

Towards Understanding and Building a Multilingual and Inclusive Web

John Samuel
CPE Lyon

Goal

Understanding open collaborative communities for building a multilingual and inclusive web experience.

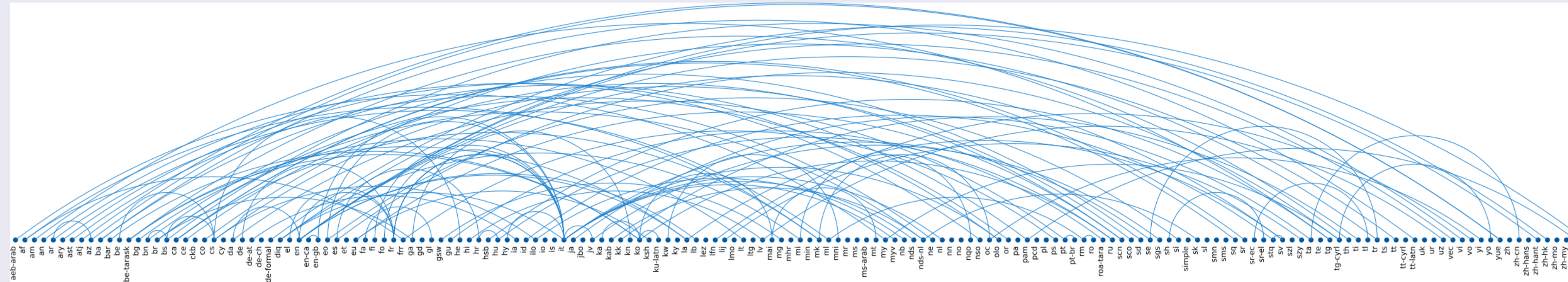


Figure: Translation path of Wikidata Property P17 (country)

Proposed Approach

- Move beyond "One model fits all" approach of building software and web applications.
- Understand collaborative aspects of multilingual ontology building [Sam18].
- Understand how language communities interact [Sam18]
- Develop language-agnostic tools for improving language and topic coverage (e.g., CSV files) [Sam21]

Focus of Study

- 1 Wikidata supported languages
- 2 Wikidata properties, lexicographical data and entity schemas
- 3 WikiProjects (e.g., WikiProject LGBT)
- 4 Wikidata tools (e.g., QuickStatements, OpenRefine)

Analyzing Translations

Count of translated labels

en (8903)	nl (8901)	fr (8376)	ca (8373)	uk (7382)	ar (7279)	de (5356)	tr (5196)	mk (4748)	es (4738)	it (4669)	ru (4352)	sr (4281)	pl (4162)
be-tarask (4040)	sv (3575)	hu (2946)	ja (2800)	pt (2765)	nb (2672)	zh-hans (2652)	da (2607)	ko (2496)	he (2479)	fa (2338)	cs (2321)	zh (2308)	eo (2271)
						zh-hant (2236)	fi (2208)	en-gb (2152)	be (1565)	nn (1532)	el (1512)	pt-br (1489)	
bn (1474)	id (1450)	cy (1383)	lv (1373)	ms (1370)	zh-tw (1273)	gl (1273)	eu (1077)	vi (1065)	nds (1009)	ro (980)	tt (933)	sco (923)	en-ca (916)
ast (848)	bs (819)	lb (794)	sq (774)	zh-hk (772)	oc (747)	sl (735)	scn (730)	zh-cn (727)	sr-ec (723)	sk (716)	ur (679)	tg (678)	sr-el (668)
									bg (650)	et (616)	ka (596)	hr (591)	ga (589)
br (564)	ba (556)	th (546)	hy (532)	vec (518)	simple (502)	zh-sg (485)	zh-mo (480)	tt-cyrl (479)	af (475)	mr (468)	ta (442)	te (432)	ia (420)
									zh-my (410)	az (408)	de-at (387)	bar (378)	hsb (374)
									de-ch (369)	lt (366)	ml (364)	gu (348)	
										ku-latn (322)	ilo (292)	yue (280)	
										ms-arab (261)	la (249)	gsw (247)	
hi (245)	tg-cyrl (245)	ckb (244)	nqo (239)	sh (208)	is (206)	fo (193)	fr (192)	or (192)	kn (176)	ne (170)	se (163)	yi (153)	yo (150)
		pa (147)	diq (146)	min (142)	fy (136)	smn (131)	ha (128)	ce (123)	mg (120)	sms (118)	tl (115)	arz (109)	io (94)
szl (90)	uz (87)	nds-nl (85)	mt (82)	ts (79)	kab (78)	rm (76)	ps (74)	ti (73)	my (69)	ksh (69)	nap (69)	gd (67)	lfn (66)
pam (64)	ky (62)	hyw (62)	ain (59)	lmo (55)	co (53)	ary (52)	am (52)	szl (52)	lij (52)	jv (46)	kw (46)	atj (44)	vo (43)
olo (40)	an (39)	sd (34)	ku (34)	kk (33)	mzn (32)	pcd (30)	si (29)	tt-latn (29)	sw (28)	mi (28)	jbo (27)	kl (27)	pap (26)

Figure: Translated label statistics of Wikidata Properties

Tools Developed

- WDPProp <https://wdprop.toolforge.org/>
- ShExStatements <https://shexstatements.toolforge.org/>

URLs

- Wikidata <https://www.wikidata.org/>
- WikiProject LGBT https://www.wikidata.org/wiki/Wikidata:WikiProject_LGBT
- WikiProject Schemas https://www.wikidata.org/wiki/Wikidata:WikiProject_Schemas
- Lexicographical data https://www.wikidata.org/wiki/Wikidata:Lexicographical_data
- Tools <https://www.wikidata.org/wiki/Wikidata:Tools>

References

- 1 John Samuel. Collaborative approach to developing a multilingual ontology: A case study of wikidata. In Emmanouel Garoufallou, Sirje Virkus, Rania Siatiri, and Damiana Koutsomiha, editors, *Metadata and Semantic Research*, pages 167–172, Cham, 2017. Springer International Publishing.
- 2 John Samuel. Analyzing and visualizing translation patterns of wikidata properties. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings*, pages 128–134, 2018.
- 3 John Samuel. ShExStatements: Simplifying Shape Expressions for Wikidata. In *Companion Proceedings of the Web Conference 2021*, pages 610–615. Association for Computing Machinery, New York, NY, USA, April 2021.