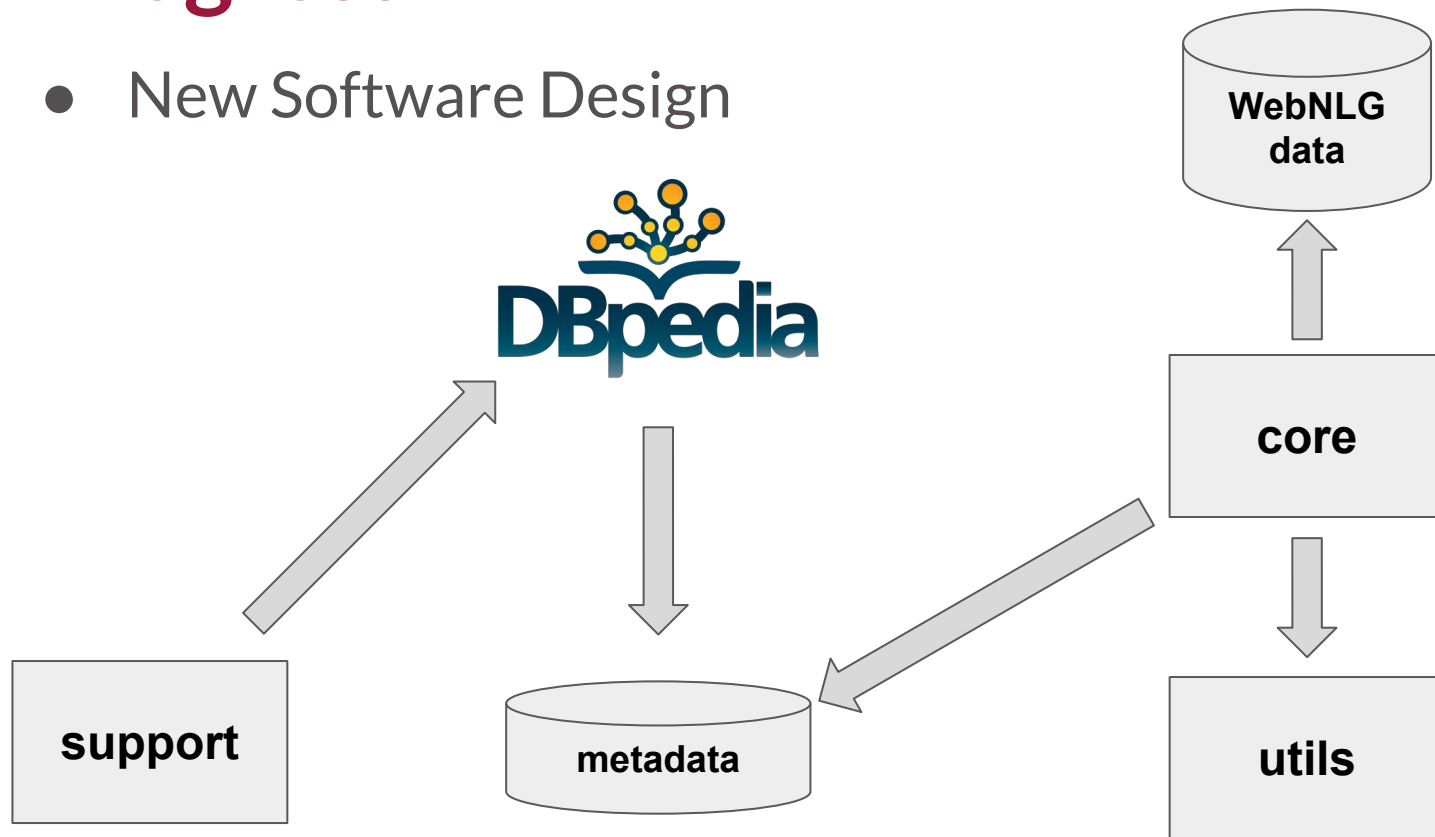# Learning to generate text from knowledge graphs with seq2seq models

Reem, François, Badr

Progress

- New Software Design

Text Delexicalization

- Simple (exact matching with lowercasing)
- Advanced
  - abbreviation handling: United States of America → USA
  - Date format handling: 2017-12-05 → Dec. 6th, 2017
  - Character-level string matching with Ratcliff/Obershelp algorithm: Lars Lokke Rasmussen → Lars Løkke Rasmussen

Semantic Types for Entities

- First option: use the domain / range of the property
- Second option: use the semantic type retrieved from DBPedia
- Third option: use the uppercased property

Domain & Range

40.3%

No Domain or Range

33.9%

5.8%

20.0%

[181, 26, 90, 152]

Domain only

Range only
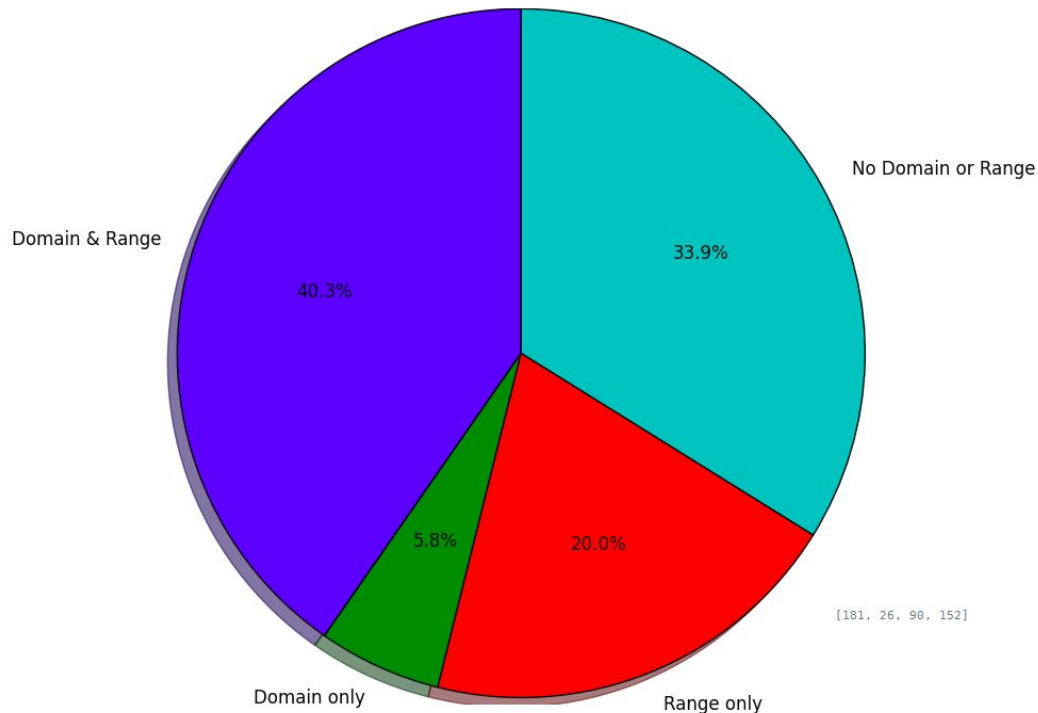
Graph representation

- Flat sequence

  ENTITY_1 FOOD country ENTITY_2 COUNTRY ENTITY_1 FOOD ingredient ENTITY_3 FOOD

- Structured

  [1] ((ENTITY_1 FOOD (ingredient (ENTITY_2 FOOD)) (country (ENTITY_3 COUNTRY))))

  [2] ((ENTITY_2 FOOD (ingredient (ENTITY_1 FOOD)))

  (ENTITY_3 COUNTRY (country (ENTITY_1 FOOD))))

Experiments 1

- Eval with BLEU on 3 references (seen and unseen)
- Impact of delex

	dev	test
simple	**48.90**	**22.02**
advanced	41.06	21.98

Experiments 2

- Graph representation

Model	dev	test
Simple flat	**48.90**	**22.02**
Structured 1	21.57	15.69
Structured 2	28.52	17.30
Structured X	30.94	18.66

Some stats

Model	Avg. Hyp. Length	Avg source Length	Avg target Length	BLEU
Simple flat	21.12	16.8	19.25	**48.90**
Advnaced flat	20.88	16.8	18.05	41.06
Structured 1	15.39	32.00	19.25	21.57
Structured 2	19.52	36.66	19.25	28.52
Structured X	19.10	33.8	19.25	30.94

Next

- Error analysis
- Extensive evaluation with different metrics (by size, category, etc.)