

Multilingual Fact Linking

Keshav Kolluru

(IIT Delhi)

Martin Rezk, Pat Verga, William Cohen, Partha Talukdar

(Google)

Task Definition

Fact Linking

Identify *known* facts present in a sentence

Delhi, the Indian capital, is known for its rich heritage

WikiData Facts

(Q987: New Delhi, P1376: capital of, Q668: India)

(Q987: New Delhi, P17: country, Q668: India)

Task Definition

Multilingual Fact Linking:

Sentence and Fact labels in different languages

भारत की राजधानी दिल्ली अपनी समृद्ध विरासत के लिए जानी जाती है

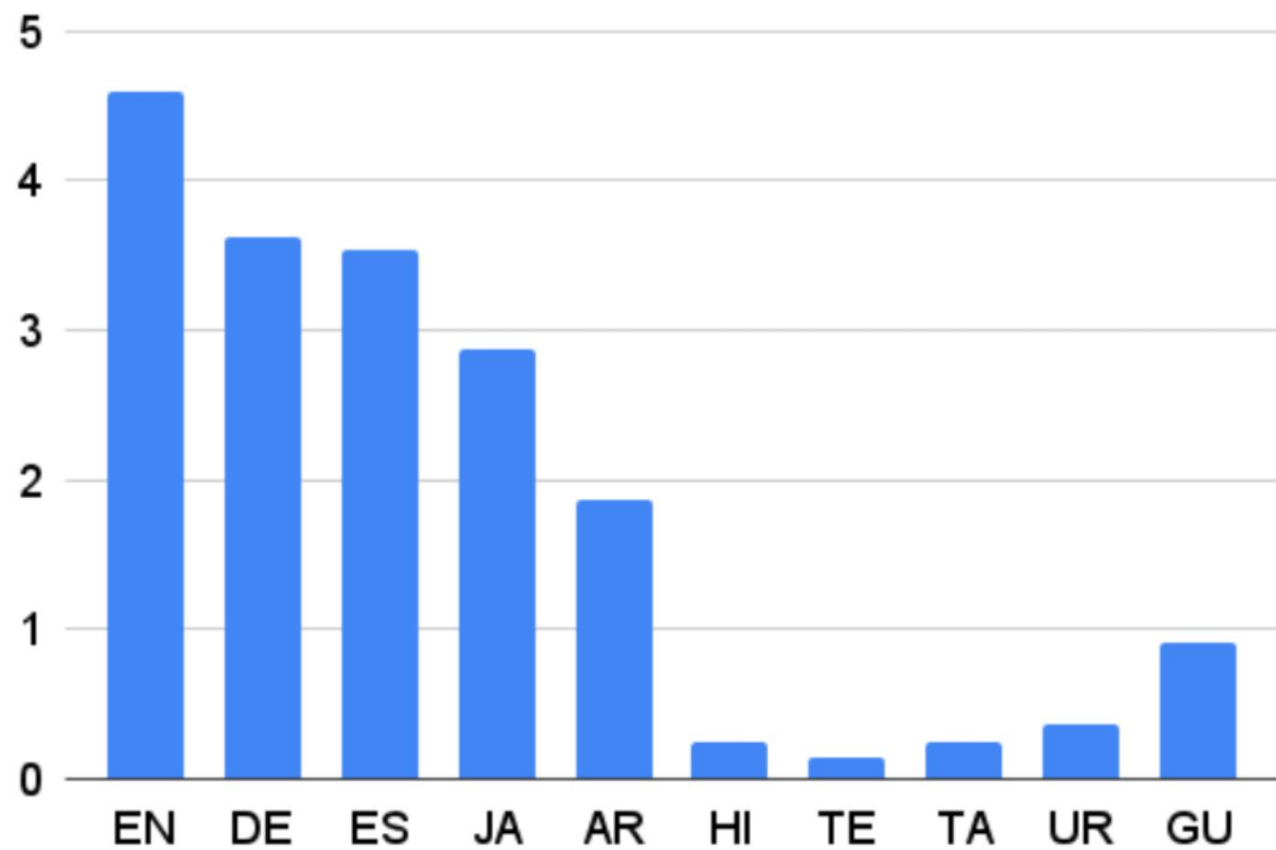
Delhi, the Indian capital, is known for its rich heritage

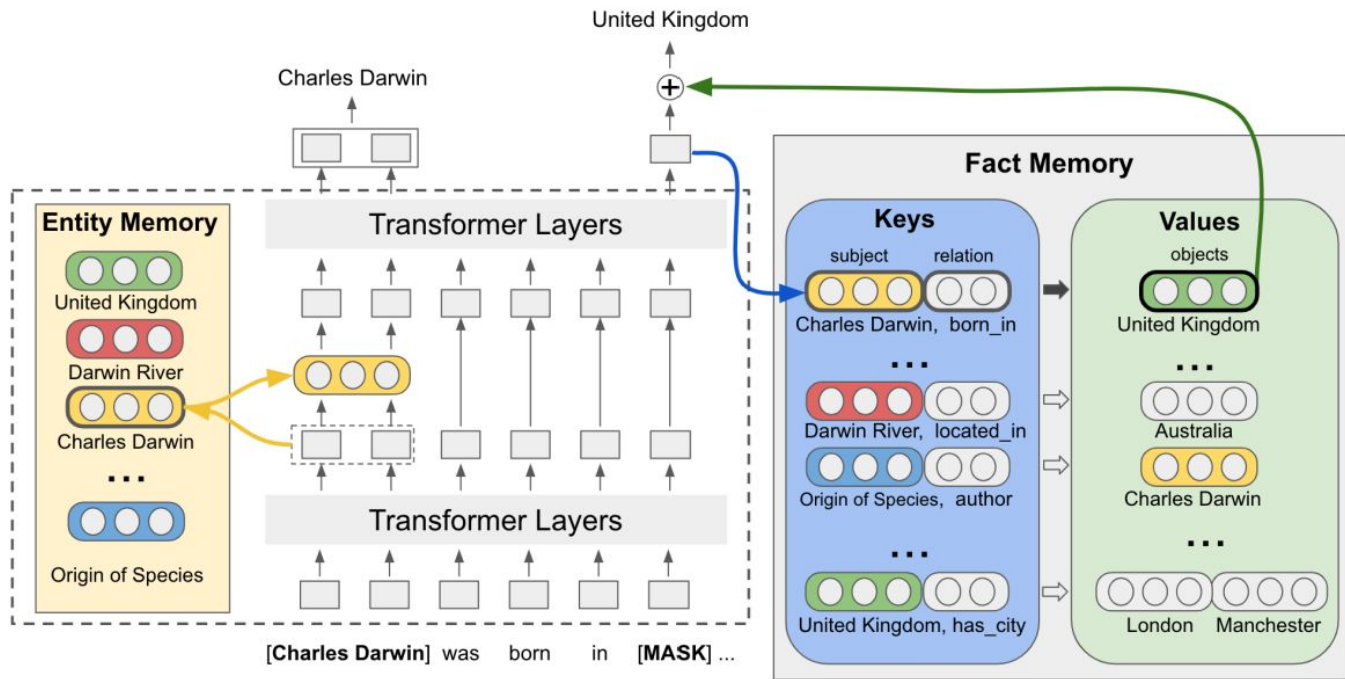
WikiData Facts

(Q987: New Delhi, P1376: capital of, Q668: India)

(Q987: New Delhi, P17: country, Q668: India)

Wikidata Fact Labels (in millions)

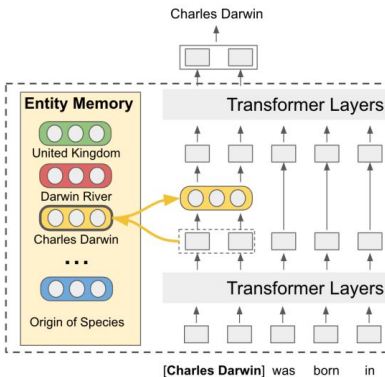




Verga et. al. (2021)

Injecting knowledge into multilingual LMs

Uni



Injecting knowledge
Verga et. al.




Linking multilingual queries
to shopping catalogs

Search bar: dell xps 14z screen

Navigation: All Shopping Images Videos News More Tools

About 1,41,000 results (0.67 seconds)

Ads · Shop dell xps 14z screen

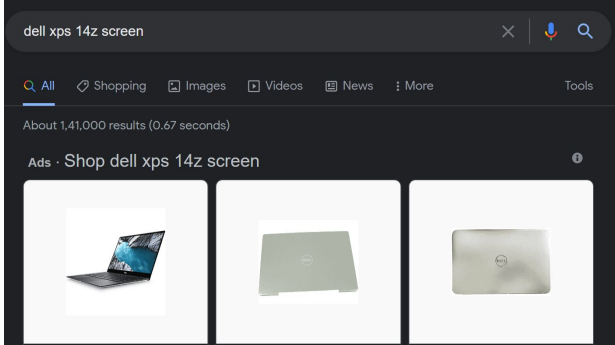
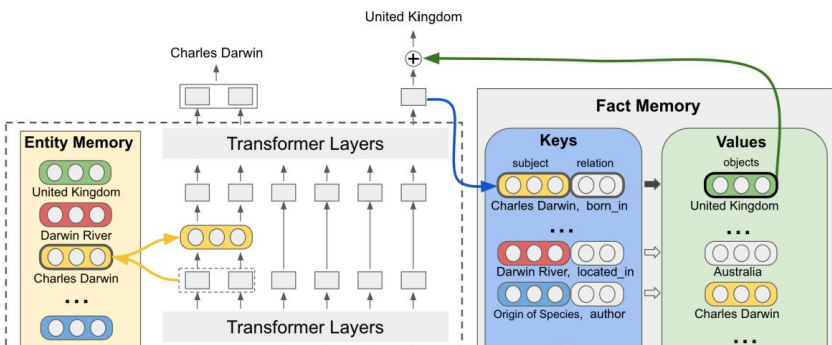
 <p>Dell XPS 13 Laptop - w/ 11th gen Intel Core - 13....</p> <p>₹1,13,990</p> <p>Dell India</p> <p>★★★★★ (392)</p>	 <p>Rubaintech Dell XPS 14z (L412z) 14" LCD Back...</p>	 <p>Jlvaa Infotech Laptop Panel LCD Back Cover f...</p>
---	---	---

Search bar: डेल स्पस 14z स्क्रीन

Navigation: All Maps Videos Shopping Images More Tools

About 1,67,000 results (0.74 seconds)

Tip: Search for **English** results only. You can specify your search language in [Preferences](#)



<https://de.wikipedia.org/wiki/Indien> · Translate this page

Indien - Wikipedia

Emperor of **India** (Kaiser(in) von Indien). Kolonialflagge Britisch-Indiens. 1885 wurde in Bombay der **Indian National Congress** (Kongresspartei) gegründet. Er ...

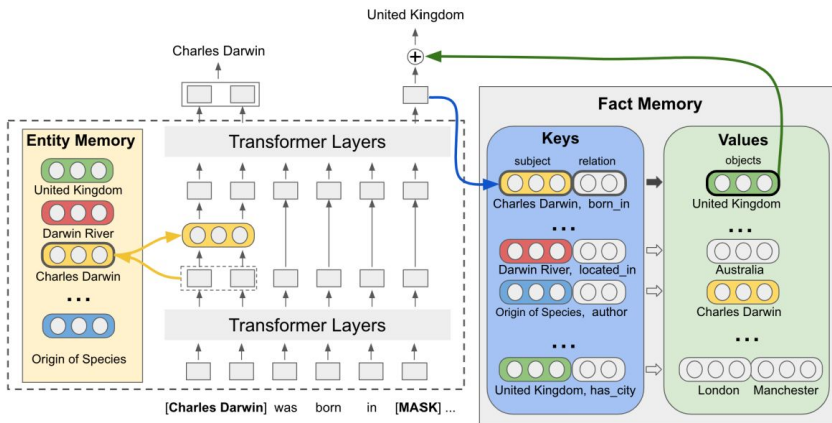
Date format: : dd-mm-yyyy;

Government: Federal parliamentary consti...

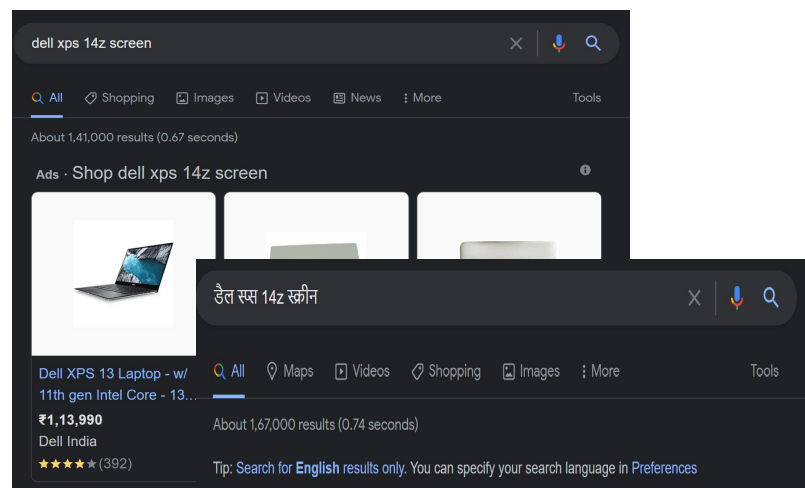
Demonym(s): Indian

Religion (2011): 79.8% Hinduism; 14.2% I...

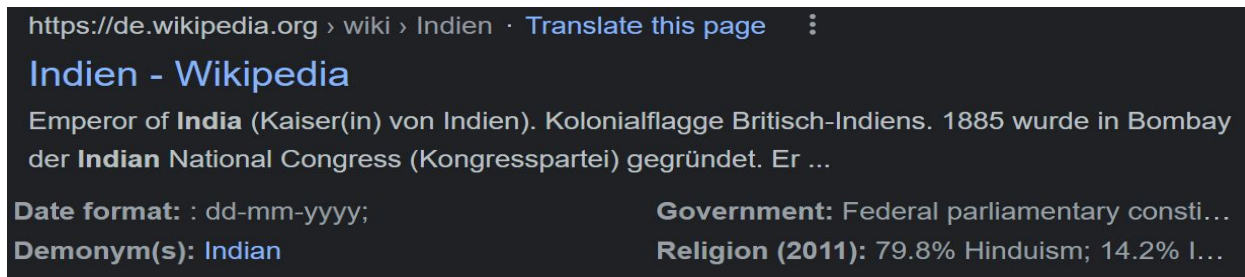
Multilingual Structured Snippets



Injecting knowledge into LMs
Verga et. al. (2021)



Linking multilingual queries to shopping catalogs



Multilingual Structured Snippets

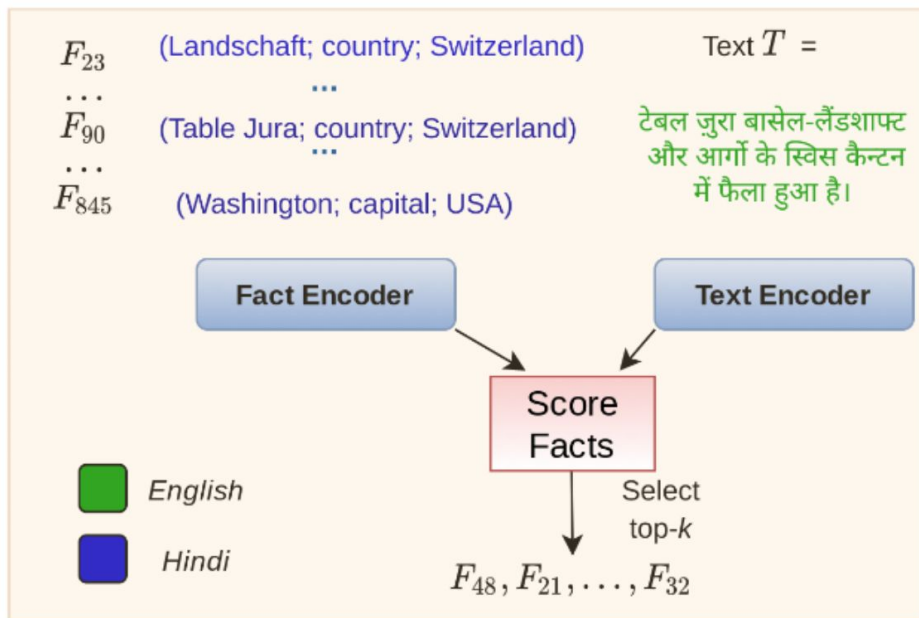
IndicLink Dataset

- New evaluation dataset for six Indian languages
- Sentences annotated with Wikidata facts

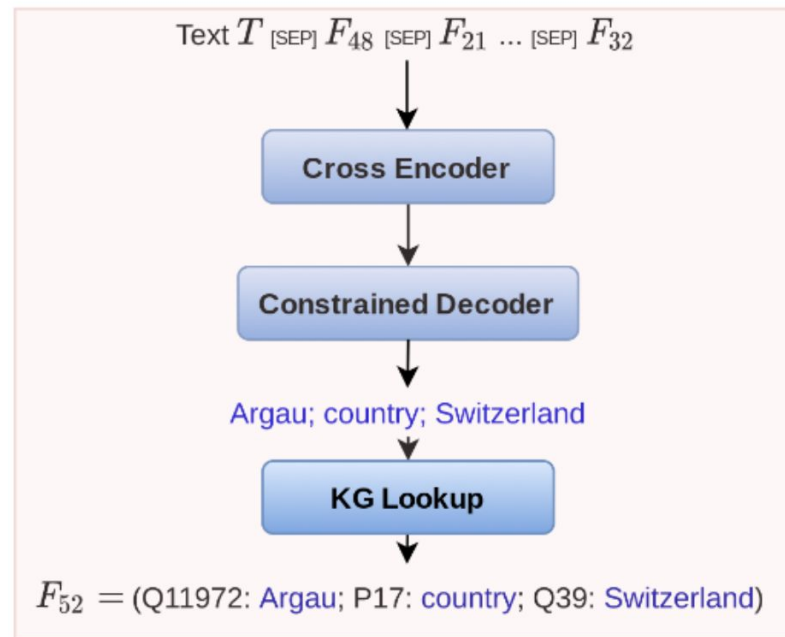
IndicLink	English (EN)	Hindi (HI)	Telugu (TE)	Tamil (TA)	Urdu (UR)	Gujarati (GU)	Assamese (AS)	Total
#Test Examples	1002	889	888	881	1001	881	887	6429
#KG Facts	4.6M	230K	145K	248K	361K	91K	257	4.6M

ReFCoG: Retrieval + Fact-Constrained Generation

Retrieval



Generation



Results

Model	EN	HI	TE	TA	UR	GU	AS	Average	
	P@1	P@1	P@1	P@1	P@1	P@1	P@1	P@1	R@5
DE _{ALL-Sum}	37.5	29.7	32.8	27.8	28.7	29.9	13.8	28.6	31.9
REFCOG _{ALL-Sum, EL}	56.4	52.4	53.4	53.2	53.6	52.5	43.1	52.1	51.5

ReFCoG code and IndicLink data available at github.com/SaiKeshav/mfl

