
Multilinguality in Wikidata

Lucie-Aimée Kaffee
kaffee@soton.ac.uk



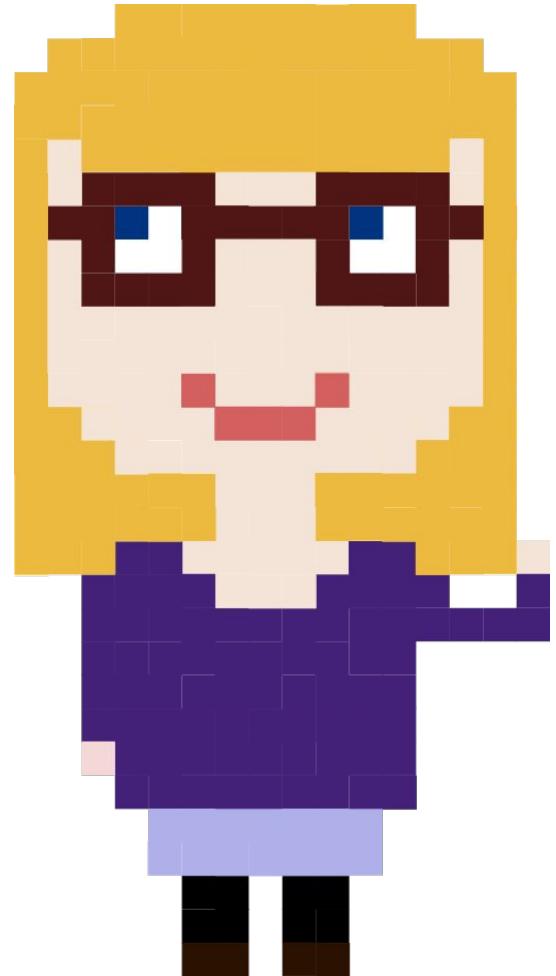
About Me

PhD Student
WAIS, University of Southampton

Previously worked as a Software
Developer at Wikimedia Deutschland, in
the Wikidata team

Interest in (under-resourced) languages

From Berlin, Germany



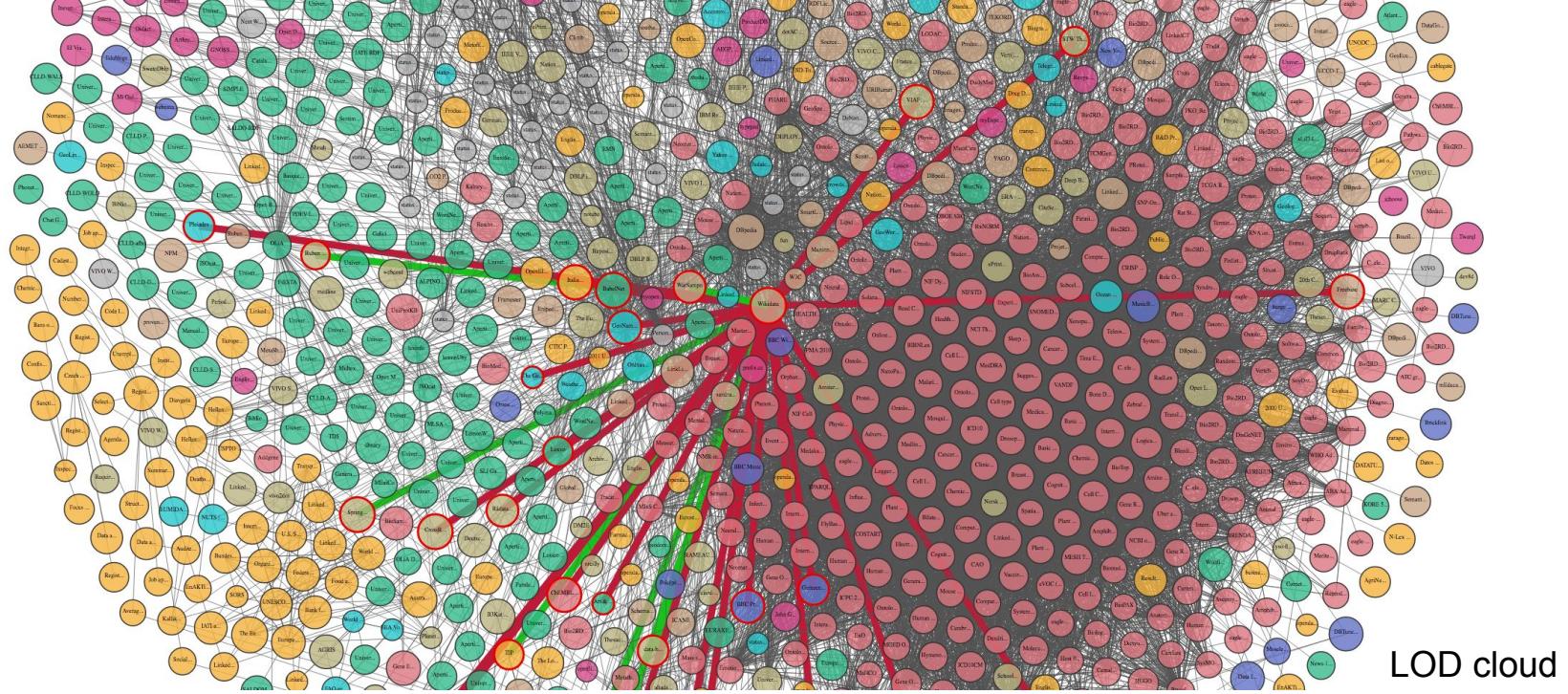


What we will talk about

Wikidata

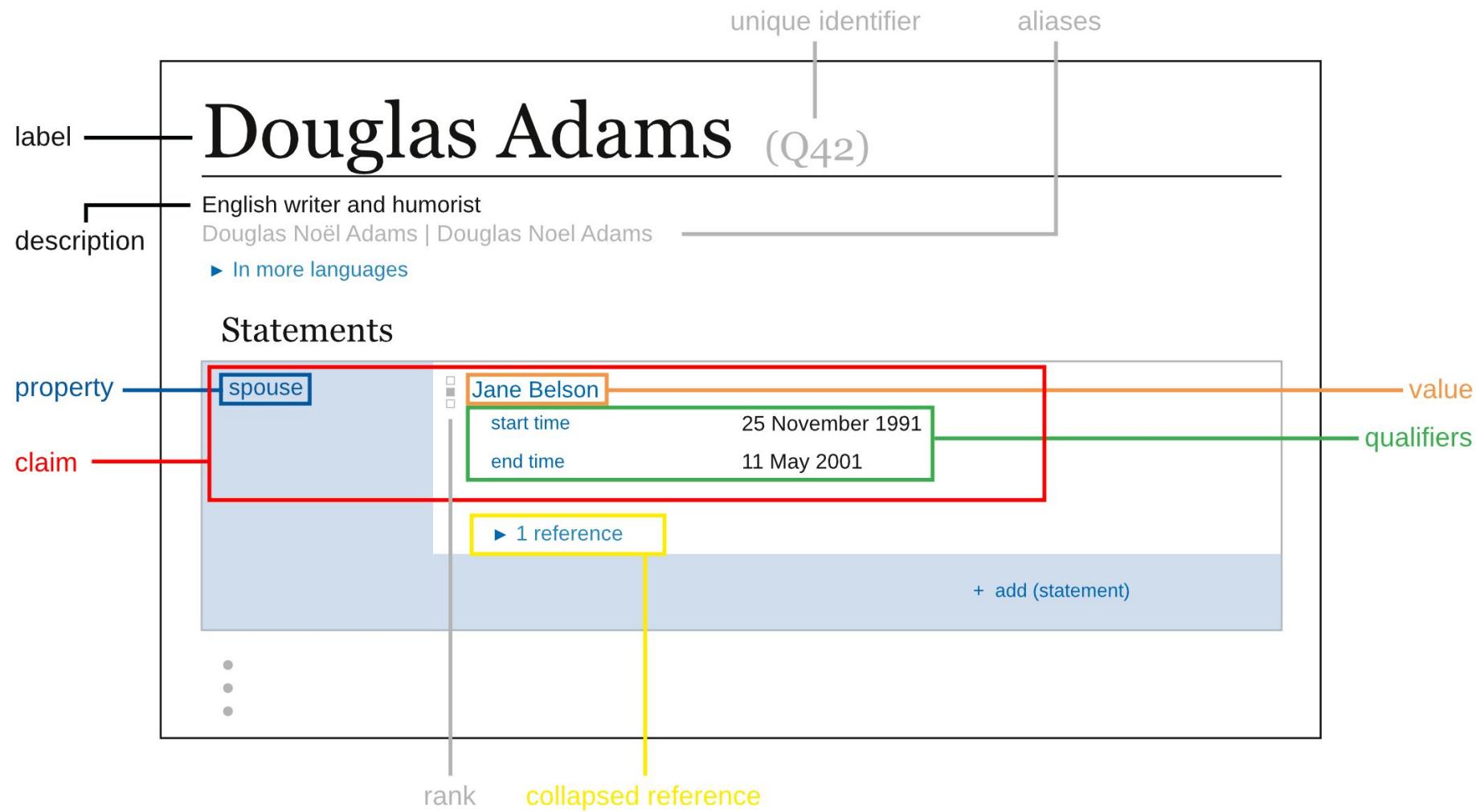
Multilinguality in Wikidata

My work



Wikidata

- Knowledge base maintained and edited by a community of users
 - 48,775,926 items
 - Each entity can have labels in >400 languages



Multilinguality in Wikidata

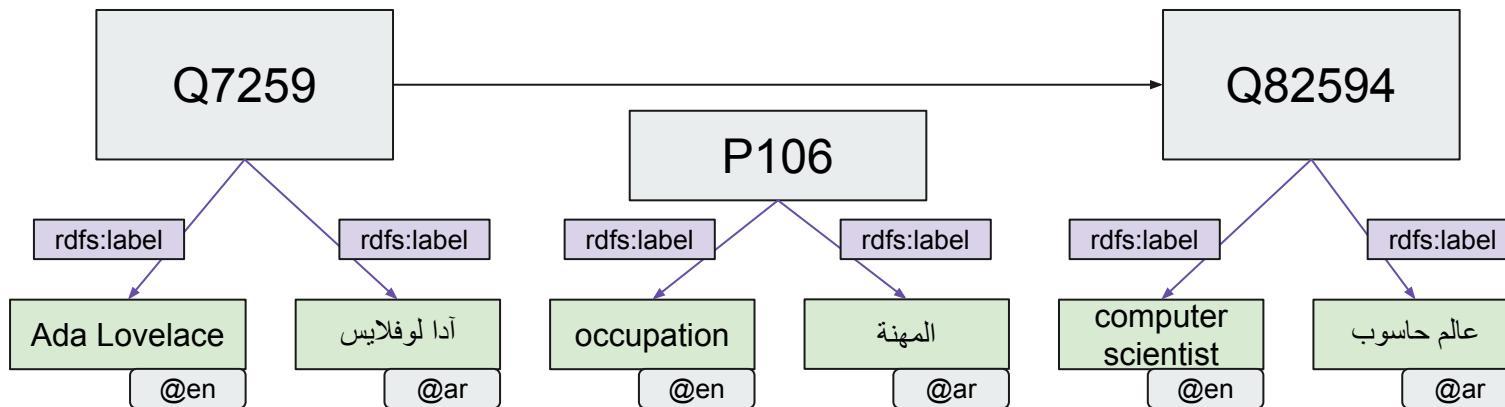
A Glimpse into Babel: An Analysis of Multilinguality in Wikidata

Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, Lydia Pintscher

OpenSym 2017



Multilinguality in Wikidata



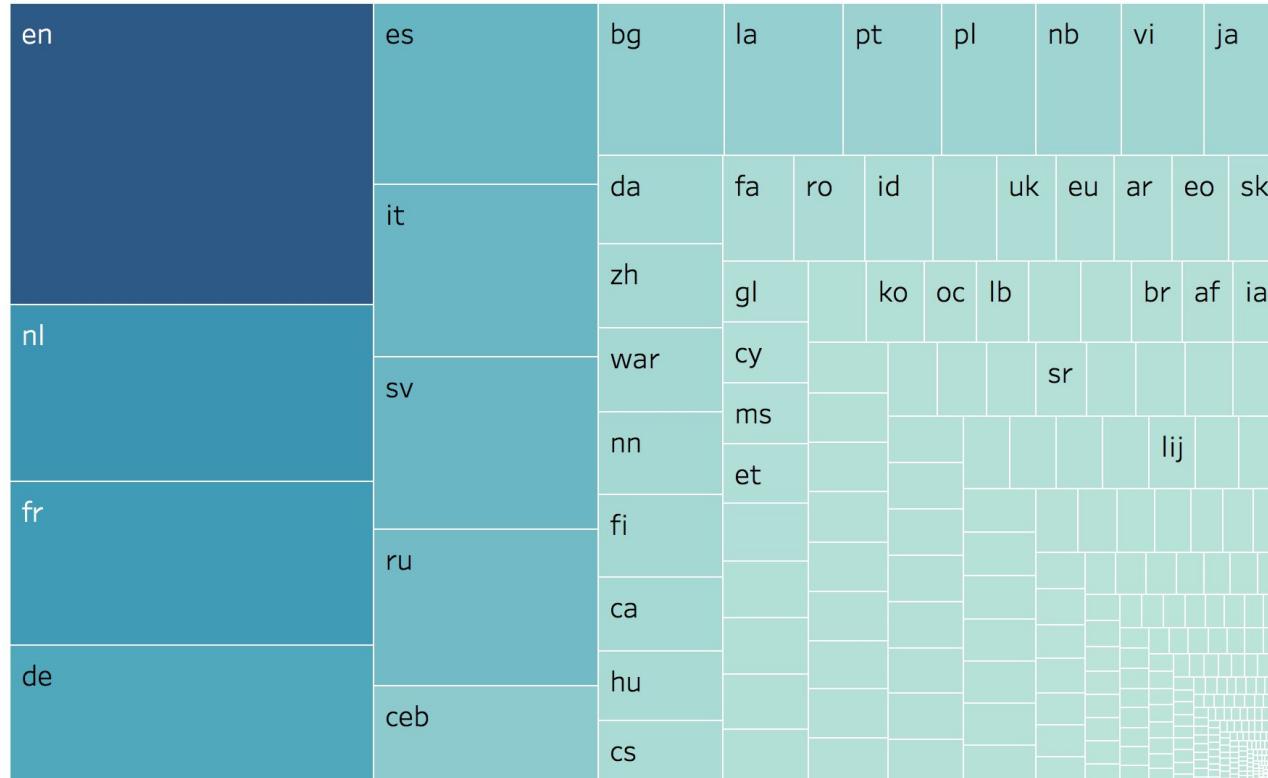
Multilinguality in Wikidata - Why do we care?

- Labels are the access point for humans
- Give language communities access to existing knowledge
- Central storage for translations for (under resourced) languages
- Semantic Web in NLP and NLG
- Reuse: Wikipedia, translation, question answering, chat bots, ...

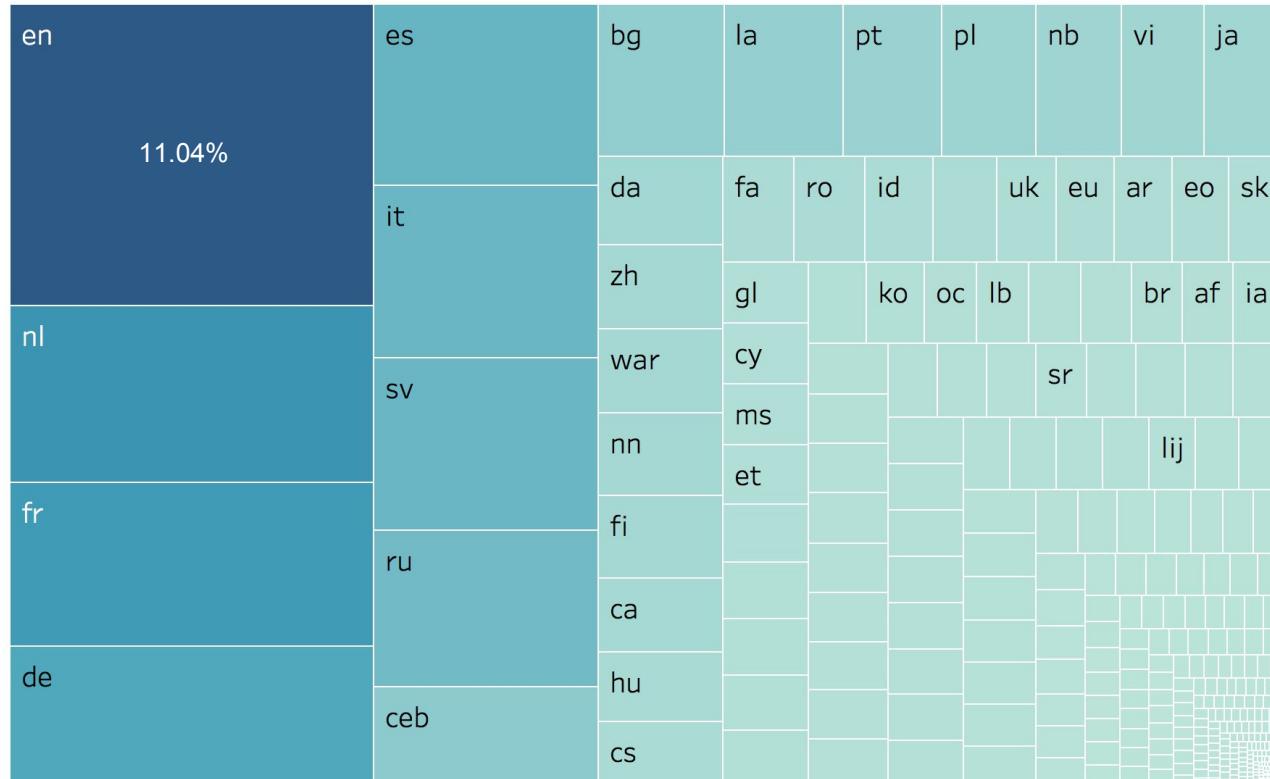
Research Questions

- What is the state of Wikidata with regard to multilinguality?
- How does Wikidata's label distribution relate to the real world and Wikipedia's language distribution?
- Is there a difference in the multilinguality of the properties, compared to the overall multilinguality of the knowledge base?

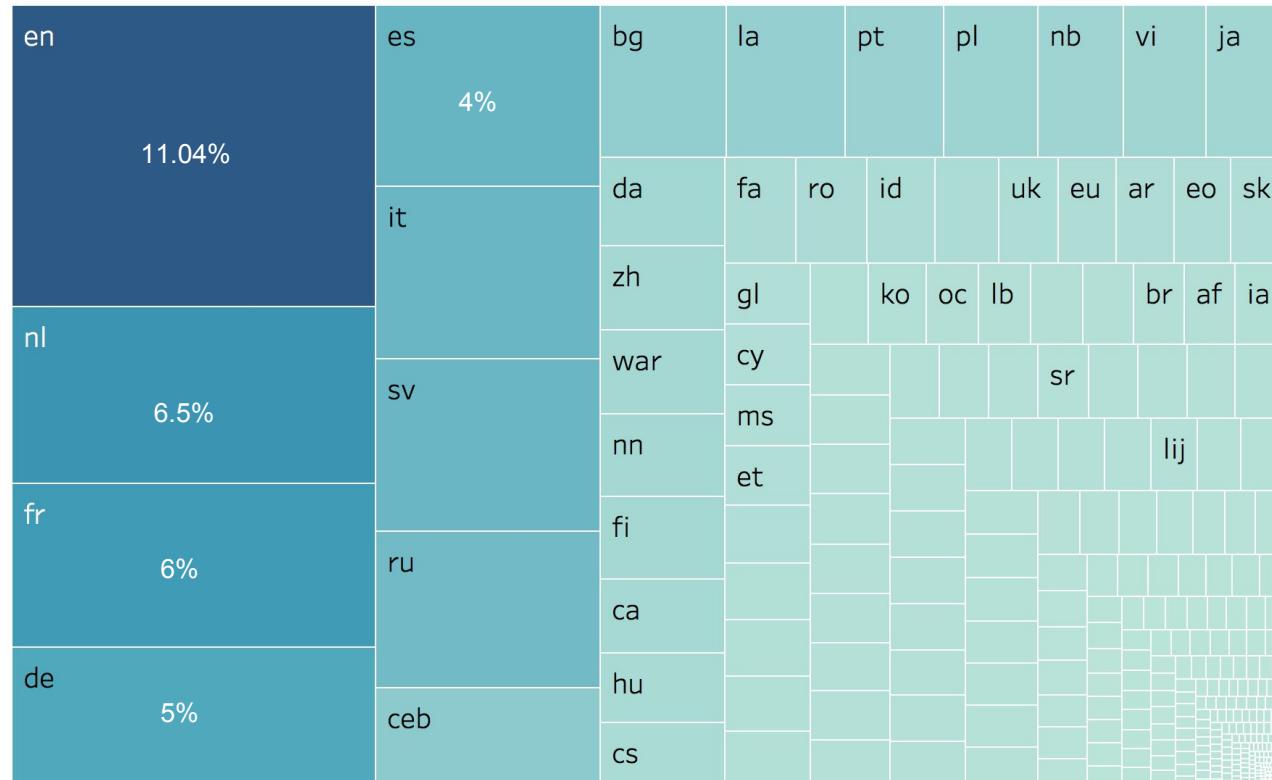
Wikidata

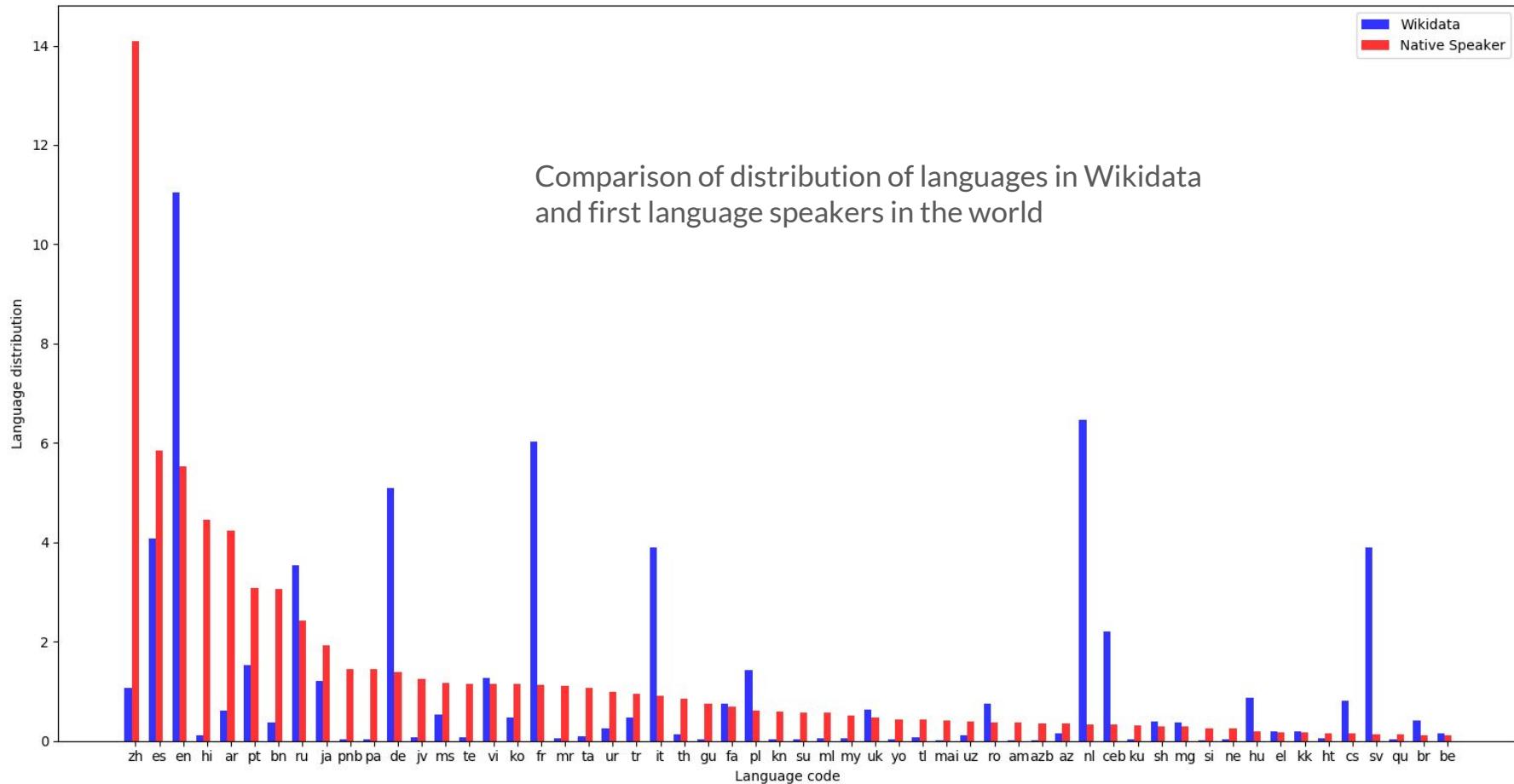


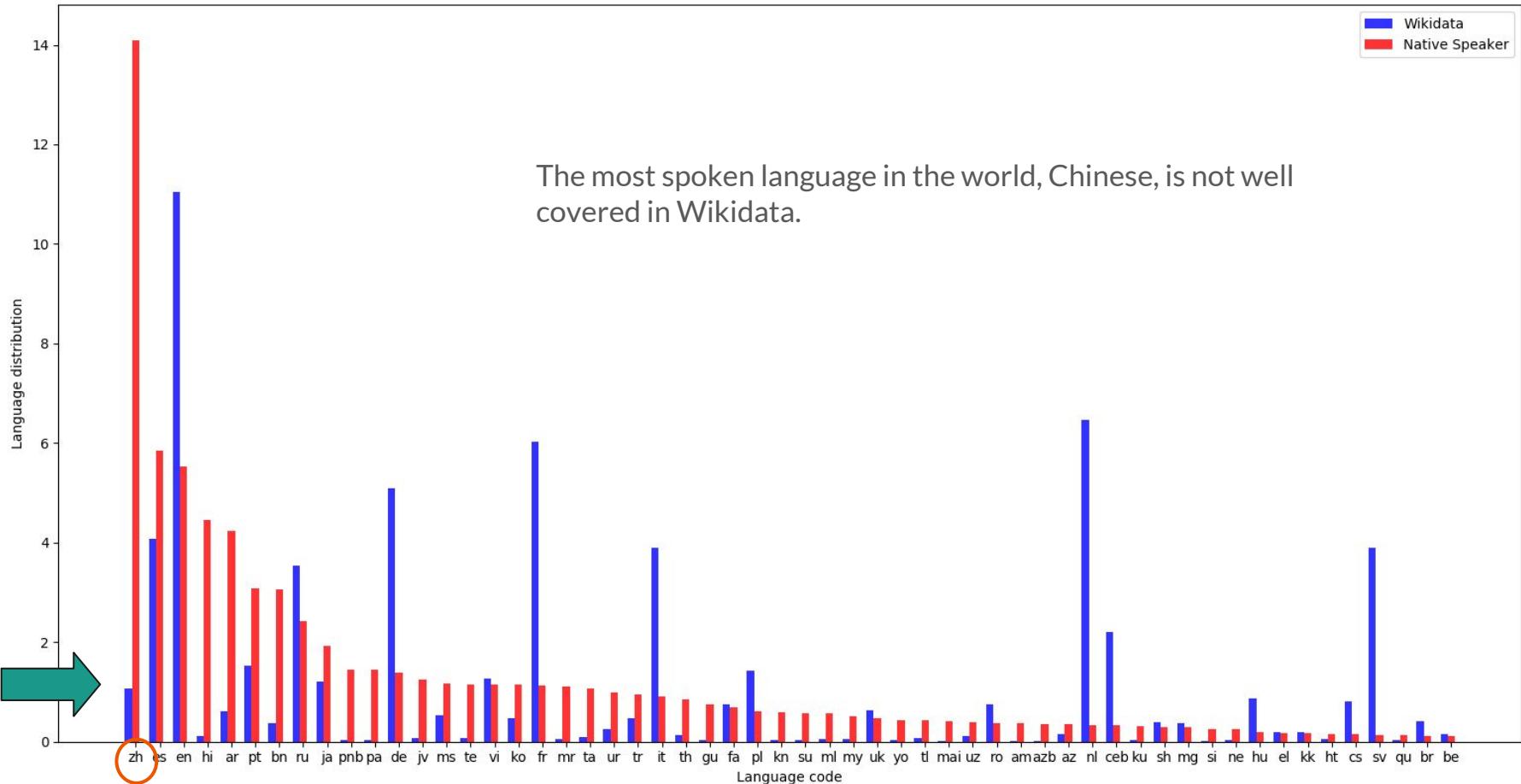
Wikidata



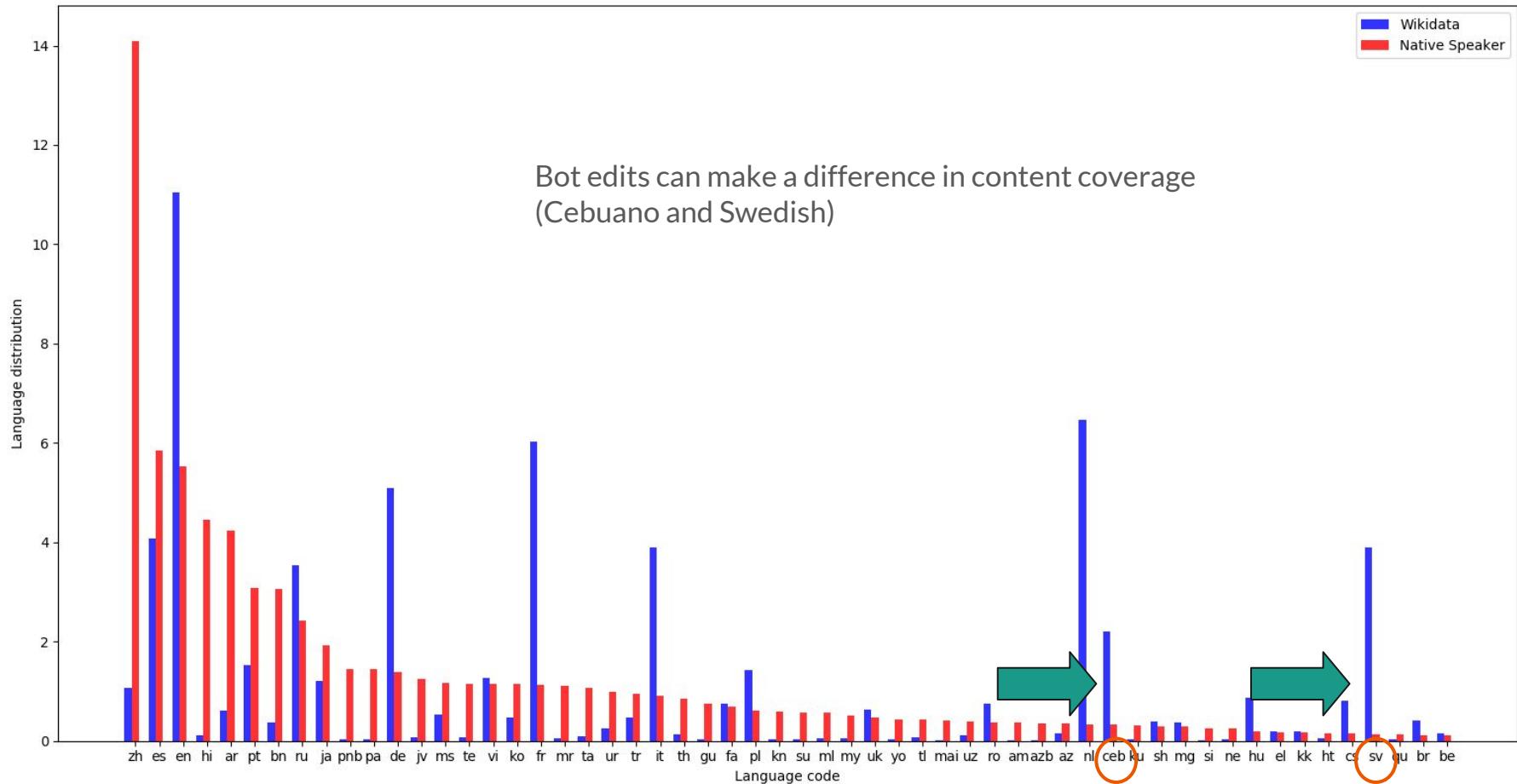
Wikidata



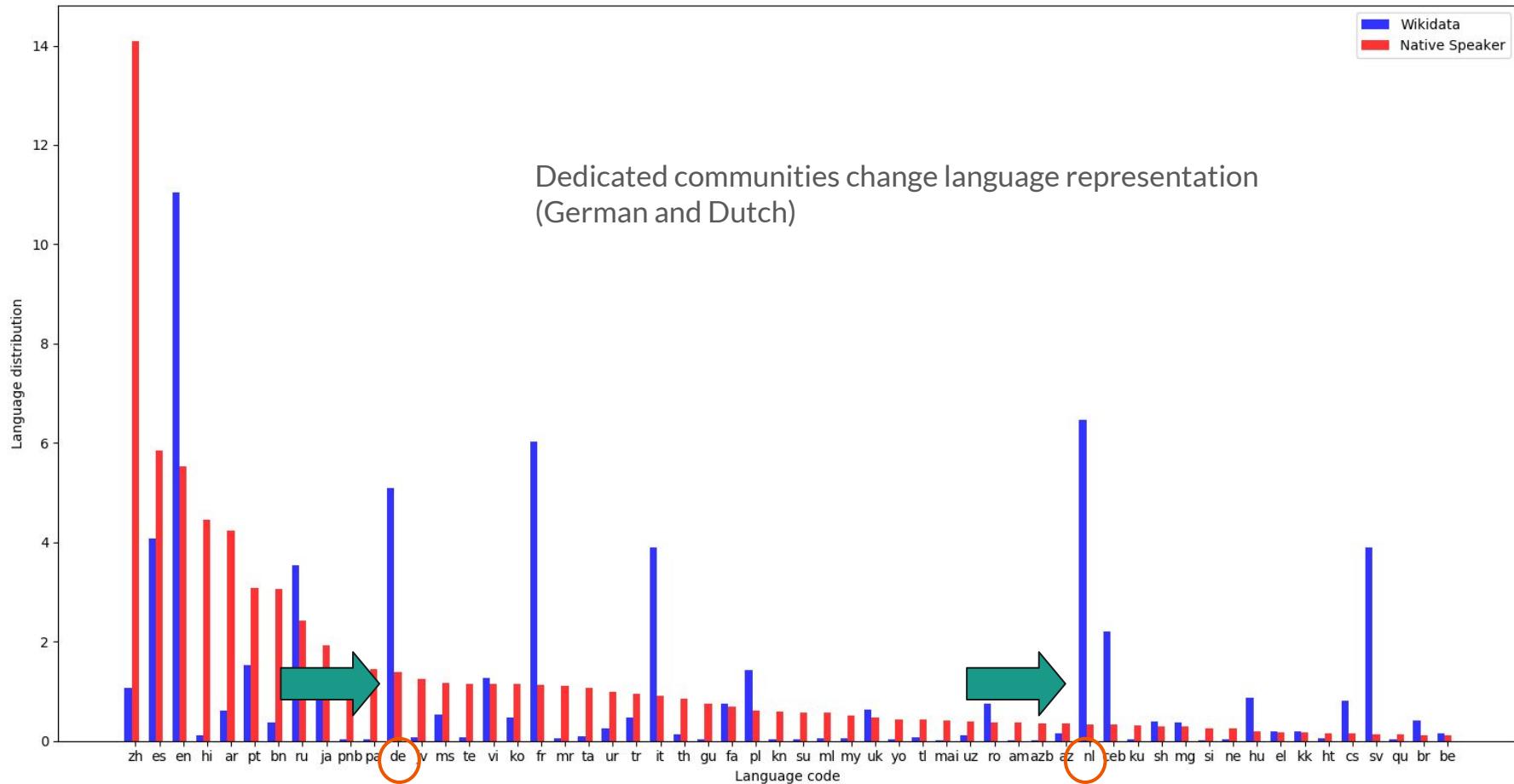




The most spoken language in the world, Chinese, is not well covered in Wikidata.

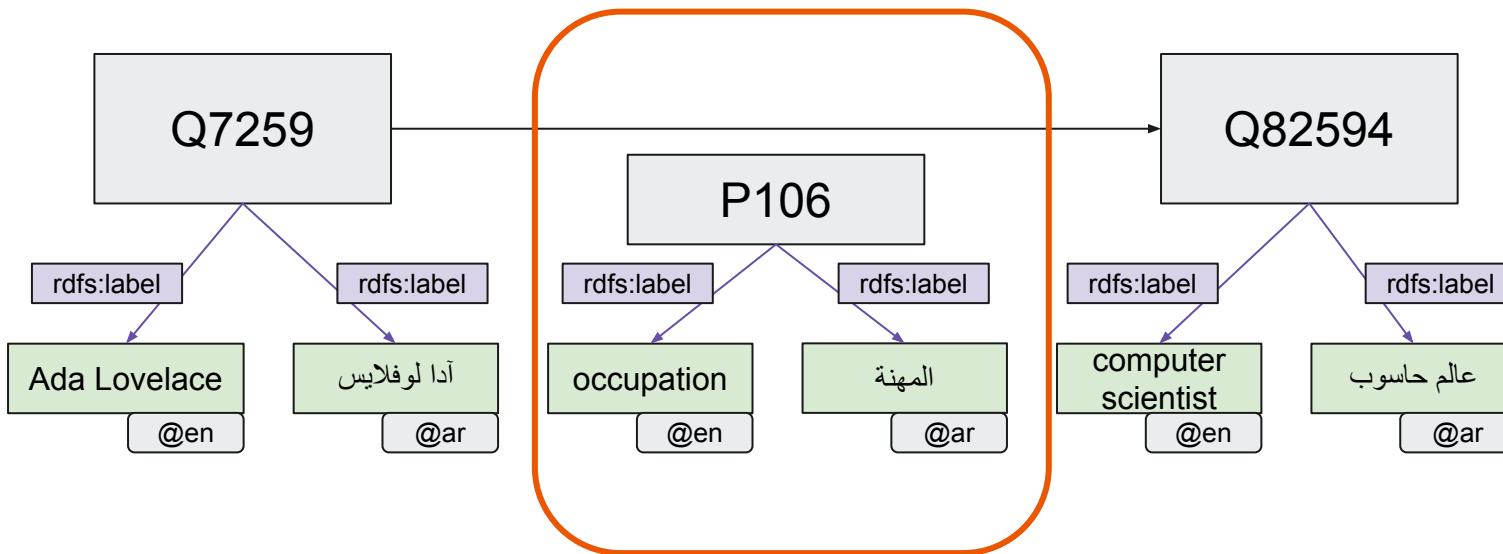


Bot edits can make a difference in content coverage
(Cebuano and Swedish)



Dedicated communities change language representation
(German and Dutch)

Wikidata Properties



Ranking of number of

Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh

Ranking of number of
Wikipedia articles by language,



Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh


Ranking of number of
Wikipedia articles by language,
all labels in Wikidata,

Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh


 Ranking of number of
 Wikipedia articles by language,
 all labels in Wikidata,
 and labels for properties in Wikidata

Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh


German is widely used in Wikipedia and Wikidata

High coverage through active community

Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	ml	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh

As German, active community that brings
high coverage of labels

Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fi	fi
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh



 High coverage in labels on Wikidata through high number of bot-imported Wikipedia articles, however low number of community edited properties

Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh

Even more extreme than Swedish: Not in
top 25 of community-edited properties by
language



Rank	Wikipedia	Wikidata	WD properties
1	en	en	en
2	ceb	nl	nl
3	sv	fr	fr
4	de	de	ru
5	nl	es	mk
6	fr	it	de
7	ru	sv	es
8	it	ru	pl
9	es	ceb	ca
10	war	bg	it
11	pl	la	sr
12	vi	pt	hu
13	ja	pl	pt
14	pt	nb	nb
15	zh	vi	ko
16	uk	ja	fa
17	ca	da	da
18	fa	zh	cs
19	ar	war	ja
20	no	nn	sv
21	sh	fi	be
22	fi	ca	el
23	hu	hu	ar
24	id	cs	uk
25	cs	fa	zh

Users in Wikidata

(Work in Progress)

Babel user information

de-N

Diese Benutzerin spricht Deutsch als Muttersprache.

en-4

This user has [near native speaker](#) knowledge of [English](#).

es-2

Esta usuaria tiene un conocimiento intermedio del [español](#).

tr-1

Bu kullanıcı [temel düzeyde Türkçe](#) bilir.

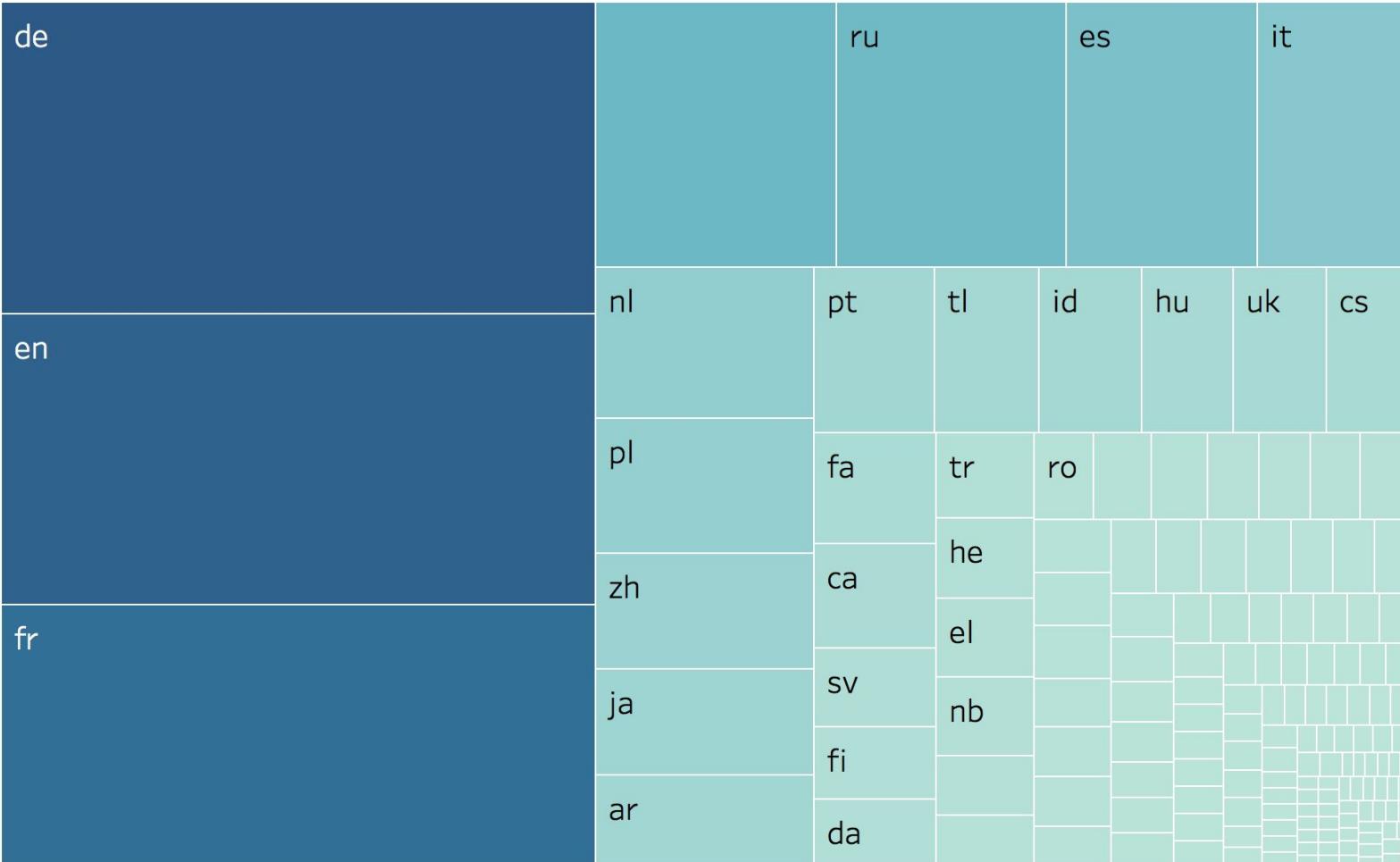
fr-1

Cette utilisatrice dispose de connaissances [de base en français](#).

ar-0

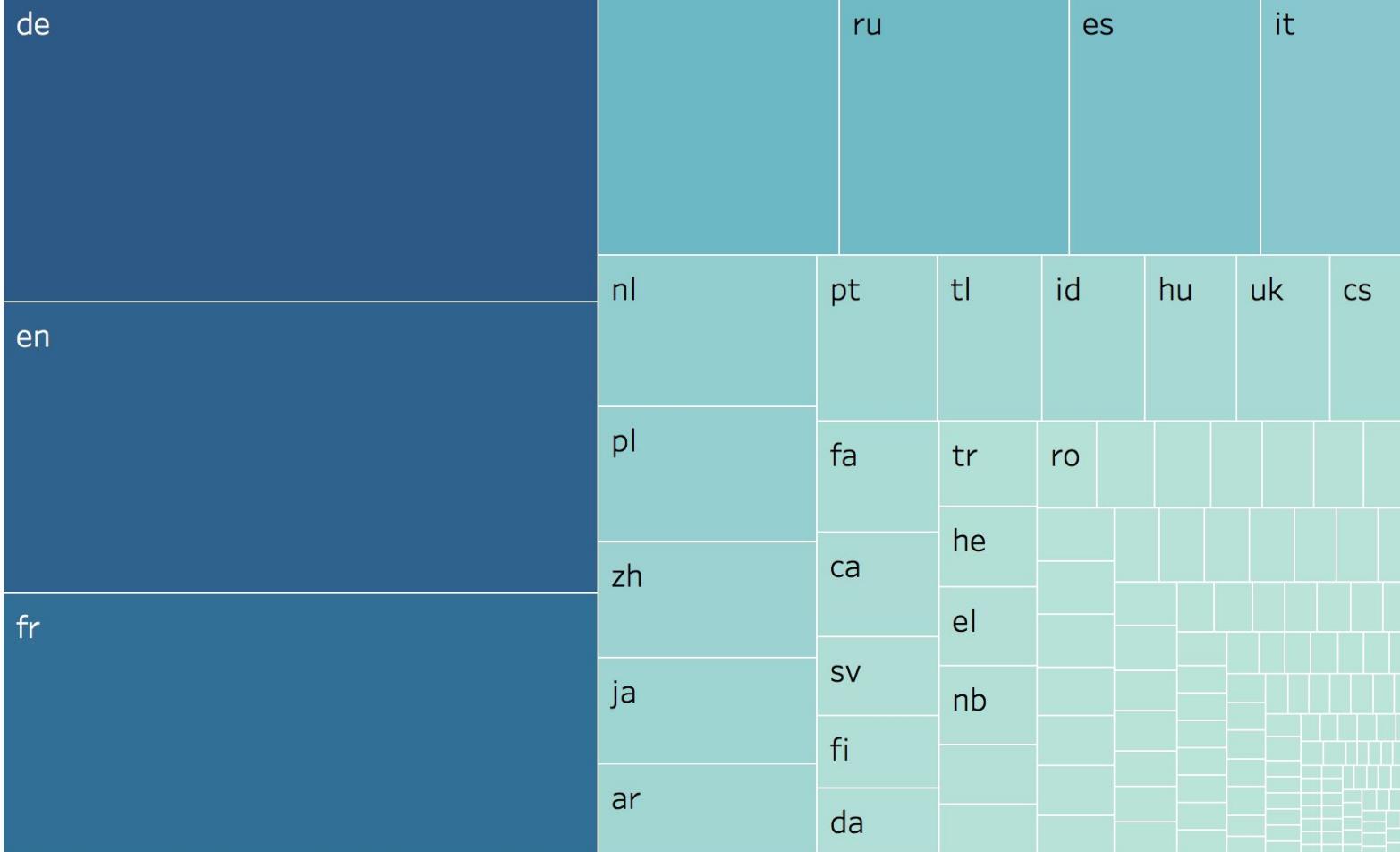
هذه المستخدمة [ليست](#) لديها معرفة [بالعربية](#) (أو تفهمها بصعوبة بالغة).

Users by language

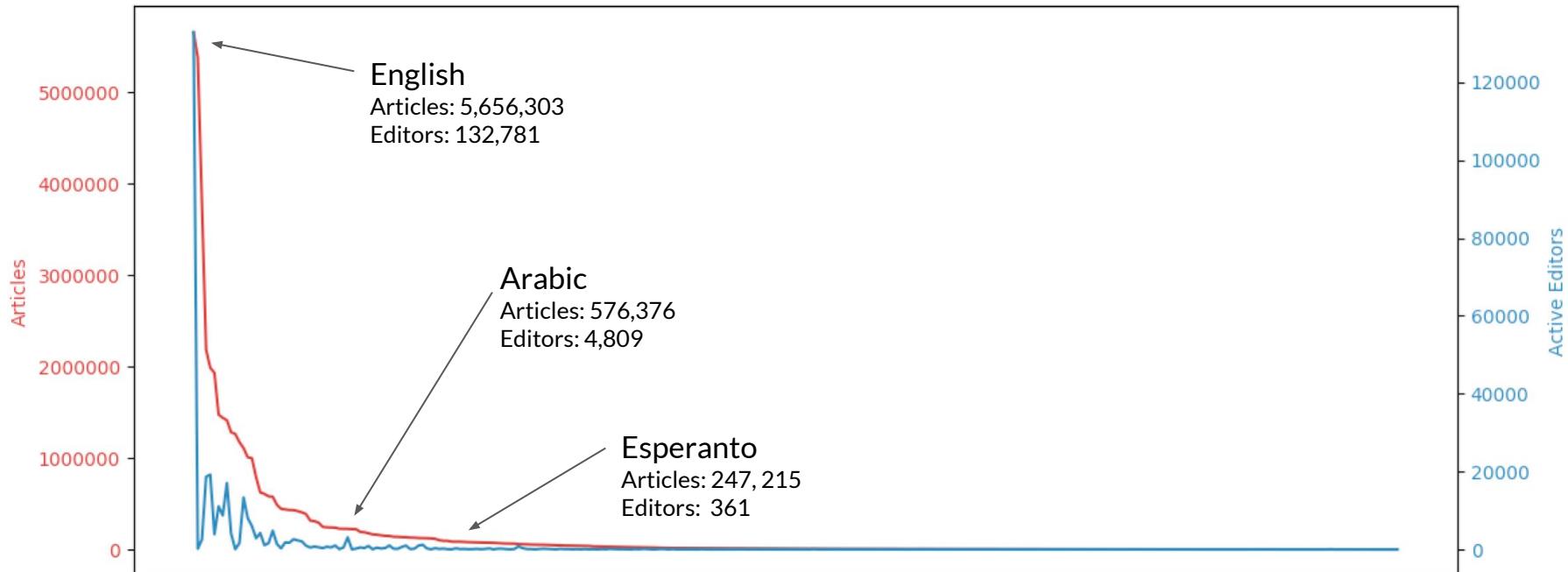


Native Languages of Wikidata users

Language coverage of labels does not reflect in languages Wikidata's users speak



From Wikidata to Wikipedia

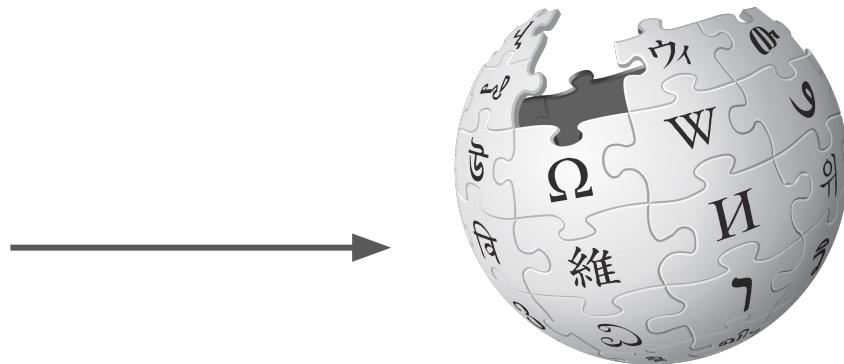


Wikipedia is available in 285 languages, but the content is unevenly distributed

Content From Wikidata to Wikipedia



WIKIDATA



WIKIPEDIA
The Free Encyclopedia

Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata

Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest,
Jonathon Hare, Elena Simperl

NAACL 2018

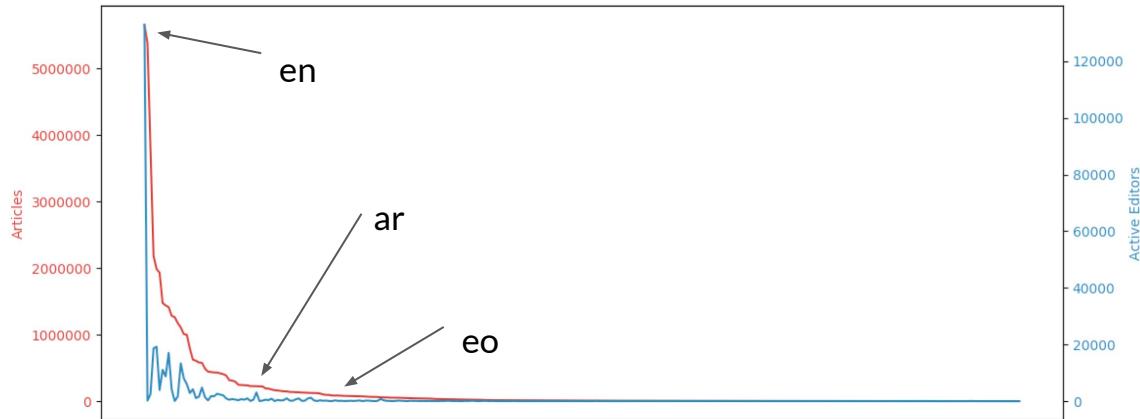


Esperanto

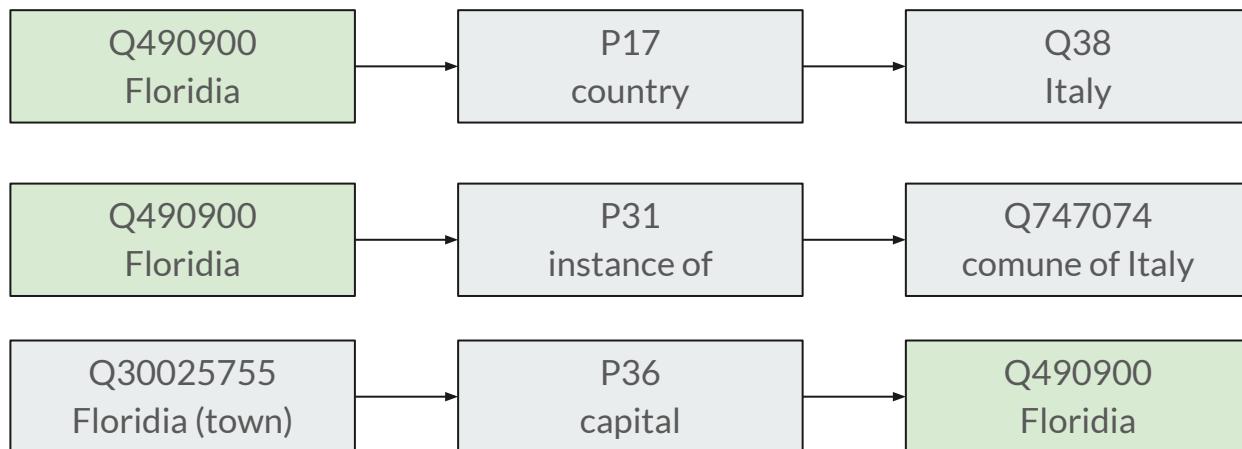
- Esperanto is an artificial language
- Easy to learn
- Engaged Wikipedia community
- A good starting point

Arabic

- Arabic is the 5th most spoken language in the world
- Content online in Arabic is sparse however



Sample Input



Neural Text Generation

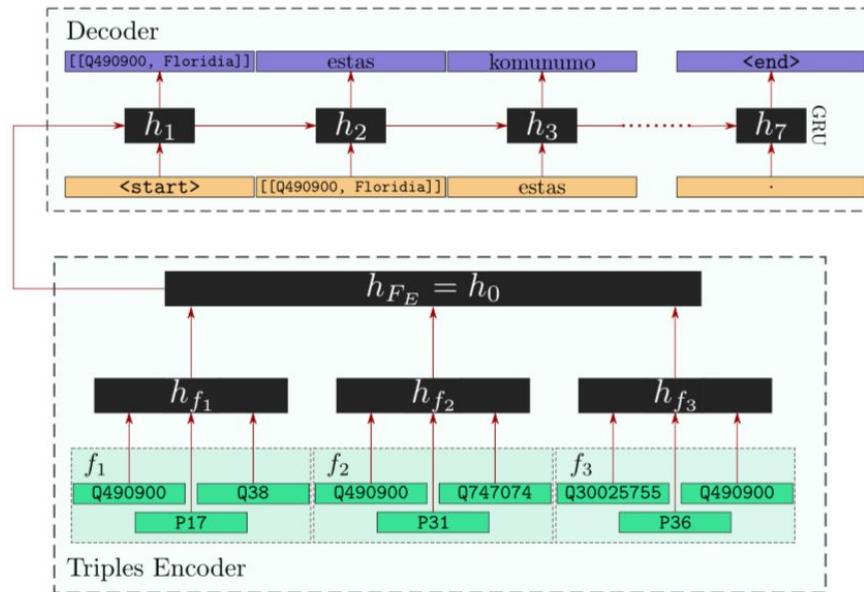
Arabic and Esperanto output text

Feed-forward architecture encodes Wikidata triples into vector of fixed dimensionality

RNN-based decoder generates text summaries, one token at a time

Property placeholder to deal with out of vocabulary words

Article-Placeholder Triples	$f_1 : \text{Q490900}$ (Floridia)	P17 (stato)	Q38 (Italio)
	$f_2 : \text{Q490900}$ (Floridia)	P31 (estas)	Q747074 (komunumo de Italio)
	$f_3 : \text{Q30025755}$ (Floridia)	P1376 (cefurbo de)	Q490900 (Floridia)
Textual Summary	Floridia estas komunumo de Italio.		
Vocab. Extended Summary	[[Q490900, Floridia]] estas komunumo de [[P17]].		





Q106693 Group 14 (chemical series)

مجموعة الكربون هي العناصر الموجودة الموجودة في الجدول الدوري للعناصر

Karbongrupo estas elemento en grupo 0 de la perioda tabelo laŭ la IUPAC-sistemo .

The carbon group is a periodic table group consisting of carbon, silicon, germanium, tin, lead, and flerovium.

Q16885 Thelxinoe (natural satellite)

ثيلكسيون هو قمر طبيعي غير نظامي يتحرك بحركة تراجعية تابع للكوكب المشتري .

Telksino estas neregula satelito de Jupitero , kiu havas retrogradan orbiton .

Thelxinoe (/θɛlk 'sɪnəʊ, i:/ /θɛlk-SIN-o-ee; Greek: Θελξινόη), also known as Jupiter XLII, is a natural satellite of Jupiter.

Automatic Evaluation

- Baselines
 - Machine Translation, Information Retrieval Based, Kneser-Ney
- Automatic Evaluation
 - BLEU 1, BLEU 2, BLEU 3, BLEU 4, METEOR, ROUGE

Model	BLEU 1		BLEU 2		BLEU 3		BLEU 4		ROUGE _L		METEOR		
	Valid.	Test	Valid.	Test	Valid.	Test	Valid.	Test	Valid.	Test	Valid.	Test	
Arabic	MT	31.12	33.48	19.31	21.12	12.69	13.89	8.49	9.11	29.96	30.51	31.05	30.1
	TP	41.39	41.73	34.18	34.58	29.36	29.72	25.68	25.98	43.26	43.58	32.99	33.33
	TP _{ext}	49.87	48.96	42.44	41.5	37.29	36.41	33.27	32.51	51.66	50.57	34.39	34.25
	Ours	53.18	52.94	45.86	45.64	40.38	40.21	35.7	35.55	57.9	57.99	39.22	39.37
Esperanto	MT	5.35	5.47	1.62	1.62	0.59	0.56	0.26	0.23	4.67	4.79	0.66	0.68
	TP	43.01	42.61	33.67	33.46	28.16	28.07	24.35	24.3	46.75	45.92	20.71	20.46
	TP _{ext}	52.75	51.66	43.57	42.53	37.53	36.54	33.35	32.41	58.15	57.62	31.21	31.04
	Ours	56.51	56.96	47.72	48.1	41.8	42.13	37.24	37.52	64.36	64.69	28.35	28.76

Results of the automatic evaluation: Our network outperforms all baselines

Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for ArticlePlaceholders

Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest,
Jonathon Hare, Elena Simperl

ESWC 2018



ArticlePlaceholder display

Wikidata triples on Wikipedia in tabular way

Currently deployed on 14
Wikipedias

Espesyal

Please contribute to the Haitian Creole Wikipedia! Am mèt tèt anseman pou Wikipédia Kreyòl Ayisyen mache douyan! Si vous plait... contribuez à Wikipédia en créole haïtien! (Por favor contribuya a la Wikipédia haitiana!)

Triceratops

dinosaure herbivore de la famille des ceratopidae

egzamp nan taxon fossile	nom scientifique du taxon Triceratops auteur taxinomique: Othniel Charles Marsh date de description scientifique: 1889	rang taxinomique genre
taxon supérieur Ceratopidae	nom vernaculaire Triceratops Trikephron 三角龍	période de disparition Maastrichtien
période d'apparition Maastrichtien	hauteur 3 Mèt (mezi)	masse 8 500 Kilogram
longueur 9 Mèt (mezi)	galerie Commons Triceratops	catégorie Commons Triceratops

Legendre de média:  (Triceratops prorsus) (Photo: A. S. L. Gishlick, Specimen at the Natural History Museum of Los Angeles)

Ressources extérieures

1. bibliothèque du Congrès, 15 janvier 2018
2. 1 sauvegarde de la base de données Freebase, 28 octobre 2013
3. Fossilworks, 13 mai 2015
4. 1 Interim Register of Marine and Nonmarine Genera, 19 avril 2018
5. 1 Encyclopédie de la Vie, 30 octobre 2014
6. 1 Répertoire d'autorité matière encyclopédique et alphabétique unifiée, 15 janvier 2018
7. 1 Système Mondial d'Information sur la Biodiversité, 10 décembre 2015



Please contribute to the Haitian Creole Wikipedia! Ann mêt tôt ansanm pou Wikipèdia Kreyòl Ayisyen mache douvan! Si vous plait... contribuez à Wikipédia en créole haïtien ! Por favor contribuya a la Wikipedia haitiana!

[masquer]

Triceratops

dinosaure herbivore de la famille des cératopsidés

egzanz nan
taxon fossile

nom scientifique du taxon
Triceratops

auteur taxonomique: Othniel Charles Marsh
date de description scientifique: 1889

rang taxinomique
genre

taxon supérieur
Ceratopsinae

nom vernaculaire
الترicerاتوبس
טריצרטופס
Трицератопс
三角龍

période de disparition
Maastrichtien

période d'apparition
Maastrichtien

hauteur
3 Mèt (mezi)

masse
8 500 Kilogram

longueur
9 Mèt (mezi)

galerie Commons
Triceratops

catégorie Commons
Triceratops



هذا عظمي لترiceratops الكامل (Triceratops prorsus) يُعرض في متحف للتاريخ الطبيعي بلوس أنجلوس، Skeletal mount of a T. prorsus specimen at the Natural History Museum of Los Angeles

Resous ekstèn

identifiant bibliothèque du Congrès sh880002 54^[1]

identifiant Encyclopædia Britannica animal/Tri en ligne ceratops

identifiant de la Grande Encyclopédie russe en ligne 4204642

identifiant Freebase /m/0f6mr [2]

identifiant Fossilworks 38862^[3]

identifiant Encyclopædia Universalis triceratop en ligne s

identifiant Gemeinsame Normdatei 4331835-6

identifiant Quora d'un sujet Tricerato

Referans

1. ↑ bibliothèque du Congrès, 15 janvye 2018
2. ↑ sauvegarde de la base de données Freebase, 28 oktiòb 2013
3. ↑ Fossilworks, 13 me 2015
4. ↑ Interim Register of Marine and Nonmarine Genera, 19 avril 2018



Please contribute to the Haitian Creole Wikipedia! Ann mêt tôt ansanm pou Wikipèdia Kreyòl Ayisyen mache douvan! Si vous plait... contribuez à Wikipédia en créole haïtien ! Por favor contribuya a la Wikipedia haitiana!

[masquer]

Triceratops

dinosaure herbivore de la famille des cérapotidés

egzanp nan
taxon fossile

nom scientifique du taxon
Triceratops

auteur taxonomique: Othniel Charles Marsh
date de description scientifique: 1889

rang taxinomique
genre



هذا عظمي لترiceratops الكامل: ترiceratops prorsus (Triceratops prorsus) في متحف للتاريخ الطبيعي بلوس أنجلوس، Skeletal mount of a T. prorsus specimen at the Natural History Museum of Los Angeles

taxon supérieur
Ceratopsinae

nom vernaculaire
الترicerاتوبس
טריצרטופס
Трицератопс
三角龍

période de disparition
Maastrichtien

période d'apparition
Maastrichtien

hauteur
3 Mèt (mezi)

masse
8 500 Kilogram

longueur
9 Mèt (mezi)

galerie Commons
Triceratops

catégorie Commons
Triceratops

Resous ekstèn

identifiant bibliothèque du Congrès sh880002 54^[1]

identifiant Encyclopædia Britannica animal/Tri en ligne ceratops

identifiant de la Grande Encyclopédie russe en ligne 4204642

identifiant Freebase /m/0f6mr [2]

identifiant Fossilworks 38862^[3]

identifiant Encyclopædia Universalis triceratop en ligne s

identifiant Gemeinsame Normdatei 4331835-6

identifiant Quora d'un sujet Tricerato

Referans

1. ↑ bibliothèque du Congrès, 15 janvye 2018
2. ↑ sauvegarde de la base de données Freebase, 28 oktiòb 2013
3. ↑ Fossilworks, 13 me 2015
4. ↑ Interim Register of Marine and Nonmarine Genera, 19 avril 2018

Enriching ArticlePlaceholder with textual summaries generated from Wikidata triples

Working with Arabic and Esperanto

Espeyal

Please contribute to the Haitian Creole Wikipedia! A mi tanzam pou Wikipédia Kreyòl Ayisyen mache douyan! Si vous plait... contribuez à Wikipédia en créole haïtien! (Por favor contribuya a la Wikipédia haitiana!)

Triceratops

déssouise herbivore de la famille des ceratopidae

egzamp nan
taxon fossile

nom scientifique du taxon
Triceratops
auteur taxonomique: Othniel Charles Marsh
date de description scientifique: 1890

rang taxinomique
genre

taxon supérieur
Ceratopidae

nom vernaculaire
Triceratops
الثريkeratosaurus
Triceratops
三角龍

période de disparition
Maastrichtien

période d'apparition
Maastrichtien

hauteur
3 Mèt (mezi)

masse
8 500 Kilogram

longueur
9 Mèt (mezi)

galerie Commons
Triceratops

catégorie Commons
Triceratops

Legend de media: (Triceratops prorsus) دریافت از موزه تاریخ طبیعی لوس آنجلس. ۱ گروه اسپرانتو at the Natural History Museum of Los Angeles

Ressous ekstén

identifiant bibliothèque du Congrès: sh80002
54[1]

identifiant Encyclopædia Britannica en ligne: animalisti

identifiant de la Grande Encyclopédie russe en ligne: 4204642

identifiant sur Freebase: /m/06smr
[2]

identifiant FossiWards: 36692[2]

identifiant Encyclopædia Universalis: triceratop

identifiant Gemeente Normdatel: 4321835-6

identifiant Quora d'un sujet: Tricerato
ps

identifiant Interim Register of Marine and Normative Genera: 1023470 [4]

References

1. bibliothèque du Congrès, 15 janvye 2018
2. 1 stevengard de la base de données Freebase, 28 oktobr 2013
3. Fossilworks, 13 me 2015
4. 1 Interim Register of Marine and Normative Genera, 19 avil 2018
5. 1 Encyclopædia de la Vie, 30 oktobr 2014
6. 1 Encyclopædia d'autorité matrice encyclopédique et alphabétique unifiée, 15 janvye 2018
7. 1 Système Mondial d'Information sur la Biodiversité, 10 desemv 2018

Community Study

- Two 15 days online surveys, aimed at readers and editors in Esperanto and Arabic
- Aiming to test our work with the actual Wikipedia community, outreach on Wikipedia platforms
- Reader:
 - ◆ Fluency: Is the text understandable and grammatically correct?
 - ◆ Appropriateness: Does the summary 'feel' like a Wikipedia article?
- Editor:
 - ◆ Editors were asked to edit the article starting from our summary (2-3 sentences)
 - ◆ How much of the text was reused?



		Fluency	Appropriateness	
		Mean SD		Part of Wikipedia
Arabic	Ours	4.7 1.2	77%	
	Wikipedia	4.6 0.9	74%	
	News	5.3 0.4	35%	
Esper.	Ours	4.5 1.5	69%	
	Wikipedia	4.9 1.2	84%	
	News	4.2 1.2	52%	

Results of the reader study

		Fluency		Appropriateness	
		Mean	SD	Part of Wikipedia	
Arabic	Ours	4.7	1.2	77%	
	Wikipedia	4.6	0.9	74%	
	News	5.3	0.4	35%	
Esper.	Ours	4.5	1.5	69%	
	Wikipedia	4.9	1.2	84%	
	News	4.2	1.2	52%	

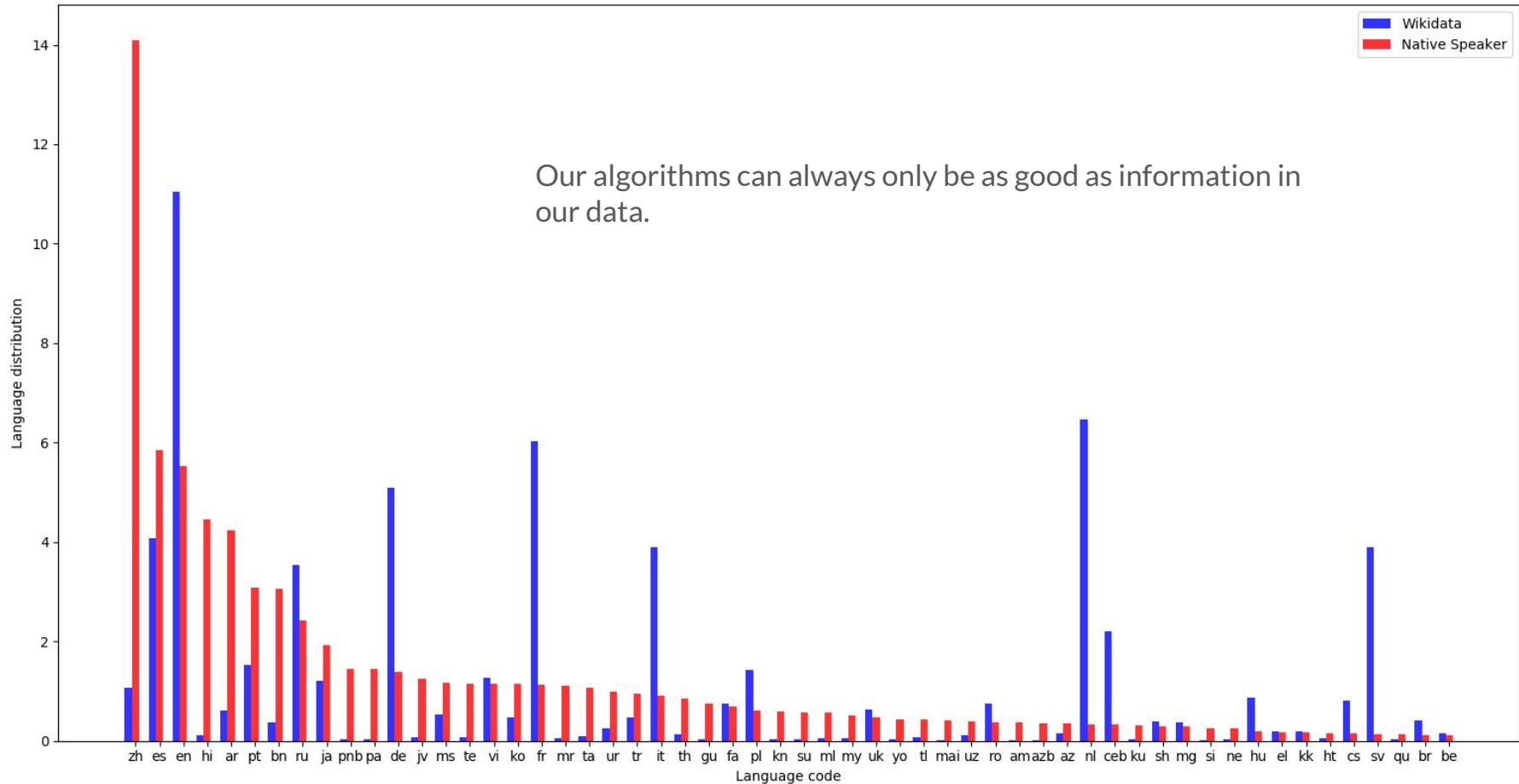
Results of the reader study: We generate sentences of comparable fluency, that “feel” like Wikipedia sentences

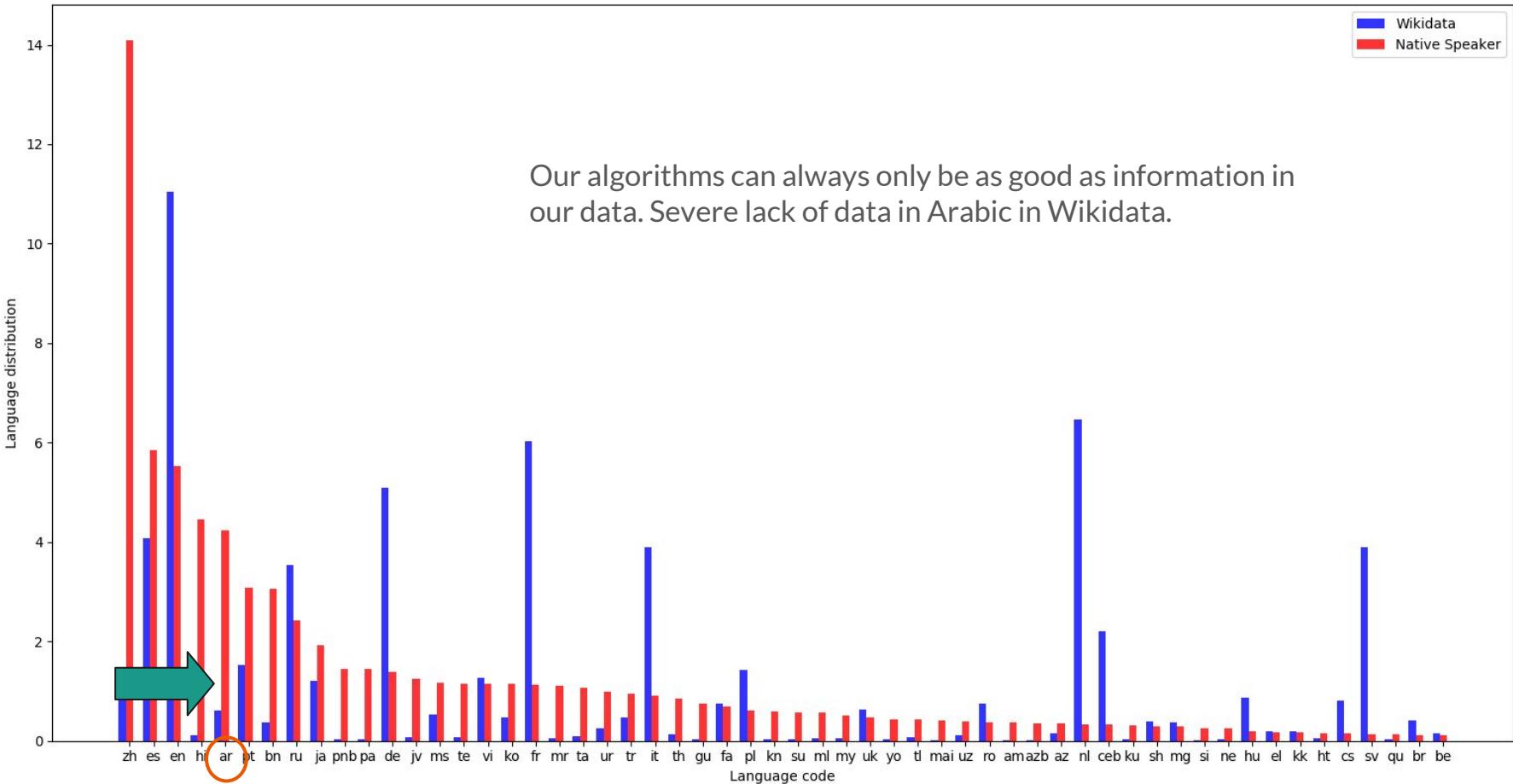
	Category	Examples	%
Arabic	wholly derived	خامسي كلوريد الزرنيخ مركب كيميائي له الصيغة (كلمة ناقصة)، ويكون على شكل بلورات بيضاء. B A	
	partially derived	خامسي كلوريد الزرنيخ هو مركب كيميائي له الصيغة (AtClu2085)، ويكون على شكل بلورات بيضاء. B A	45.45%
	non derived	بيتش باتوم (بالإنجليزية (كلمة ناقصة) Ohio (هي منطقة سكنية تقع في الولايات المتحدة في (كلمة ناقصة). C D بيتش باتوم (بالإنجليزية: Beach Batom) هي قرية تقع في الولايات المتحدة الأمريكية في برووك كاونتي. C D E	33.33%
Esperanto	wholly derived	Zederik estas komunumo en la nederlanda provinco Zuid-Holland. Zederik estas komunumo en la nederlanda provinco Zuid-Hooland kaj estas ĉirkaŭata de la municipoj Lopik kaj Zederik.	78.98%
	partially derived	Nova Pádua estas municipio en la brazila subŝtato Suda Rio-Grando, kiu havis (manka nombro) loĝantojn en (jaro). Nova Pádua estas municipio en la brazila subŝtato Suda Rio-Grando.	15.79%
	non derived	Ibiúna estas municipio de la brazila subŝtato San-Paúlio, kiu taksis (manka nombro) enloĝantojn en (jaro). Ibiúna estas brazila [[municipio]] kiu troviĝas en la administra unuo [[San-Paúlo]].	5.26%

Results of the editor study: We generate sentences that are highly reused by editors

	Category	Examples	%
Arabic	wholly derived	خامسي كلوريد الزرنيخ مركب كيميائي له الصيغة (كلمة ناقصة)، ويكون على شكل بلورات بيضاء. B A	45.45%
	partially derived	خامسي كلوريد الزرنيخ هو مركب كيميائي له الصيغة (AtClu2085)، ويكون على شكل بلورات بيضاء. B A	33.33%
	non derived	بيتش ياتوم (بالإنجليزية (كلمة ناقصة) Ohio (هي منطقة سكنية تقع في الولايات المتحدة في (كلمة ناقصة). C D E بيتش ياتوم (بالإنجليزية: Beach Batom) هي قرية تقع في الولايات المتحدة الأمريكية في برووك كاونتي. C D E F	21.21%
Esperanto	wholly derived	Zederik estas komunumo en la nederlanda provinco Zuid-Holland. Zederik estas komunumo en la nederlanda provinco Zuid-Hooland kaj estas ĉirkaŭata de la municipoj Lopik kaj Zederik.	78.98%
	partially derived	Nova Pádua estas municipo en la brazila subŝtato Suda Rio-Grando, kiu havis (manka nombro) loĝantojn en (jaro). Nova Pádua estas municipo en la brazila subŝtato Suda Rio-Grando.	15.79%
	non derived	Ibiúna estas municipo de la brazila subŝtato San-Paúlio, kiu taksis (manka nombro) enloĝantojn en (jaro). Ibiúna estas brazila [[municipo]] kiu troviĝas en la administra unuo [[San-Paúlo]].	5.26%

Results of the editor study: We generate sentences that are highly reused by editors

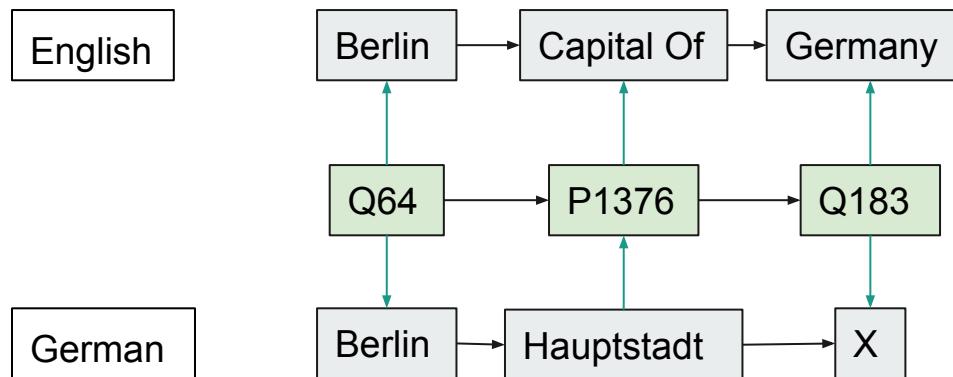




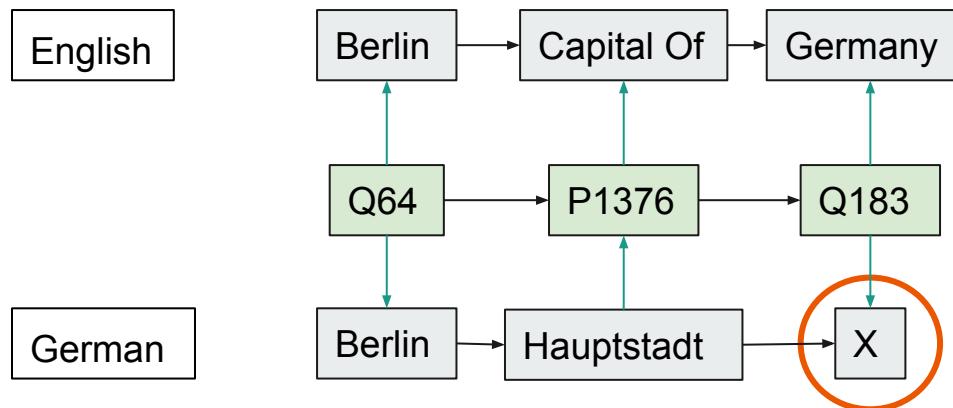
Our algorithms can always only be as good as information in our data. Severe lack of data in Arabic in Wikidata.

Future Work: Label Extraction From Wikipedia For Wikidata

Example Wikidata Triple



Example Wikidata Triple





WIKIPEDIA
Die freie Enzyklopädie

Hauptseite
Themenportale
Von A bis Z
Zufälliger Artikel

Mitmachen
Artikel verbessern
Neuen Artikel anlegen
Autorenportal
Hilfe
Letzte Änderungen
Kontakt
Spenden

Werkzeuge

Links auf diese Seite
Änderungen an verlinkten Seiten
Datei hochladen
Spezialseiten
Permanenter Link
Seiteninformationen
Wikidata-Datenobjekt
Artikel zitieren

Drucken/exportieren

Buch erstellen
Als PDF herunterladen
Druckversion

In anderen Projekten

Commons
Wikinews
Wikiquote
Wikisource
★ Wikivoyage

In anderen Sprachen

Artikel Diskussion

Lesen Bearbeiten Quelltext bearbeiten Versionsgeschichte Mehr Wikipedia durchsuchen Koordinaten: 52° 31' N, 13° 24' O I B

Berlin

Dieser Artikel befasst sich mit der deutschen Hauptstadt. Zu weiteren Bedeutungen siehe Berlin (Begriffsklärung).

Berlin [beg'lɪn] ist die **Bundeshauptstadt** der **Bundesrepublik Deutschland**[14] und zugleich eines ihrer **Länder**. Die Stadt Berlin ist mit fast 3,6 Millionen Einwohnern die **bevölkerungsreichste** und die **flächengröße Gemeinde Deutschlands**. Sie bildet das Zentrum des **Metropolraums Berlin** und der **Agglomeration Berlin** (4,5 Millionen Einw.). Der **Stadtstaat Berlin** umfasst die **Spree** und **Havel** befinden sich im Stadtgebiet kleinere Fließgewässer sowie zahlreiche Seen.

Urkundlich erstmals im 13. Jahrhundert erwähnt, war Berlin im Verlauf der Jahrhunderte eine Residenz- und Hauptstadt Brandenburgs, Preußens und später der Stadt Hauptstadt der Deutschen Demokratischen Republik. Im Zuge der deutschen Wiedervereinigung wurde Berlin wieder gesamtdeutsche Hauptstadt und in der Folge Sitz des **Deutschen Bundestages**, des **Bundesrates** sowie zahlreicher **Bundesministerien**.

Zu den bedeutenden Wirtschaftszweigen in Berlin gehören unter anderem die **Kulturwirtschaft**, die **Biotechnologie** und **Gesundheitswirtschaft** sowie die **Informationstechnologie**, die **Informations- und Kommunikationstechnologien**, die **Bau- und Immobilienwirtschaft**, die **Energietechnik** sowie die **Messe- und Kongresswirtschaft**. Die Stadt Berlin ist ein wichtiger **Verkehrsknotenpunkt** und hat einen hohen Luftverkehrs. Berlin zählt zu den aufstrebenden, internationalen Metropolen. Die Stadt ist überdurchschnittlich erfolgreich und verzeichnet jährlich hohe Zuwachsraten bei der Zahl der Erwerbstätigen. Berlin gilt als **Weltstadt** der Kultur, Politik, Medien und Wissenschaft. Es gibt zahlreiche **Forschungseinrichtungen**, **Sportereignisse**[20] und **Museen** Berlins. Berlin ist seit 2009 mit dem UNESCO-Titel *Stadt des Designs* und ist eines der meistbesuchten Zentren des Kontinents.[21] Berlins Architektur, Museen, Festivals, Nachtleben und vielfältige Lebensbedingungen sind weltweit bekannt.[22]



Wikidata:Q183

Englisch: Germany
country in central Europe
ist ein(e) – souveräner Staat
Gründung/Erstellung – 23. Mai 1949
ist Teil von – (en) Central Europe

Approve

Reject



Land Berlin	
	Einwohnerzahl: 3.6 Mio. Fläche: 891 km² Gegründet: 17.03.1920 Teil von: Deutschland Wappen: Berliner Bär
Landesflagge	
	Landeswappen
Basisdaten	
Staatsform	parlamentarische Republik, teilsouveräner Gliedstaat eines Bundesstaates
Sprache	Deutsch[1]
Postleitzahlen	10115–14199
Vorwahl	030
Kfz-Kennzeichen	B
Bruttoinlandsprodukt (BIP)	129,424 Mrd. € (Februar 2017)[2]
BIP pro Kopf	35.627 € (Februar 2016)[3]
Schulden	56,935 Mrd. € (31. Dezember 2017)[4]
ISO 3166-2	DE-BE

Inhaltsverzeichnis [Verbergen]

- 1 Geographie
 - 1.1 Klima
 - 1.2 Gliederung
 - 1.3 Parkanlagen, Zoos, Schutzgebiete
 - 1.4 Nachbarstädte und -gemeinden
- 2 Geschichte
 - 2.1 Namensbildung
 - 2.2 Markgrafschaft und Kurfürstentum
 - 2.3 Das Stadtkönigreich



WIKIPEDIA
Die freie Enzyklopädie

Hauptseite
Themenportale
Von A bis Z
Zufälliger Artikel

Mitmachen
Artikel verbessern
Neuen Artikel anlegen
Autorenportal
Hilfe
Letzte Änderungen
Kontakt
Spenden

Werkzeuge

Links auf diese Seite
Änderungen an verlinkten Seiten
Datei hochladen
Spezialseiten
Permanenter Link
Seiteninformationen
Wikidata-Datenobjekt
Artikel zitieren

Drucken/exportieren

Buch erstellen
Als PDF herunterladen
Druckversion

In anderen Projekten

Commons

Wikinews

Wikiquote

Wikisource

Wikivoyage

In anderen Sprachen

Artikel Diskussion

Lesen Bearbeiten Quelltext bearbeiten Versionsgeschichte Mehr Wikipedia durchsuchen Koordinaten: 52° 31' N, 13° 24' O I

Berlin

Dieser Artikel befasst sich mit der deutschen Hauptstadt. Zu weiteren Bedeutungen siehe Berlin (Begriffsklärung).

Berlin [beg'lɪn] ist die **Bundeshauptstadt** der **Bundesrepublik Deutschland**[14] und zugleich eines ihrer **Länder**. Die Stadt Berlin ist mit fast 3,6 Millionen Einwohnern die **bevölkerungsreichste** und die **flächengrößte** Gemeinde Deutschlands. Sie bildet das Zentrum des **Metropolraums Berlin** und der **Agglomeration Berlin** (4,5 Millionen Einw.). Der **Stadtstaat Berlin** umfasst die **Spree** und **Havel** befinden sich im Stadtgebiet kleinere Fließgewässer sowie zahlreiche Seen.

Urkundlich erstmals im 13. Jahrhundert erwähnt, war Berlin im Verlauf der Jahrhunderte eine Residenz- und Hauptstadt Brandenburgs, Preußens und später der Stadt. Hauptstadt der Deutschen Demokratischen Republik. Im Zuge der deutschen Wiedervereinigung 1990 wurde Berlin wieder gesamtdeutsche Hauptstadt und in der Folge Sitz des **Deutschen Bundestages**, des **Bundesrates** sowie zahlreicher **Bundesministerien**.

Zu den bedeutenden Wirtschaftszweigen in Berlin gehören unter anderem die **Kunst- und Kulturgüter**, die **Kulturwirtschaft**, die **Biotechnologie** und **Gesundheitswirtschaft** sowie die **Informationstechnologien**, die **Informations- und Kommunikationstechnologien**, die **Bau- und Immobilienwirtschaft**, die **Energietechnik** sowie die **Messe- und Kongresswirtschaft**. Die Stadt Berlin ist ein wichtiger **Verkehrsknotenpunkt** im Eisenbahn-, Straßen- und Luftverkehrs. Berlin zählt zu den aufstrebenden, internationalen Metropolen. Die Stadt ist überregional bekannt für ihre hohe Lebensqualität und verzeichnet jährlich hohe Zuwachsrate bei der Zahl der Erwerbstätigen.

Berlin gilt als **Weltstadt** der Kultur, Politik, Medien und Wissenschaft. Es gibt zahlreiche **Forschungseinrichtungen**, **Sportereignisse**[20] und **Museen** Berlins. Berlin ist seit 2009 mit dem UNESCO-Titel *Stadt des Designs* und ist eines der meistbesuchten Zentren des Kontinents.[21] Berlins Architektur, Museen, Festivals, Nachtleben und vielfältige Lebensbedingungen sind weltweit bekannt.[22]



Wikidata:Q183

Englisch: Germany
country in central Europe
ist ein(e) – souveräner Staat
Gründung/Erstellung – 23. Mai 1949
ist Teil von – (en) Central Europe

Approve

Reject



Land Berlin	
	Landesflagge
	Landeswappen

Basisdaten

Staatsform	parlamentarische Republik, teilsouveräner Gliedstaat eines Bundesstaates
Sprache	Deutsch[1]
Postleitzahlen	10115–14199
Vorwahl	030
Kfz-Kennzeichen	B
Bruttoinlandsprodukt (BIP)	129,424 Mrd. € (Februar 2017)[2]
BIP pro Kopf	35.627 € (Februar 2016)[3]
Schulden	56,935 Mrd. € (31. Dezember 2017)[4]
ISO 3166-2	DE-BE

Inhaltsverzeichnis [Verbergen]

- 1 Geographie
 - 1.1 Klima
 - 1.2 Gliederung
 - 1.3 Parkanlagen, Zoos, Schutzgebiete
 - 1.4 Nachbarstädte und -gemeinden
- 2 Geschichte
 - 2.1 Namensbildung
 - 2.2 Markgrafschaft und Kurfürstentum
 - 2.3 Das „Schaus-Königreich“



WIKIPEDIA
Die freie Enzyklopädie

Hauptseite
Themenportale
Von A bis Z
Zufälliger Artikel

Mitmachen
Artikel verbessern
Neuen Artikel anlegen
Autorenportal
Hilfe
Letzte Änderungen
Kontakt
Spenden

Werkzeuge
Links auf diese Seite
Änderungen an verlinkten Seiten
Datei hochladen
Spezialseiten
Permanenter Link
Seiteninformationen
Wikidata-Datenobjekt
Artikel zitieren

Drucken/exportieren
Buch erstellen
Als PDF herunterladen
Druckversion

In anderen Projekten

Commons
Wikinews
Wikiquote
Wikisource
★ Wikivoyage

In anderen Sprachen

Artikel Diskussion

Lesen Bearbeiten Quelltext bearbeiten Versionsgeschichte Mehr Wikipedia durchsuchen Koordinaten: 52° 31' N, 13° 24' O I B

Berlin

Dieser Artikel befasst sich mit der deutschen Hauptstadt. Zu weiteren Bedeutungen siehe Berlin (Begriffsklärung).

Berlin [beg'lɪn] ist die **Bundeshauptstadt** der **Bundesrepublik Deutschland**[¹] und zugleich eines ihrer Länder. Die Stadt Berlin ist mit fast 3,6 Millionen Einwohnern die **bevölkerungsreichste** und die **flächengrößte** Gemeinde Deutschlands. Sie bildet das Zentrum des **Metropolraums Berlin** und der **Agglomeration Berlin** (4,5 Millionen Einw.). Der Stadtstaat Berlin umfasst die **Spree** und **Havel** befinden sich im Stadtgebiet kleinere Fließgewässer sowie zahlreiche Seen.

Urkundlich erstmals im 13. Jahrhundert erwähnt, war Berlin im Verlauf der Jahrhunderte eine Residenz- und Hauptstadt Brandenburgs, Preußens und später der Stadt Hauptstadt der Deutschen Demokratischen Republik. Im Zuge der deutschen Wiedervereinigung 1990 wurde Berlin wieder gesamtdeutsche Hauptstadt und in der Folge Sitz des **Deutschen Bundestages**, des **Bundesrates** sowie zahlreicher **Bundesministerien**.

Zu den bedeutenden Wirtschaftszweigen in Berlin gehören unter anderem die **Kunst- und Kulturgüter**, die **Kulturwirtschaft**, die **Biotechnologie** und **Gesundheitswirtschaft** sowie die **Informationstechnologien**, die **Informations- und Kommunikationstechnologien**, die **Bau- und Immobilienwirtschaft**, die **Energietechnik** sowie die **Messe- und Kongresswirtschaft**. Die Stadt Berlin ist ein wichtiger **Verkehrsknotenpunkt** im Eisenbahn- und Luftverkehrsnetz. Berlin zählt zu den aufstrebenden, internationalen **Metropolen** Europas. Die Stadt ist mit über 400.000 Betrieben verzeichnet jährlich hohe Zuwachsraten bei der Zahl der Erwerbstätigen.

Berlin gilt als **Weltstadt** der Kultur, Politik, Medien und Wissenschaft. Es verfügt über zahlreiche **Forschungseinrichtungen**, **Sportereignisse**[²⁰] und **Museen** Berlins Architektur, Kultur und Geschichte. Berlin ist der einzige Ort, der den UNESCO-Titel *Stadt des Designs* und ist eines der meistbesuchten Zentren des Kontinents.[²¹] Berlins Architektur, Festivals, Nachtleben und vielfältige Lebensbedingungen sind weltweit bekannt.[²²]



Wikidata:Q183

Englisch: Germany
country in central Europe
ist ein(e) – souveräner Staat
Gründung/Erstellung – 23. Mai 1949
ist Teil von – (en) Central Europe

Approve

Reject



Land Berlin	
	Landesflagge
	Landeswappen
Basisdaten	
Staatsform	parlamentarische Republik, teilsouveräner Gliedstaat eines Bundesstaates
Sprache	Deutsch[¹¹]
Postleitzahlen	10115–14199
Vorwahl	030
Kfz-Kennzeichen	B
Bruttoinlandsprodukt (BIP)	129,424 Mrd. € (Februar 2017)[²]
BIP pro Kopf	35.627 € (Februar 2016)[³]
Schulden	56,935 Mrd. € (31. Dezember 2017)[⁴]
ISO 3166-2	DE-BE

Inhaltsverzeichnis [Verbergen]

- 1 Geographie
 - 1.1 Klima
 - 1.2 Gliederung
 - 1.3 Parkanlagen, Zoos, Schutzgebiete
 - 1.4 Nachbarstädte und -gemeinden
- 2 Geschichte
 - 2.1 Namensbildung
 - 2.2 Markgrafschaft und Kurfürstentum
 - 2.3 Das „Schaus-Königreich“



WIKIPEDIA
Die freie Enzyklopädie

Hauptseite
Themenportale
Von A bis Z
Zufälliger Artikel

Mitmachen
Artikel verbessern
Neuen Artikel anlegen
Autorenportal
Hilfe
Letzte Änderungen
Kontakt
Spenden

Werkzeuge
Links auf diese Seite
Änderungen an verlinkten Seiten
Datei hochladen
Spezialseiten
Permanenter Link
Seiteninformationen
Wikidata-Datenobjekt
Artikel zitieren

Drucken/exportieren
Buch erstellen
Als PDF herunterladen
Druckversion

In anderen Projekten

Commons
Wikinews
Wikiquote
Wikisource
★ Wikivoyage

In anderen Sprachen

Artikel Diskussion

Lesen Bearbeiten Quelltext bearbeiten Versionsgeschichte Mehr Wikipedia durchsuchen Koordinaten: 52° 31' N, 13° 24' O

Berlin

Dieser Artikel befasst sich mit der deutschen Hauptstadt. Zu weiteren Bedeutungen siehe Berlin (Begriffsklärung).

Berlin [begˈli:n] ist die **Bundeshauptstadt** der **Bundesrepublik Deutschland**[¹] und zugleich eines ihrer Länder. Die Stadt Berlin ist mit fast 3,6 Millionen Einwohnern die **bevölkerungsreichste** und die **flächengrößte** Gemeinde Deutschlands. Sie bildet das Zentrum des **Metropolraums Berlin** und der **Agglomeration Berlin** (4,5 Millionen Einw.). Der Stadtstaat Berlin umfasst die **Spree** und **Havel** befinden sich im Stadtgebiet kleinere Fließgewässer sowie zahlreiche Seen.

Urkundlich erstmals im 13. Jahrhundert erwähnt, war Berlin im Verlauf der Jahrhunderte eine Residenz- und Hauptstadt Brandenburgs, Preußens und des Deutschen Kaiserreichs. Von 1949 bis 1990 war Berlin wieder gesamtdeutsche Hauptstadt und in der Folge Sitz des **Deutschen Bundestages**, des **Bundesrates** sowie zahlreicher **Bundesministerien**.

Zu den bedeutenden Wirtschaftszweigen in Berlin gehören unter anderem die **Kulturwirtschaft**, die **Biotechnologie** und **Gesundheitswirtschaft** sowie die **Informationstechnologien**, die **Informations- und Kommunikationstechnologien**, die **Bau- und Immobilienwirtschaft**, die **Energietechnik** sowie die **Messe- und Kongresswirtschaft**. Die Stadt Berlin ist ein wichtiger **Verkehrsknotenpunkt** und hat einen guten **Flughafen** sowie einen **großen Bahnhof**. Berlin zählt zu den aufstrebenden, internationalen **Metropolen** Europas. Die Stadt ist überregional bekannt für ihre hohe Lebensqualität und verzeichnet jährlich hohe Zuwachsraten bei der Zahl der Erwerbstätigen.

Berlin gilt als **Weltstadt** der Kultur, Politik, Medien und Wissenschaft. Es gibt zahlreiche **Forschungseinrichtungen**, **Sportereignisse**[²⁰] und **Museen** Berlins. Berlin ist seit 2009 mit dem UNESCO-Titel **Stadt des Designs** und ist eines der meistbesuchten Zentren des Kontinents.[²¹] Berlins Architektur, Museen, Festivals, Nachtleben und vielfältige Lebensbedingungen sind weltweit bekannt.[²²]

Inhaltsverzeichnis [Verbergen]

- 1 Geographie
 - 1.1 Klima
 - 1.2 Gliederung
 - 1.3 Parkanlagen, Zoos, Schutzgebiete
 - 1.4 Nachbarstädte und -gemeinden
- 2 Geschichte
 - 2.1 Namensbildung
 - 2.2 Markgrafschaft und Kurfürstentum
 - 2.3 Das Stadtkönigreich



Wikidata:Q183

Englisch: Germany
country in central Europe
ist ein(e) – souveräner Staat
Gründung/Erstellung – 23. Mai 1949
ist Teil von – (en) Central Europe

Approve

Reject



Land Berlin	
	Landesflagge
	Landeswappen
Basisdaten	
Staatsform	parlamentarische Republik, teilsouveräner Gliedstaat eines Bundesstaates
Sprache	Deutsch[¹¹]
Postleitzahlen	10115–14199
Vorwahl	030
Kfz-Kennzeichen	B
Bruttoinlandsprodukt (BIP)	129,424 Mrd. € (Februar 2017)[²]
BIP pro Kopf	35.627 € (Februar 2016)[³]
Schulden	56,935 Mrd. € (31. Dezember 2017)[⁴]
ISO 3166-2	DE-BE

- *Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for ArticlePlaceholders*, Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon Hare, Elena Simperl, ESWC 2018
- *Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata*, Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon Hare, Elena Simperl, NAACL 2018
- *The Human Face of the Web of Data: A Cross-Sectional Study of Labels*, Lucie-Aimée Kaffee, Elena Simperl, SEMANTiCS 2018
- *Property Label Stability in Wikidata*, Thomas Pellissier Tanon, Lucie-Aimée Kaffee, Wiki Workshop at The Web Conference 2018
- *A Glimpse into Babel: An Analysis of Multilinguality in Wikidata*, Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, Lydia Pintscher, OpenSym 2017
- *Provenance Information in a Collaborative Knowledge Graph: an Evaluation of Wikidata External References*, Alessandro Piscopo, Lucie-Aimée Kaffee, Chris Phethean, Elena Simperl, ISWC 2017
- *What do Wikidata and Wikipedia Have in Common?: An Analysis of their Use of External References*, Alessandro Piscopo, Pavlos Vougiouklis, Lucie-Aimée Kaffee, Christopher Phethean, Jonathon Hare, Elena Simperl, OpenSym 2017
- *Neural Wikipedian: Generating Textual Summaries from Knowledge Base Triples*, Pavlos Vougiouklis, Hady Elsahar, Lucie-Aimée Kaffee, Christophe Gravier, Frédérique Laforest, Jonathon Hare, Elena Simperl (under submission)

Thanks!

kaffee@soton.ac.uk
luciekaffee.github.io
@frimelle

Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for ArticlePlaceholders, Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon Hare, Elena Simperl, ESWC 2018

Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata, Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon Hare, Elena Simperl, NAACL 2018

Property Label Stability in Wikidata, Thomas Pellissier Tanon, Lucie-Aimée Kaffee, Wiki Workshop at The Web Conference 2018

A Glimpse into Babel: An Analysis of Multilinguality in Wikidata, Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, Lydia Pintscher, OpenSym 2017