

WikiCite als kollaborativer Ansatz für offene Zitationsdaten

Philipp Zumstein (UB Mannheim) Jakob Voß (VZG)



{ } wikicite

13. Juni 2018, Bibliothek*tag Berlin



Einführung & Motivation







Termine

- ▶ Geburtstag von Wikidata
- ▶ Wikibase-Workshop an der DNB
- ▶ Dritte WikiCite-Konferenz
- ▶ ...

<https://wiki.dnb.de/display/GND/Authority+Control+meets+Wikibase>

Was ist WikiCite?

- Kontext: Wikipedia, **Wikidata**, Wikisource...

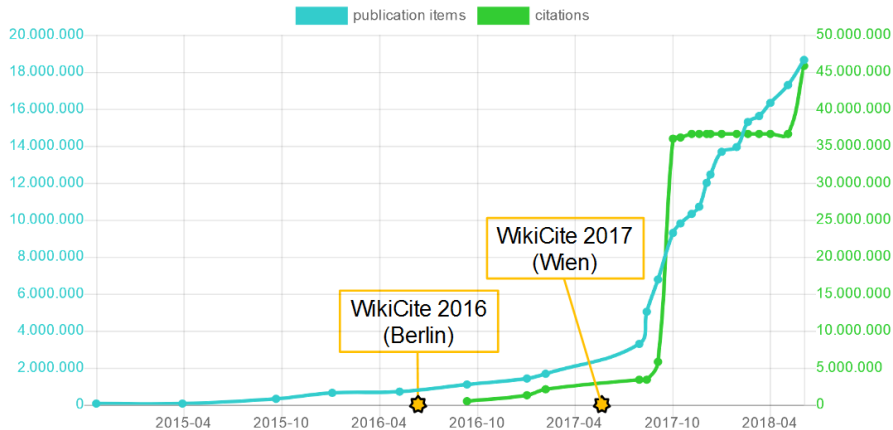
ist ein(e)	 Projekt  bearbeiten
	▼ 0 Fundstelle
	 Konferenzreihe  bearbeiten
	▼ 0 Fundstelle
	 Online-Community  bearbeiten
	▼ 0 Fundstelle
+ Wert hinzufügen	

Ziele von WikiCite

- ▶ Freie bibliographische Datenbank in Wikidata
 - ▶ Bibliographische Metadaten
 - ▶ Zitationsdaten
- ▶ Selbstständige + unselbstständiger Werke + Beziehungen
 - ▶ Kollaborative Katalogisierung (Social Cataloging)
 - ▶ offene Daten (CC0)

Umfang

- ▶ **Publikationen:** 19 Millionen Wikidata-Einträgen (39%)
- ▶ **Zitationen:** 46 Millionen Zitationen in Wikidata (10%)



Motivation für Wikipedia

- Aussagen mit Belegen versehen, z.B.

Berlin gilt als [Weltstadt](#) der Kultur, Politik, Medien und Wissenschaften.^{[15][16][17][18]} Die Universitäten, Forschungseinrichtungen, Sportereignisse und Museen Berlins genießen internationalen Ruf.^[19] Die [Metropole](#) trägt den UNESCO-Titel *Stadt des Designs* und ist eines der meistbesuchten Zentren des Kontinents.^[20] Berlins Architektur, Festivals, Nachtleben und vielfältige Lebensbedingungen sind weltweit bekannt.^[21]

15. ↑ [Global Power City Index 2015](#) (PDF; 1,2 MB) Abgerufen am 10. April 2016.
16. ↑ [Revealed: Cities that rule the world – and those on the rise.](#) (CNN (englisch), abgerufen am 14. April 2010).
17. ↑ [Leading cities in cultural globalisations/Media](#) (englisch), GaWC, abgerufen am 14. April 2010.
18. ↑ [Berlin – Stadt der Nobelpreisträger.](#) In: [luisse-berlin.de](#). Abgerufen am 3. September 2015.
19. ↑ [London crowned as SportBusiness' 2012 Ultimate Sports City](#) (Memento vom 18. August 2012 im *Internet Archive*). In: [sportsbusiness.com](#), abgerufen am 26. November 2012.
20. ↑ [Tourismusreport 2014.](#) IHK Berlin.
21. ↑ [Berlin, the Big Canvas.](#) In: [nytimes.com](#). [The New York Times](#), abgerufen am 16. April 2010 (englisch).

- Bibl. Metadaten zentral für alle Artikel und Sprachversionen

Motivation für Wikidata

- Fakten in Wikidata mit Fundstellen belegen, z.B. höchster Punkt in Berlin

höchster Punkt

Arkenberge ...

bearbeiten

▼ eine Fundstelle

kopieren

URL der Fundstelle	http://www.tagesspiegel.de/berlin/teufelsberg-oder-arkenberge-zum-wettstreit-um-den-hoechsten-gipfel-berlins/11413932.html
Titel	Zum Wettstreit um den höchsten Gipfel Berlins (Deutsch)

+ Fundstelle hinzufügen

+ Wert hinzufügen

Motivation für bibliometrische Auswertungen

► Auswertung und Visualisierung von Publikationsdaten

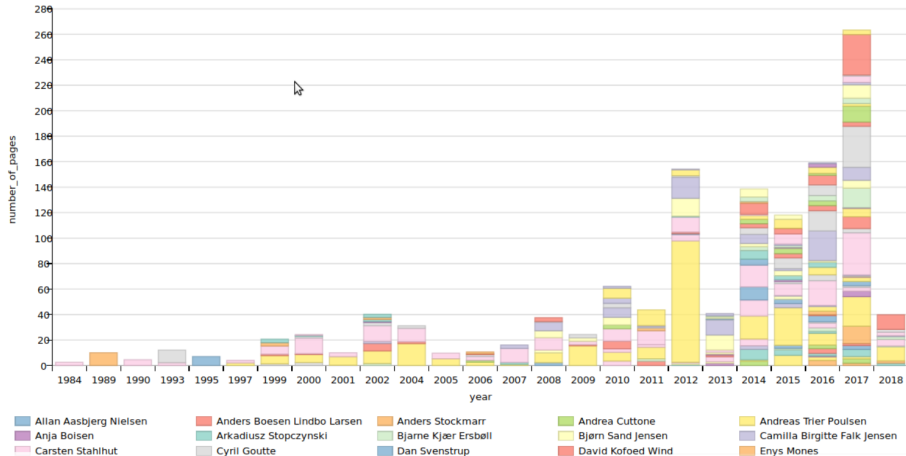
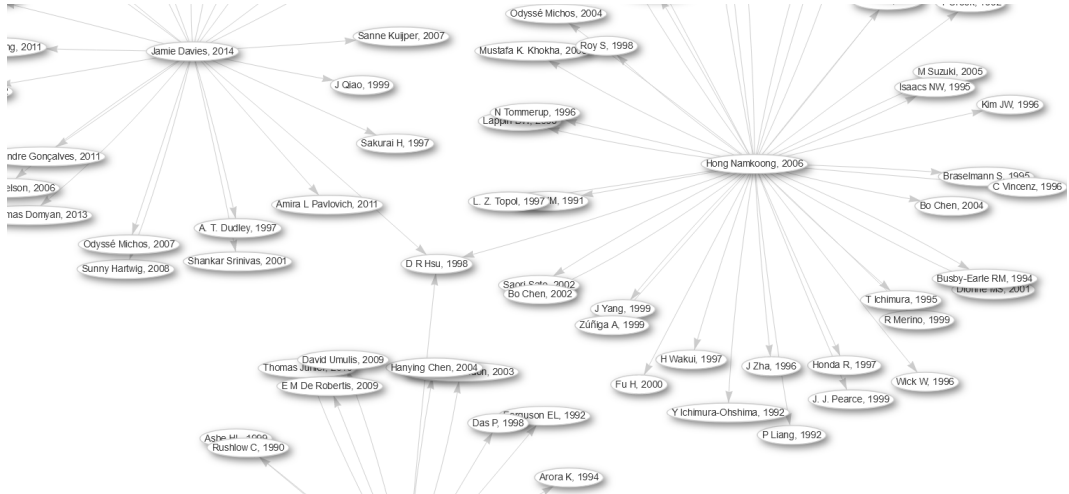


Abb. Screenshot von Scholia zur Anzahl produzierter Seiten von Publikationen mit Autoren der DTU

Motivation für bibliometrische Auswertungen

► Auswertung und Visualisierung von Zitationsdaten



Umsetzung von WikiCite

Bezeichnung: The Synergistic Activity of Thyroid Transcription Factor 1 and Pax 8 Relies

ID: (Q30000000)...

Beschreibung: wissenschaftlicher Artikel aus dem Jahr 2002

[In weiteren Sprachen](#)

mehrsprachig

Bearbeitungsmöglichkeiten

Aussagen

1. Aussage:

ist ein(e)



wissenschaftlicher Artikel ...



[bearbeiten](#)

↳ 0 Fundstellen



[Fundstelle hinzufügen](#)



[Wert hinzufügen](#)

2. Aussage:

Titel



The Synergistic Activity of Thyroid Transcription Factor 1 and Pax 8 Relies on the Promoter/Enhancer Interplay (Englisch)



[bearbeiten](#)

↳ eine Fundstelle

Fundstellen für Aussagen



[Wert hinzufügen](#)

3. Aussage:

Schlagwort



Transkription ...



[bearbeiten](#)

↳ eine Fundstelle



[Wert hinzufügen](#)

...

Wikidata-Datenmodell

- ▶ Item-Identifizier
 - ▶ Q300000000
- ▶ Eigenschaften
 - ▶ P31 *ist ein(e) / instance of / est ...*
 - ▶ P50 *Autor / author / scriptor ...*
 - ▶ ...
- ▶ Datenmodell ist dynamisch
 - ▶ Neue Eigenschaften vorschlagen + diskutieren
 - ▶ Fragen und Diskussion auf spezifischen Projektseiten
Metadata, Books, Periodicals...

Tools zur Bearbeitung

Direkte Bearbeitung auf Wikidata oder...

- ▶ **sourceMD/fatameh** Import anhand DOI/PMID
- ▶ **zotkat** Import aus Zotero (Literaturverwaltung)
- ▶ **Mix'n'Match:** Abgleich mit externen Identifikatoren
- ▶ **OpenRefine 3.0** mit erweiterter Wikidata-Anbindung
- ▶ ...

Qualitätssicherung

- ▶ Eigene Beobachtungslist für Wikidata-Einträge mit Benachrichtigung bei Änderungen
- ▶ Einschränkungen von Werten von Eigenschaften
 - ▶ *ISBN* (P212/P957) folgen bestimmten Mustern
 - ▶ *Autoren* (P50) müssen Personen oder Organisation sein
- ▶ Regelverletzungen werden protokolliert (constraint reports) und können dann verbessert werden
- ▶ bottom-up

Normdaten-Verlinkung

- ▶ Wikidata ist selbst eine universelle Normdatei
- ▶ Verknüpfungen mit externen Einträgen
 - ▶ DOI, ISBN, PMID, arXiv-ID, ... (Publikationen)
 - ▶ GND, VIAF, ORCID, dblp-ID, ... (Autoren...)
 - ▶ ...
- ▶ 2824 von 4741 Wikidata-Eigenschaften dienen der Verlinkung mit anderen Datenbanken

Abfragemöglichkeiten

- ▶ <https://query.wikidata.org/> (SPARQL-Editor)
- ▶ Beispielabfragen
 - ▶ Bücher zusammen von Elter(n) und Kinder(n) geschrieben:
<http://tinyurl.com/y7oe8uo8>
 - ▶ Autoren, deren Werke 2018 gemeinfrei wurden:
<http://tinyurl.com/yc8v6l9c>
 - ▶ Meist zitierte Wissenschaftlerinnen:
<http://tinyurl.com/y8ell8nw>
- ▶ Hilfreich zur Qualitätssicherung

Perspektiven für Bibliotheken

Nachnutzung von Daten aus Wikidata

- ▶ Kataloganreicherungen bei Normdaten
- ▶ mehrsprachige Suche nach Schlagwörtern
- ▶ Artikelindex für Discovery aus Wikidata?
- ▶ Kataloganreicherungen mit offenen Zitationsdaten?

Ideen für Katalogisierungstools

- ▶ Können Ansätze zur Datenerfassung in Wikidata auch für die Katalogisierung in Bibliotheken übernommen werden?
- ▶ Sind nicht auch stärker Tools zur Batch-Bearbeitungen gefragt?
- ▶ Was ist mit Kontroll- und Prüftools?

Umfang der Katalogisierung

- ▶ Aufsatzkatalogisierung
 - ▶ Thema für Bibliotheken? FIDs?
 - ▶ WikiCite zeigt auf, dass es möglich ist
- ▶ Katalogisierung von Zitationen
 - ▶ Können Bibliotheken zu offenen Zitationsdaten beitragen?
 - ▶ Zitationsdaten in Verbundkatalog speichern?
 - ▶ Linked Open Citation Database (LOC-DB): Clip

Ausblick

WikiCite ist schneller

	GBV/CBS	Wikidata/Wikibase
Mit-Katalogisieren	Monate-Jahre	Minuten Tage-Wochen
Neue Kategorie	Monate-Jahre	Tage-Wochen
Gesamtabzug	Wochen-Monate	

Qualität durch Qualitätskontrollen

- ▶ Regeln zur Datenanalyse (was ist drin) statt Regeln (was soll rein)

Präskriptiv vs. Deskript

**Fragen / Diskussion /
Antworten ?**