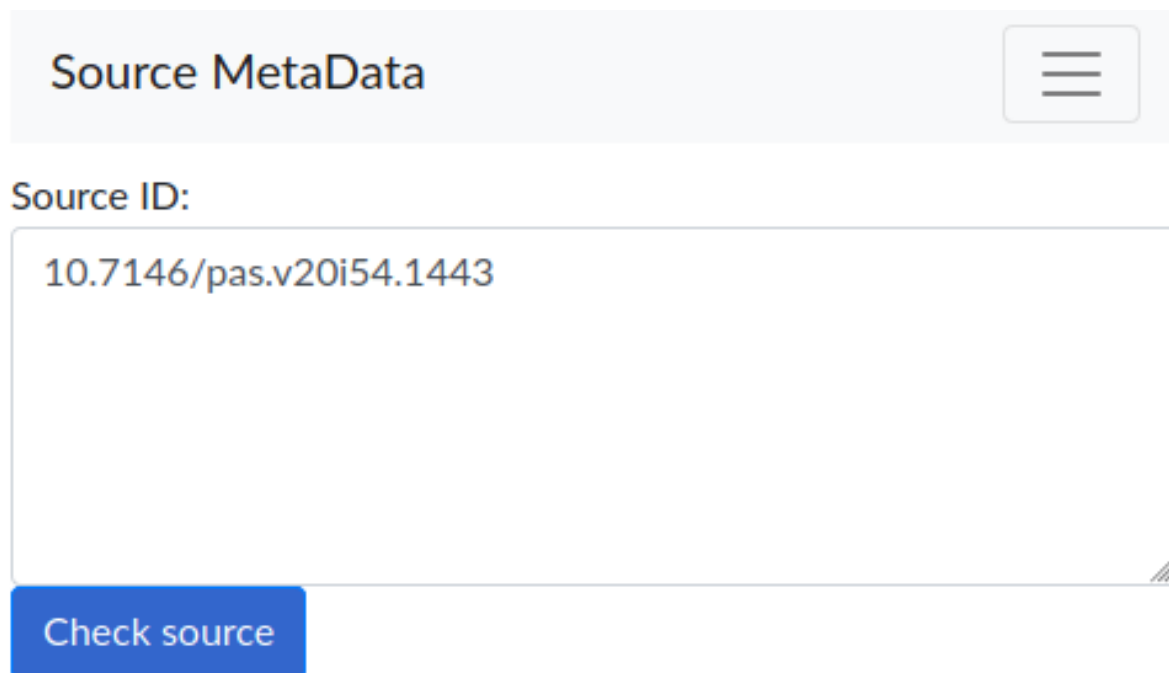
abgerufen am
Datum, an dem die referenzierte Information zuletzt geprüft wurde

astronomischer Filter
Durchlassband um bestimmte Teile des elektromagnetischen Spektrums einzugrenzen

Any change in data should be edited in Wikidata, and direct links to Wikidata are provided from Scholia.

Changes in Wikidata will not show up immediately in Scholia due to cache and delay between Wikidata and the Wikidata Query Service SPARQL endpoint.

2: Sourcemd



Source MetaData

Source ID:

10.7146/pas.v20i54.1443

Check source

Magnus Manske's *Source MetaData* (sourcemd) tool for entry of scientific article metadata via an identifier, – either DOI, the PubMed identifier (PMID) or the PubMed Central identifier: “PMID/DOI/PMCID, one per row” .

https://tools.wmflabs.org/sourcemd/index_old.php

Works entirely independent of Scholia.

Here with the DOI for the Danish article *Alt er undergivet Forvandling*.

3: Author-disambiguator

Author Disambiguator

(Log in to your Wikimedia account to use OAuth instead of Quickstatements for updates - still experimental.)

Author name: Fuzzy match Wikibase search

Specify name strings Limit:

Additional SPARQL filters separated by semicolons (eg. for papers on Zika virus, enter wdt:P921 wd:Q202864):

Filter potential authors as well?

Potential publications

22 publications found

[Click here to create clusters based on exact author strings rather than rougher matches.](#)

Group #1

[Check all](#) | [Uncheck all](#)

Title	Authors (identified)	Published In	Identifier(s)	Topic	Published Date	Match?
Plasma malondialdehyde as biomarker for oxidative stress: reference interval and effects of life-style factors.	[1] <input checked="" type="checkbox"/> Nielsen F, [2] Mikkelsen BB, [3] Nielsen JB, [4] Andersen HR, [5] Philippe Grandjean [Full author list]	Clinical Chemistry [missing]	PubMed: 9216458 [ORCID]	biomarker [missing]	1997-07-01	

Arthur P. Smith's (and Magnus Manske's AFAIK) **Author Disambiguator** tool for semi-automatically resolving author strings to author items.

Works independently of Scholia, but links are provided back and forth between Scholia and Author Disambiguator

Here searching on “Nielsen F” and the tool clusters matched articles.

4: Quickstatements

Magnus Manske's Quickstatement format for mass-entry in Wikidata.

Example with metadata for an article

```
CREATE
```

```
LAST P356 "10.7146/PAS.V20I54.1443"
```

```
LAST P31 Q13442814
```

```
LAST P433 "54"
```

```
LAST P478 "20"
```

```
LAST P577 +2005-12-01T00:00:00Z/11
```

```
LAST P1433 Q50398672
```

```
LAST P2093 "Stefan Iversen" P1545 "1"
```

Associated webservice at <https://tools.wmflabs.org/quickstatements/>

Works independently of Scholia.

5: “Missing” in Scholia

author
Finn Årup Nielsen (Q20980928)

Missing information with respect to the author.

Author name strings to be resolved

The authorship may be represented as an author name string rather than a Wikidata item. To try to resolve the author name strings, follow the links below to use the Author disambiguator tool, which also has a [dedicated page for this author](#).

Show entries

Search:

Count	Author name	Author resolver url
34	F Nielsen	https://tools.wmflabs.org/author-disambiguator/?doit=Look+for+author&name=F%20Nielsen
6	Finn Nielsen	https://tools.wmflabs.org/author-disambiguator/?doit=Look+for+author&name=Finn%20Nielsen
0	F A Nielsen	https://tools.wmflabs.org/author-disambiguator/?doit=Look+for+author&name=F%20A%20Nielsen
0	Finn A Nielsen	https://tools.wmflabs.org/author-disambiguator/?doit=Look+for+author&name=Finn%20A%20Nielsen
0	Finn Å. Nielsen	https://tools.wmflabs.org/author-disambiguator/?doit=Look+for+author&name=Finn%20%C3%85.%20Nielsen

Subpages for items in Scholia indicating data that might be missing.

Somewhat “hidden” in the interface, but you can try and append “/missing” to the URL, e.g., missing for authors is listing possible data that are missing in Scholia such as author disambiguation and main topic specification on articles.

Example: <https://tools.wmflabs.org/scholia/author/Q20980928/missing>

6: Scholia's arxiv identifier

arXiv to Quickstatements

Input

Copy and paste an ID from the [arXiv](#) preprint repository. Bare IDs (such as "1703.06103") and URLs (such as "https://arxiv.org/abs/1703.06103") both work.

The input ID will be queried in Wikidata and Quickstatements will not be generated if the input ID is found in Wikidata. However, note that new items may not be immediately found because of caching.

Result

```
CREATE
LAST P818 "2005.08101"
LAST P31 Q13442814
LAST Len "The Missing Path: Diagnosing Incompleteness in Linked Data"
LAST P1476 en:"The Missing Path: Diagnosing Incompleteness in Linked Data"
LAST P577 +2020-05-16T00:00:00Z/11
LAST P953 "https://arxiv.org/pdf/2005.08101.pdf"
LAST P820 "cs.HC"
LAST P2093 "Marie Destandau" P1545 "1"
LAST P2093 "Jean-Daniel Fekete" P1545 "2"
```

[Forward to Magnus Manske's quickstatements](#)

Scholia's has the ability to extract metadata from the arXiv.org preprint server and format the metadata to Quickstatements.

Copy-and-paste an arxiv.org identifier into an edit field, press the button, forward to Magnus Manske's quickstatement service with mass-entry there.

Should also work from the command line version of Scholia:

```
$ python -m scholia arxiv-to-quickstatements 1710.04099
```


7: NeurIPS scraping

Conference on Neural Information Processing Systems (Q1961016)

The Conference and Workshop on Neural Information Processing Systems (abbreviated as NeurIPS and formerly NIPS) is a machine learning and computational neuroscience conference held every December. The conference is currently a double-track meeting (single-track until 2015) that includes invited talks as well as oral and poster presentations of refereed papers, followed by parallel-track workshops that up to 2013 were held at ski resorts. ... (from the [English Wikipedia](#))

[@NipsConference](#)

List of events

Show entries Search:

Year	Short name	Event	Proceedings
2020	NeurIPS 2020	Thirty-fourth Conference on Neural Information Processing Systems	
2019	NeurIPS 2019	Thirty-third Conference on Neural Information Processing Systems	Advances in Neural Information Processing Systems 32
2018	NeurIPS 2018	Thirty-second Annual Conference on Neural Information Processing Systems	Advances in Neural Information Processing Systems 31
2017	NIPS 2017	Thirty-first Conference on Neural Information Processing Systems	Advances in Neural Information Processing Systems 30
2016	NIPS 2016	Thirtieth Annual Conference on Neural Information Processing Systems	Advances in Neural Information Processing Systems 29
2015	NIPS 2015	Twenty-ninth Conference on Neural Information Processing Systems	Advances in Neural Information Processing Systems 28

Articles from the proceedings of the machine learning conference *Neural Information Processing Systems* have been scraped with the functionality implemented in the file `/scholia/scrape/nips.py` with output in the Quickstatements format.

Example with *Quaternion Knowledge Graph Embeddings* conference paper:

```
$ python -m scholia.scrape.nips paper-url-to-quickstatements https://papers.nip ...  
# Q63295515 is https://papers.nips.cc/paper/8541-quaternion-knowledge-graph-em ...
```

(Does not generate Quickstatement because it is already entered).

8. Other scraping

Command-line version of Scholia has the ability of scrape information from some other websites, — implemented in modules in `/scholia/scrape/`

Bespoke scrapers may be necessary and not yet fully working :(

Licenses may be an issue, c.f. discussion at the Association for Computational Linguistics, e.g., [Clarify the copyright of ACL Anthology metadata #580](#).

9: Bespoke crawling and mass editing

Hope that other Wikidata editors will do the work:

Large-scale topic annotation by Daniel Mietchen.

Scientific article references from CrossRef.

Clinical trials data.

ORCID information and scraping of university websites.

See, e.g., *How the world's first Wikidata Visiting Scholar created linked open data for five thousand works of art* for Simon Cobb

Some can be gamified (Magnus Manske's tool)

10: Ordia lexemes

Text to lexemes

Eine durch die Digitalisierung veränderte und auf Open Science ausgerichtete Wissenschaftspraxis benötigt angepasste Infrastrukturen und Services. Daraus ergeben sich verschiedene neue oder veränderte Aktionsfelder für wissenschaftliche Bibliotheken und Infrastruktureinrichtungen. Zu nennen sind zum Beispiel die nicht-textuellen Materialien wie Forschungsdaten, AV-Medien oder Software und die Umsetzung der FAIR-Prinzipien. Hinzu kommen neue Aufgaben im Bereich der Forschungsinformationen, zum Beispiel in der Unterstützung institutioneller Forschungsinformationssysteme, die Gestaltung von Open Access, die Unterstützung kollaborativen wissenschaftlichen Arbeitens sowie die Schaffung von offenen Infrastrukturen. In diesem Artikel werden diese Felder kurz vorgestellt und sich daraus abzeichnende Anforderungen an das bibliothekarische Berufsbild skizziert.

Language:

Deutsch - de - German

Casing:

Lowercase first sentence letters

Submit

Extraction

Search:

Word	Form	Lexeme	Lexical category	Fea
AV-Medien				
Access				
Aktionsfelder				

Ordia's text-to-lexemes facility from <https://tools.wmflabs.org/ordia/text-to>

Scholia-inspired web application, but for Wikidata's multilingual lexicographic information.

Here with abstract from *Open Science und die Bibliothek – Aktionsfelder und Berufsbild* and unknown words extracted, e.g., *Aktionsfelder*.

Link between Scholia and Ordia, e.g., dictionaries describing words, articles about words, see, e.g., *gehen* at <https://tools.wmflabs.org/scholia/lexeme/L1026>.

Thanks