Entity Linking using MAG

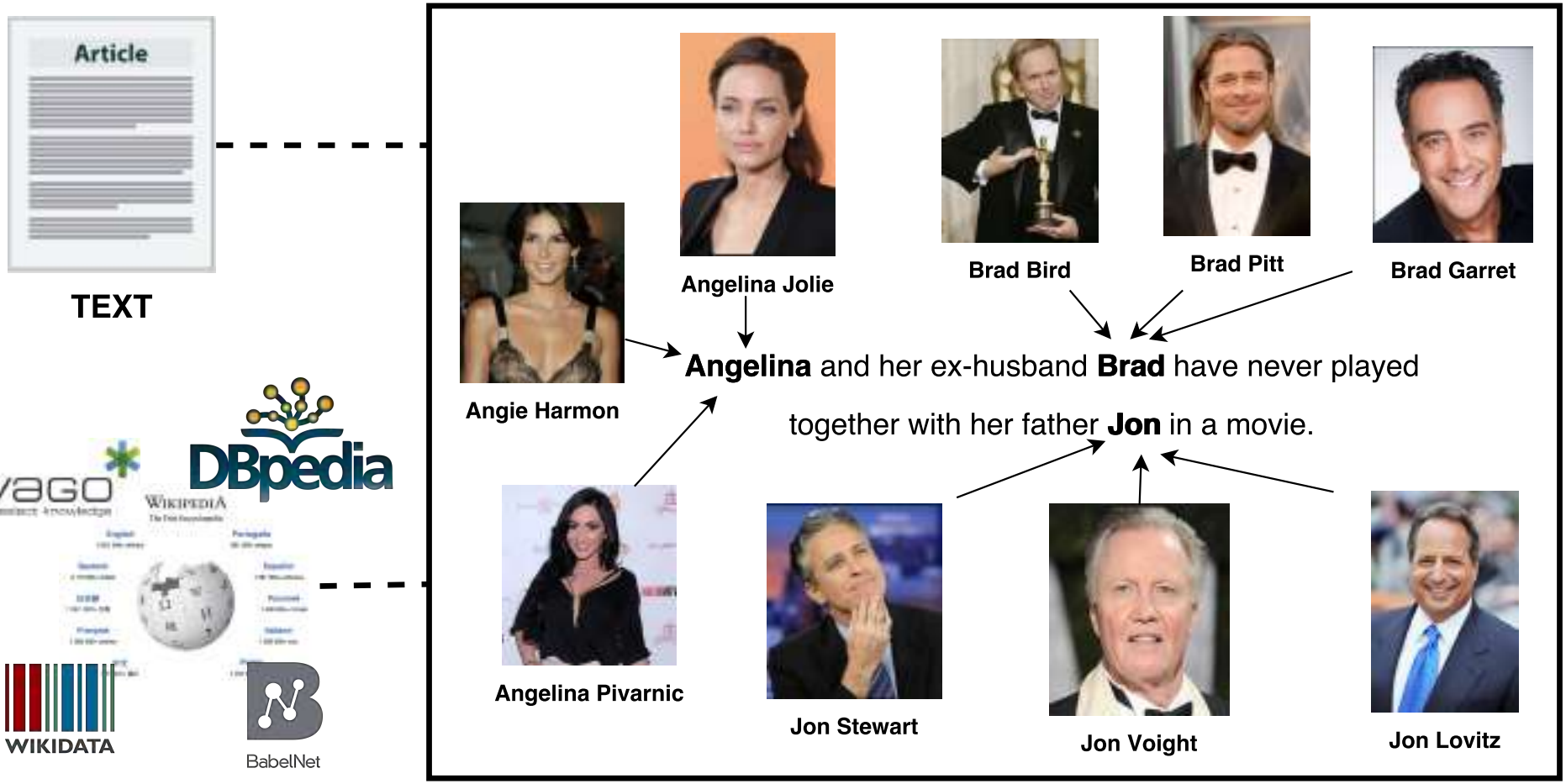
Diego Moussallem[1,2]

[1]Department of Research & Development
University of Leipzig, Germany

[2]Data Science Group
University of Paderborn, Germany

Introduction

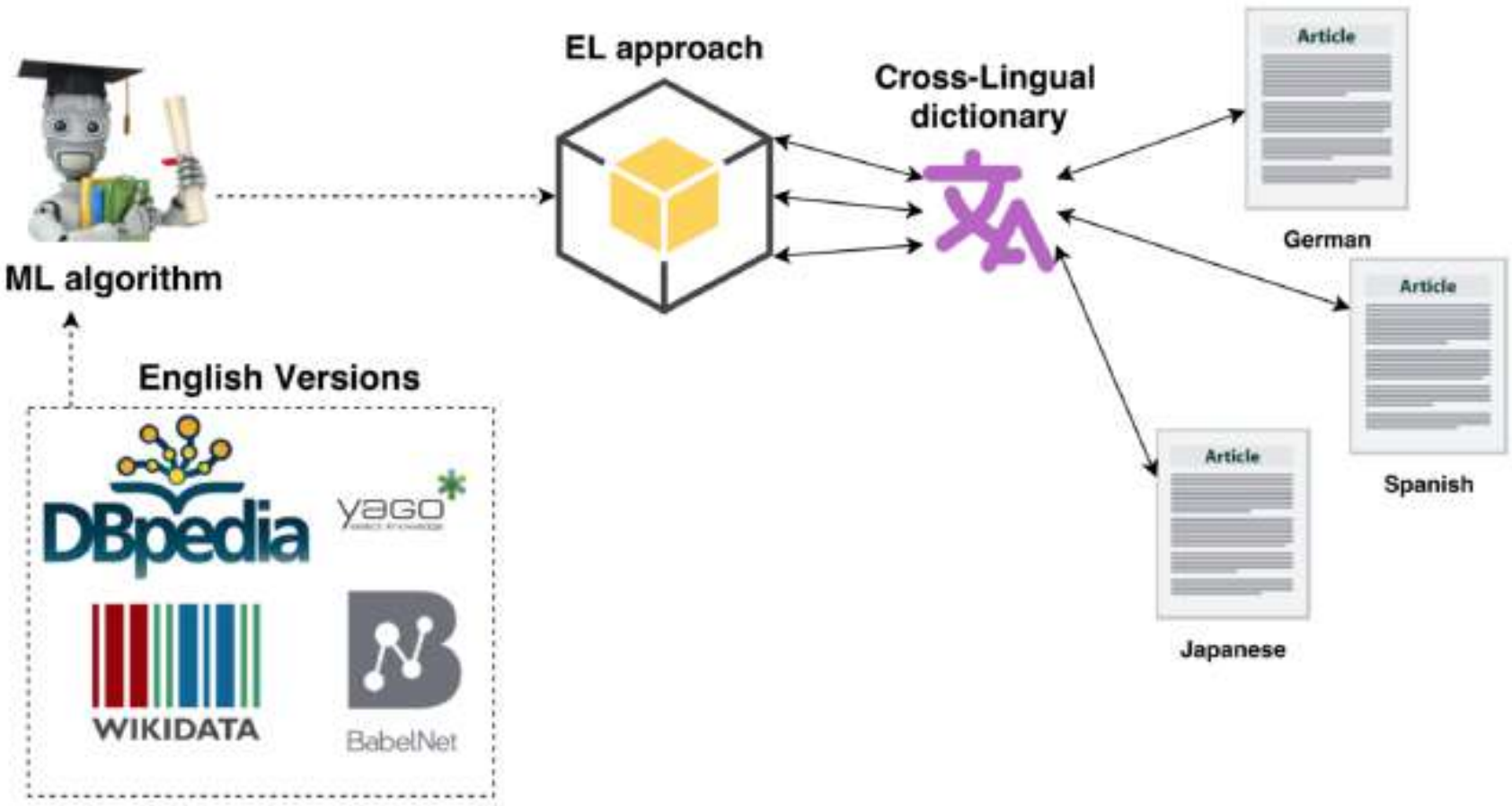
More than 1 exabyte per day


Article
TEXT
yago
WIKIPEDIA
DBpedia
WIKIDATA
BabelNet
Angelina Jolie
Brad Bird
Brad Pitt
Brad Garret
Angie Harmon
Angelina and her ex-husband Brad have never played together with her father Jon in a movie.
Angelina Pivarnic
Jon Stewart
Jon Voight
Jon Lovitz

- AGDISTIS - Graph-Based Disambiguation of Named Entities Using Linked Data by Usbeck et al. published at ISWC in 2014
- MAG - A Multilingual, Knowledge-base Agnostic and Deterministic Entity Linking Approach by Moussallem et al published at K-CAP in 2017
- Entity Linking in 40 Languages using MAG by Moussallem et al. published at ESWC in 2018

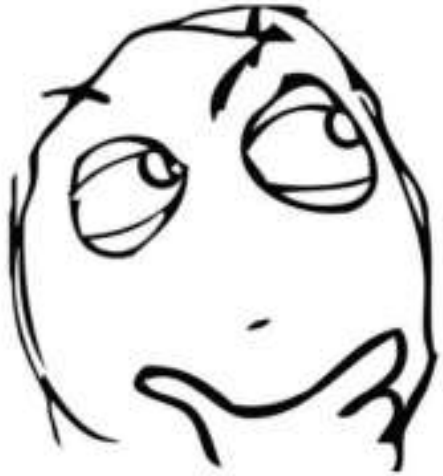
Motivation

Current Drawback: Multilingual Approaches



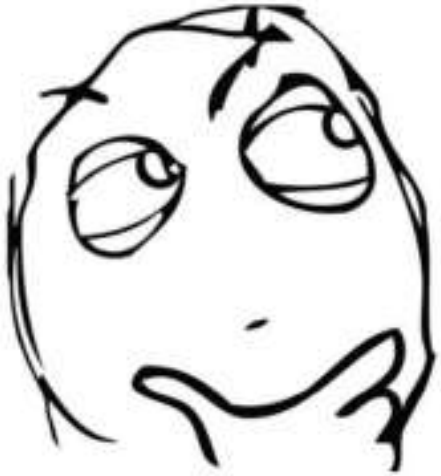
Goal

Create an EL approach which (1) achieves state of the art performance and (2) can easily be ported to different languages.



Goal

Create an EL approach which (1) achieves state of the art performance and (2) can easily be ported to different languages.

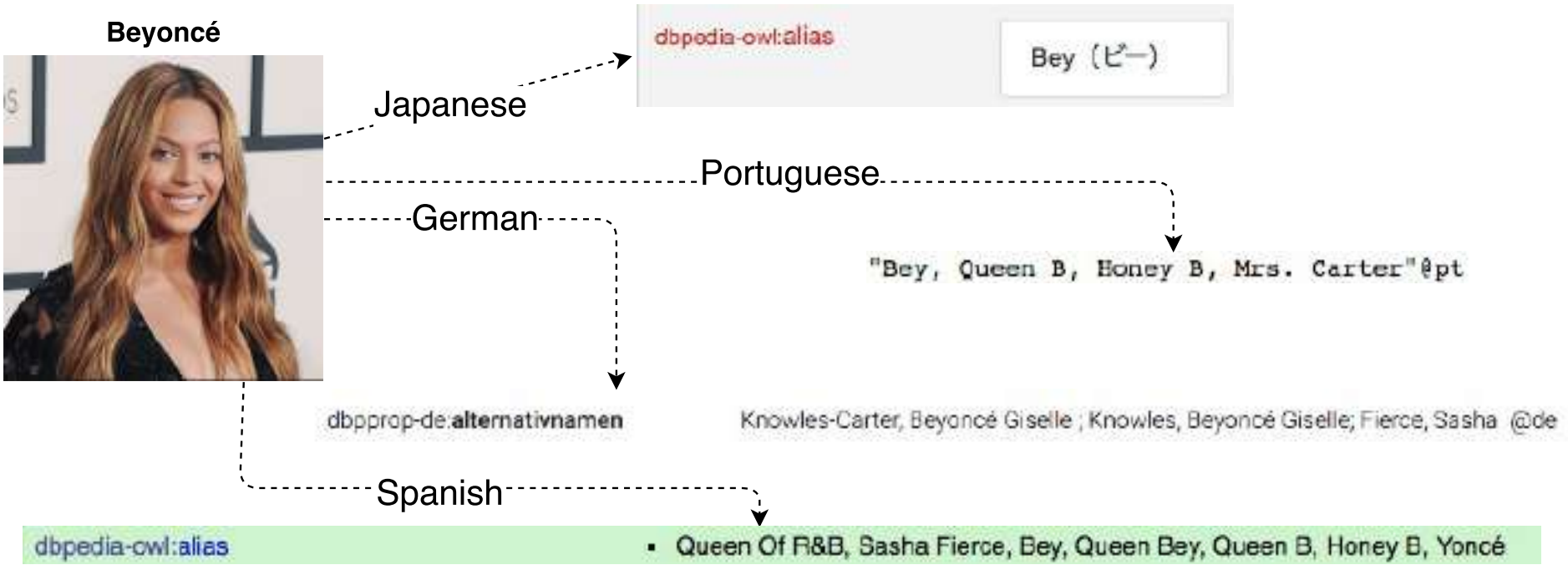


Solution

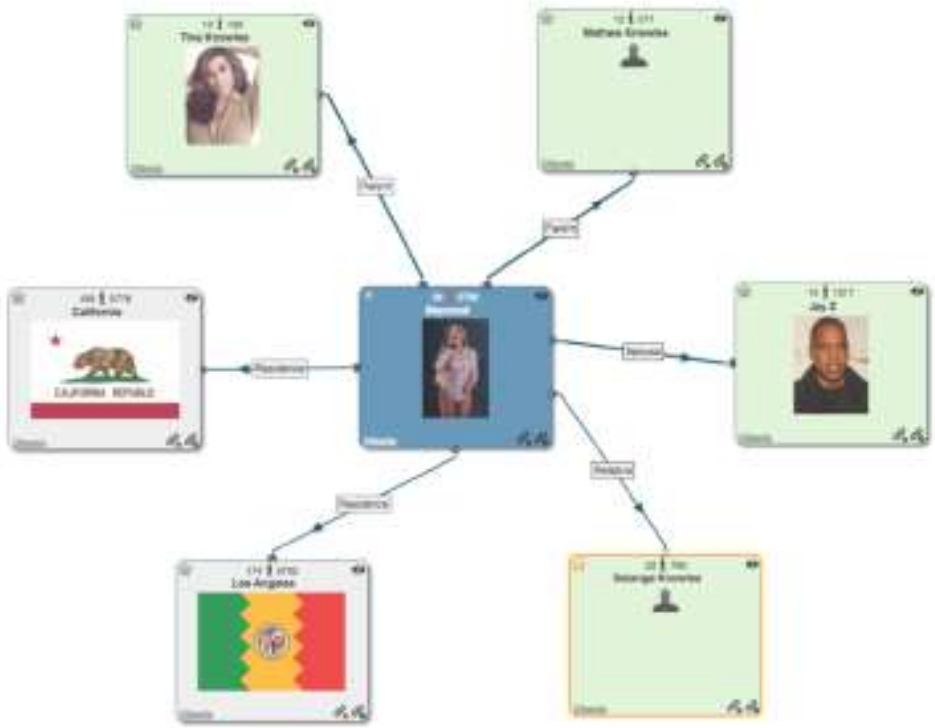
MAG

Index creation

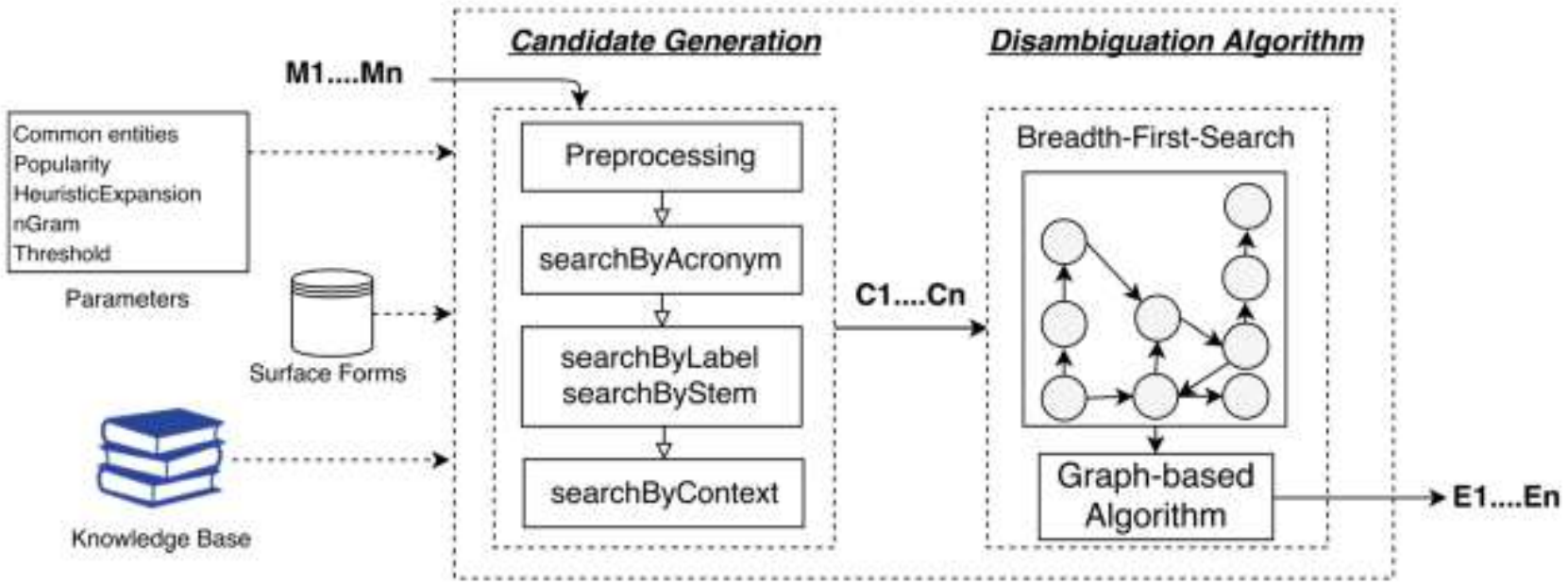
- Surface forms (rdfs:label, schema:name, skos:prefLabel, skos:altLabel and others.)
- Acronyms (list of acronyms provided by Stands4.com)
- Person's name (foaf:name, foaf:nickname, foaf:givenName, dbo:alias and others.)
- Rare references (rdfs:comments, rdfs:description properties.)
- Context (Concise Bounded Descriptions (CBD) of resources)


Beyoncé
Japanese
dbpedia-owl:alias
Bey（ビー）
Portuguese
German
"Bey, Queen B, Honey B, Mrs. Carter"@pt
dbpprop-de:alternativnamen
Knowles-Carter, Beyoncé Giselle ; Knowles, Beyoncé Giselle; Fierce, Sasha @de
Spanish
dbpedia-owl:alias
Queen Of R&B, Sasha Fierce, Bey, Queen Bey, Queen B, Honey B, Yoncé


Beyoncé
Spanish
Beyoncé Giselle Knowles , conocida como Beyoncé (Houston, Texas, Estados Unidos, 4 de septiembre de 1981), es una cantante
Japanese
ビヨンセ・ジゼル・ノウルズ、通称名ビヨンセ（Beyoncé Giselle Knowles, 1981年9月4日 - ）は、アメリカ合衆国テキサス州ヒューストン出身のシンガーソ
English
Beyoncé Giselle Knowles-Carter (/biːˈjɒnseɪ/; born September 4, 1981) is an American singer
Portuguese
"Beyoncé Giselle Knowles Carter (Houston, 4 de setembro de 1981), mais conhecida simplesmente como Beyoncé é uma cantora

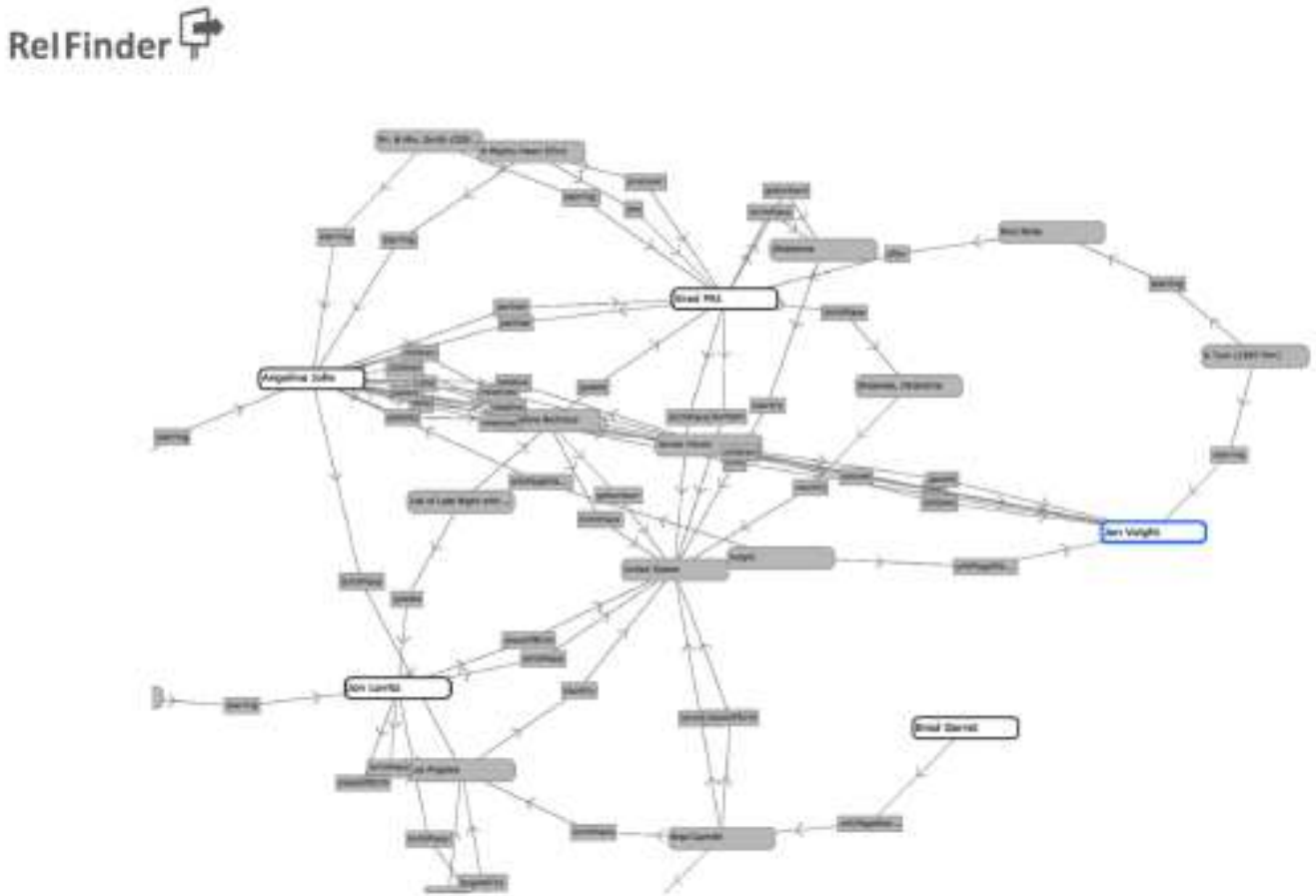


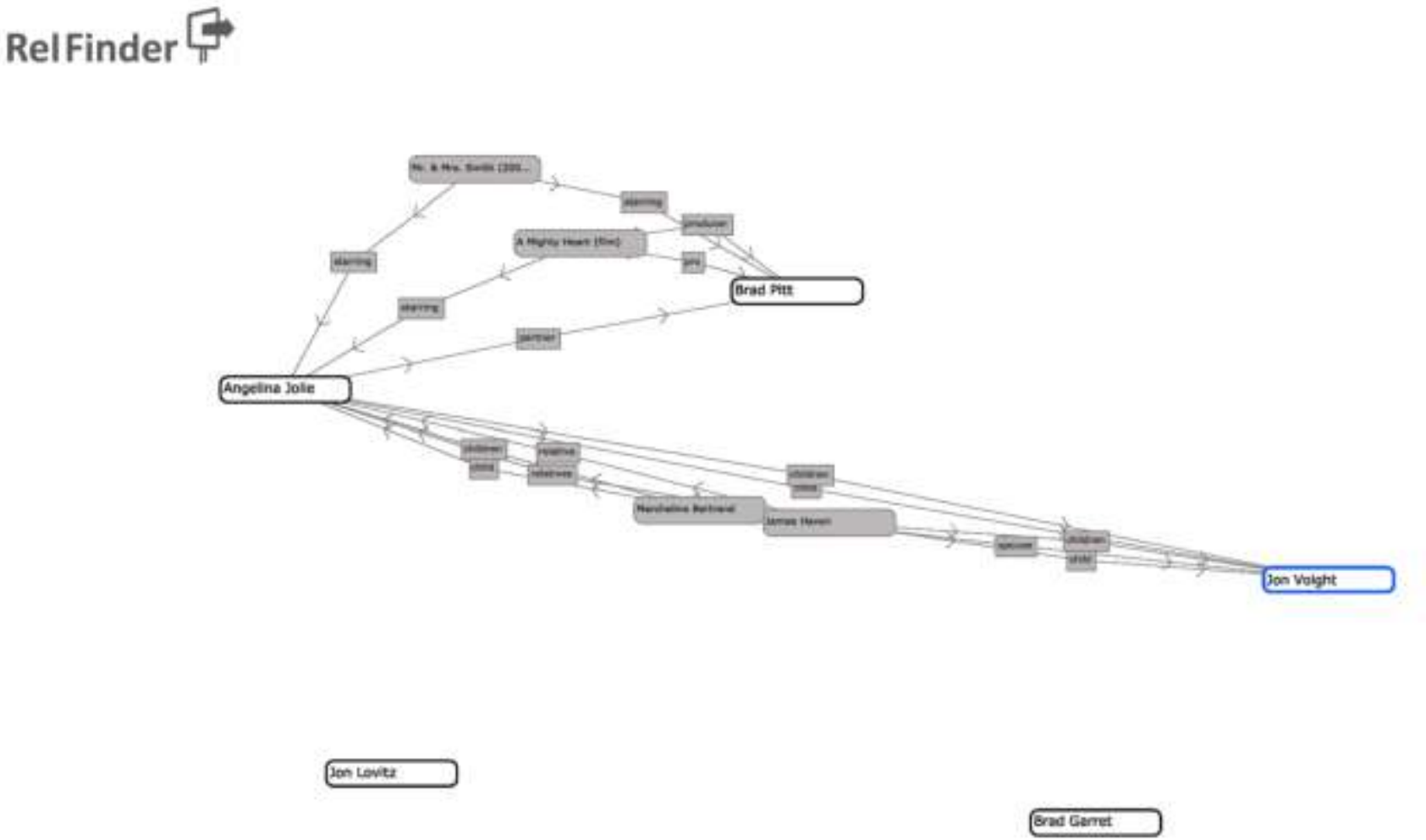
dbr:Beyoncé — [Beyoncé spouse Jay Z sister Solange Knowles parent Tina Knowles residence California residence Los Angeles + N words]



Architecture of MAG

[Angelina] and [Brad] have never played in a movie with her father [Jon].




RelFinder
Brad Pitt
Angelina Jolie
Jon Voight
Jon Lovitz
Brad Garret

Results - K-CAP

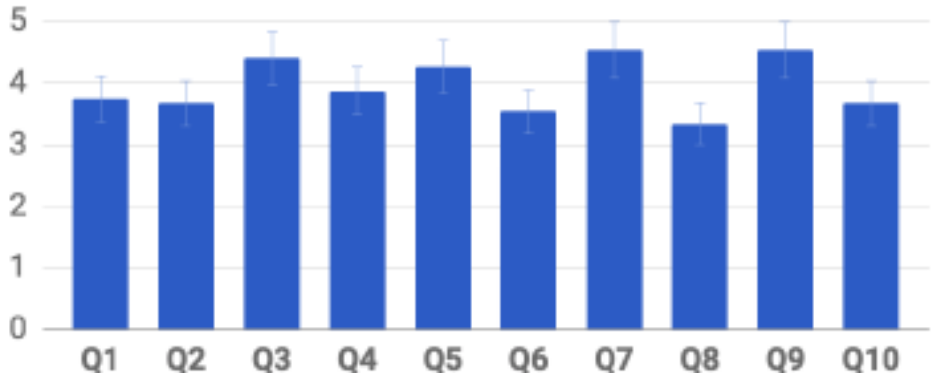
Red entries are the top scores while blue represents the second best scores

Language	Tools/Data sets	AGDISTS	AIDA	Babelfy	DBpedia	DoSer	entityclassifier.eu	FRED	Kea	NERD-ML	PBOH	WAT	xLisa	MAG + HITS	MAG + PR
English	ACE2004	0.65	0.70	0.53	0.48	0.75	0.50	0.00	0.66	0.58	0.72	0.66	0.70	0.69	0.60
	AIDA/CoNLL-Complete	0.55	0.68	0.66	0.50	0.69	0.50	0.00	0.61	0.20	0.75	0.71	0.48	0.59	0.54
	AIDA/CoNLL-Test A	0.54	0.67	0.65	0.48	0.69	0.48	0.00	0.61	0.00	0.75	0.7	0.45	0.59	0.54
	AIDA/CoNLL-Test B	0.52	0.69	0.68	0.52	0.69	0.48	0.00	0.61	0.00	0.75	0.72	0.47	0.57	0.52
	AIDA/CoNLL-Training	0.55	0.69	0.66	0.50	0.69	0.52	0.00	0.61	0.28	0.75	0.71	0.48	0.60	0.55
	AQUAINT	0.52	0.55	0.68	0.53	0.82	0.41	0.00	0.78	0.60	0.81	0.73	0.76	0.67	0.68
	Spotlight	0.27	0.25	0.52	0.71	0.81	0.25	0.04	0.74	0.56	0.79	0.67	0.71	0.65	0.66
	IITB	0.47	0.18	0.37	0.30	0.43	0.14	0.00	0.48	0.43	0.38	0.41	0.27	0.52	0.43
	KORE50	0.27	0.70	0.74	0.46	0.52	0.30	0.06	0.60	0.31	0.63	0.62	0.51	0.24	0.24
	MSNBC	0.73	0.69	0.71	0.42	0.83	0.51	0.00	0.78	0.62	0.82	0.73	0.5	0.79	0.75
	Microposts2014-Test	0.33	0.42	0.48	0.50	0.76	0.41	0.05	0.64	0.52	0.73	0.60	0.55	0.45	0.44
	Microposts2014-Train	0.42	0.51	0.51	0.48	0.77	0.00	0.31	0.65	0.52	0.71	0.63	0.59	0.49	0.44
	N3-RSS-500	0.66	0.45	0.44	0.20	0.48	0.00	0.00	0.44	0.38	0.53	0.44	0.45	0.69	0.67
	N3-Reuters-128	0.61	0.47	0.45	0.33	0.69	0.00	0.41	0.51	0.41	0.65	0.52	0.39	0.69	0.64
	OKE 2015 Task 1 evaluation set	0.59	0.56	0.59	0.31	0.59	0.00	0.46	0.63	0.61	0.63	0.57	0.62	0.58	0.55
	OKE 2015 Task 1 example set	0.50	0.60	0.40	0.22	0.55	0.00	0.60	0.55	0.00	0.50	0.60	0.50	0.67	0.50
	OKE 2015 Task 1 training set	0.62	0.67	0.71	0.25	0.78	0.00	0.61	0.78	0.77	0.76	0.72	0.75	0.72	0.70
Multilingual	N^3 news.de	0.61	0.52	0.50	0.48	0.56	0.28	0.00	0.61	0.33	0.30	0.59	0.36	0.76	0.63
	Italian Abstracts	0.22	0.28	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.80	0.80
	Spanish Abstracts	0.25	0.33	0.26	0.00	0.24	0.27	0.00	0.47	0.00	0.31	0.33	0.31	0.75	0.68
	Japanese Abstracts	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.54	0.54
	Dutch Abstracts	0.33	0.36	0.36	0.28	0.36	0.22	0.00	0.40	0.00	0.5	0.40	0.25	0.66	0.67
	French Abstracts	0.00	0.00	0.28	0.22	0.00	0.25	0.00	0.00	0.00	0.20	0.28	0.28	0.80	0.80
	Average	0.45	0.48	0.50	0.36	0.55	0.24	0.11	0.53	0.31	0.59	0.54	0.45	0.63	0.59
	Standard Deviation	0.19	0.22	0.18	0.19	0.27	0.21	0.21	0.23	0.26	0.20	0.22	0.20	0.13	0.13

Extension for 40 languages using DBpedia

ESWC 2018, we released MAG demo in 40 languages.

- Feedback from the community
  - Good usability according to SUS score
  - Extension to all languages provided by DBpedia
  - Combination with other approaches + Machine Learning



SUS score

http://agdistis.aksw.org/mag-demo/

Future work

- All languages provided by DBpedia (KB completion)
- New graph-based algorithms
- Conventional Machine Learning algorithms + Neural Networks

Thank you for your Attention!

Diego Moussallem
moussallem@informatik.uni-leipzig.de
http://aksw.org/DiegoMoussallem
Follow me on Twitter @DiegoMoussallem

⇒ GITHUB: https://github.com/dice-group/AGDISTIS

⇒ WEBSITE: https://agdistis.aksw.org/mag-demo