From Strings to Things:
KELVIN in TAC KBP and EDL

Tim Finin[1], Dawn Lawrie[2], James Mayfield, Paul McNamee and Craig Harman

Human Language Technology Center of Excellence
Johns Hopkins University

[1]University of Maryland, Baltimore County
[2]Loyola University Maryland

December 2017

Kelvin

- **KELVIN**: **K**nowledge **E**xtraction, **L**inking, **V**alidation and **In**ference
- Developed at the *Human Language Technology Center of Excellence* at JHU and used in TAC KBP (2010-17), EDL (2015-17) and other projects
- Takes English, Chinese & Spanish documents and produce a knowledge graph in several formats
- We'll review its monolingual processing, look at the multi-lingual use case

NIST TAC

NIST Text Analysis Conference

- Annual evaluation workshops since 2008 on natural language processing & related applications with large test collections and common evaluation procedures
- **Knowledge Base Population** (KBP) tracks focus on building KBs from information extracted from text
  - **Cold Start KBP**: construct a KB from text
  - **Entity discovery & linking**: cluster and link entity mentions
  - Slot filling
  - Slot filler validation
  - Sentiment
  - Events: discover and cluster events in text

http://nist.gov/tac

NIST TAC Cold Start

When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated european adult movies available on their cable channels.

After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that

When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated european adult movies available on their cable channels.

Springfield 

After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that

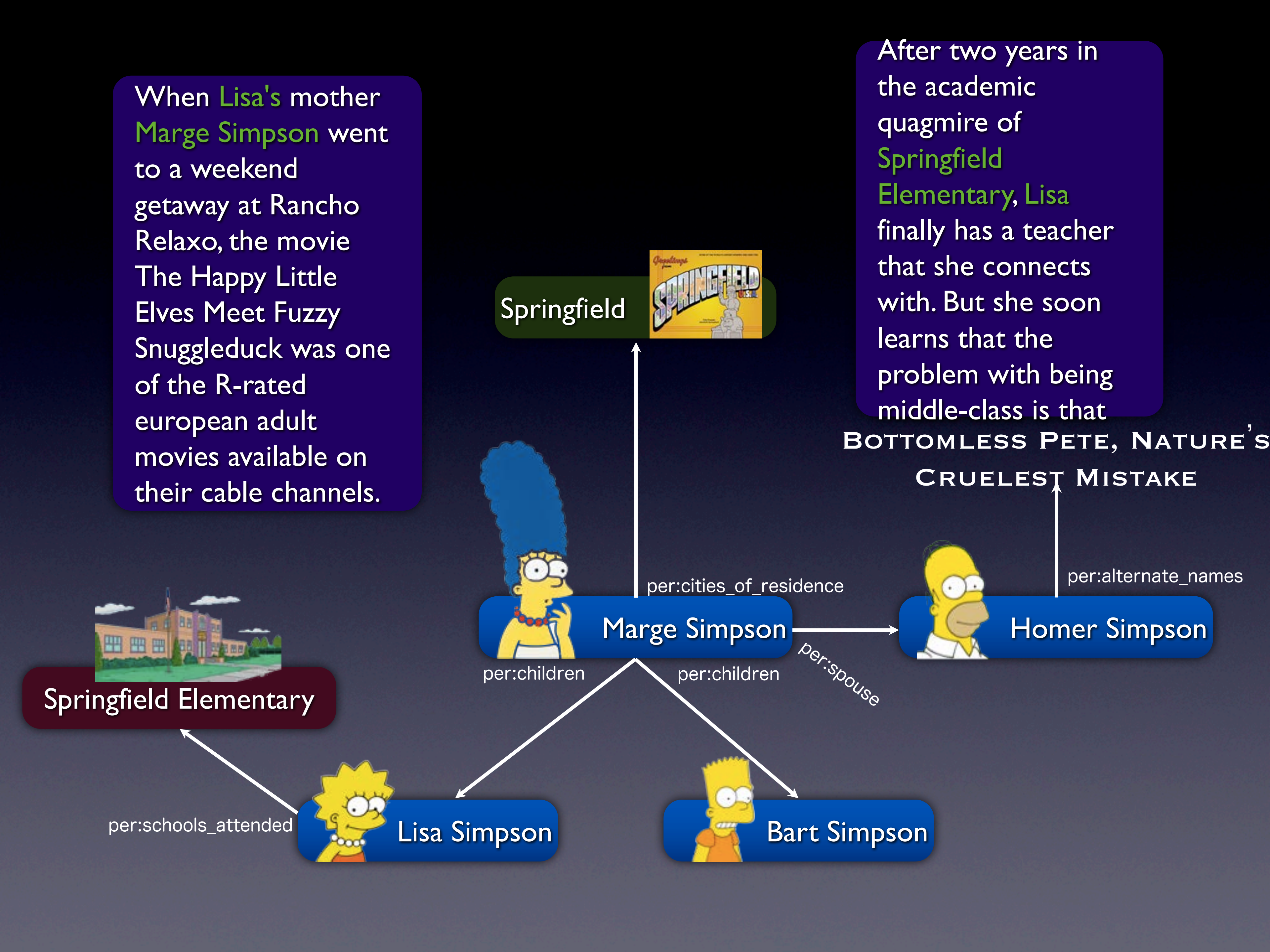

Marge Simpson


Homer Simpson


Springfield Elementary


Lisa Simpson


Bart Simpson


When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated european adult movies available on their cable channels.
After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that
Bottomless Pete, Nature's Cruelest Mistake
Springfield
per:cities_of_residence
per:alternate_names
Marge Simpson
Homer Simpson
per:spouse
per:children
per:children
Springfield Elementary
per:schools_attended
Lisa Simpson
Bart Simpson

Entity-Valued Relations

Relation	Inverse(s)
per:children	per:parents
per:other_family	per:other_family
per:parents	per:children
per:siblings	per:siblings
per:spouse	per:spouse
per:employee_of	{org,gpe}:employees*
per:member_of	org:membership*
per:schools_attended	org:students*
per:city_of_birth	gpe:births_in_city*
per:stateorprovince_of_birth	gpe:births_in_stateorprovince*
per:country_of_birth	gpe:births_in_country*
per:cities_of_residence	gpe:residents_of_city*
per:statesorprovinces_of_residence	gpe:residents_of_stateorprovince
per:countries_of_residence	gpe:residents_of_country*
per:city_of_death	gpe:deaths_in_city*
per:stateorprovince_of_death	gpe:deaths_in_stateorprovince*
per:country_of_death	gpe:deaths_in_country*
org:shareholders	{per,org,gpe}:holds_shares_in*
org:founded_by	{per,org,gpe}:organizations_founded*
org:top_members_employees	per:top_member_employee_of*
{org,gpe}:member_of	org:members
org:members	{org,gpe}:member_of
org:parents	{org,gpe}:subsidiaries
org:subsidiaries	org:parents
org:city_of_headquarters	gpe:headquarters_in_city*
org:stateorprovince_of_headquarters	gpe:headquarters_in_stateorprovince*
org:country_of_headquarters	gpe:headquarters_in_country*

String-Filled Relations

per:alternate_names

per:date_of_birth

per:age

per:origin

per:date_of_death

per:cause_of_death

per:title

per:religion

per:charges

org:alternate_names

org:political_religious_affiliation

org:number_of_employees_members

org:date_founded

org:date_dissolved

org:website

Cold Start

Schema

per:children
per:other_family
per:parents
per:siblings
per:spouse
per:employee_of
per:member_of
per:schools_attended
per:city_of_birth
per:stateorprovince_of_birth
per:country_of_birth
per:cities_of_residence
per:statesorprovinces_of_residence
per:countries_of_residence
per:city_of_death
per:stateorprovince_of_death
per:country_of_death
org:shareholders
org:founded_by

The Task

You are given:

Schema

per:children
per:other_family
per:parents
per:siblings
per:spouse
per:employee_of
per:member_of
per:schools_attended
per:city_of_birth
per:stateorprovince_of_birth
per:country_of_birth
per:cities_of_residence
per:statesorprovinces_of_residence
per:countries_of_residence
per:city_of_death
per:stateorprovince_of_death
per:country_of_death

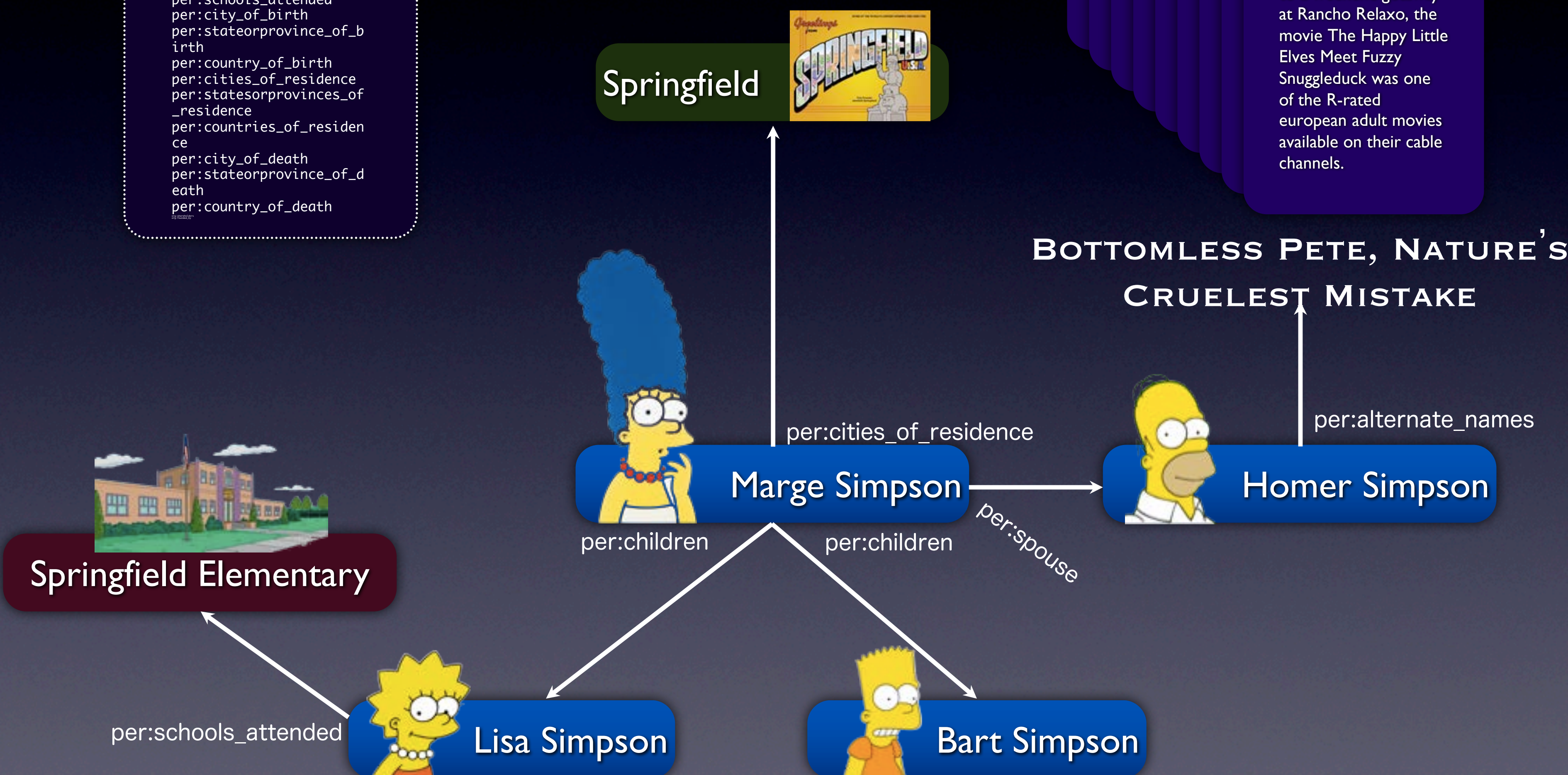
When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated european adult movies available on their cable channels.

Schema

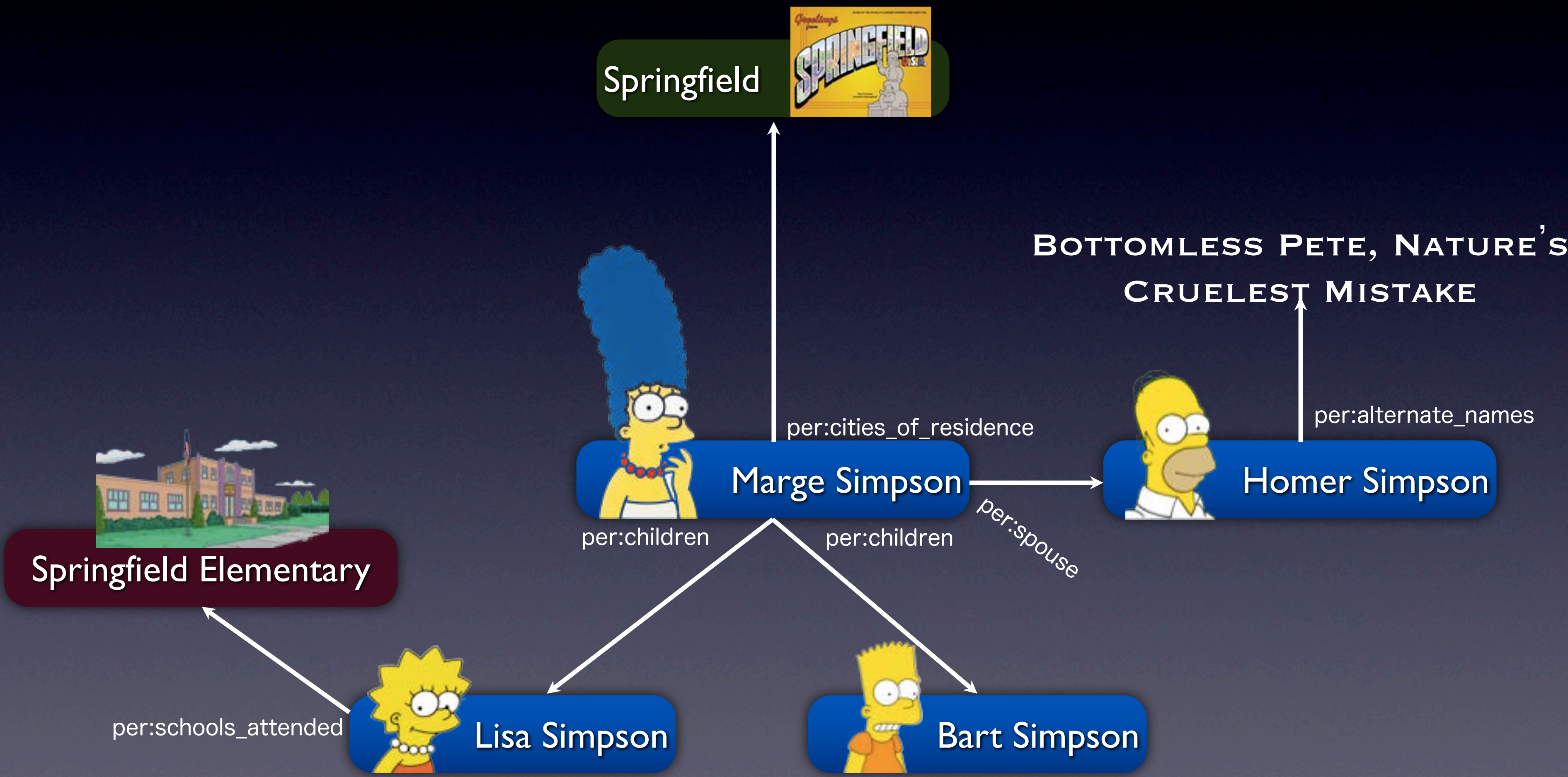
per:children
per:other_family
per:parents
per:siblings
per:spouse
per:employee_of
per:member_of
per:schools_attended
per:city_of_birth
per:stateorprovince_of_birth
per:country_of_birth
per:cities_of_residence
per:statesorprovinces_of_residence
per:countries_of_residence
per:city_of_death
per:stateorprovince_of_death
per:country_of_death

You Must Produce:

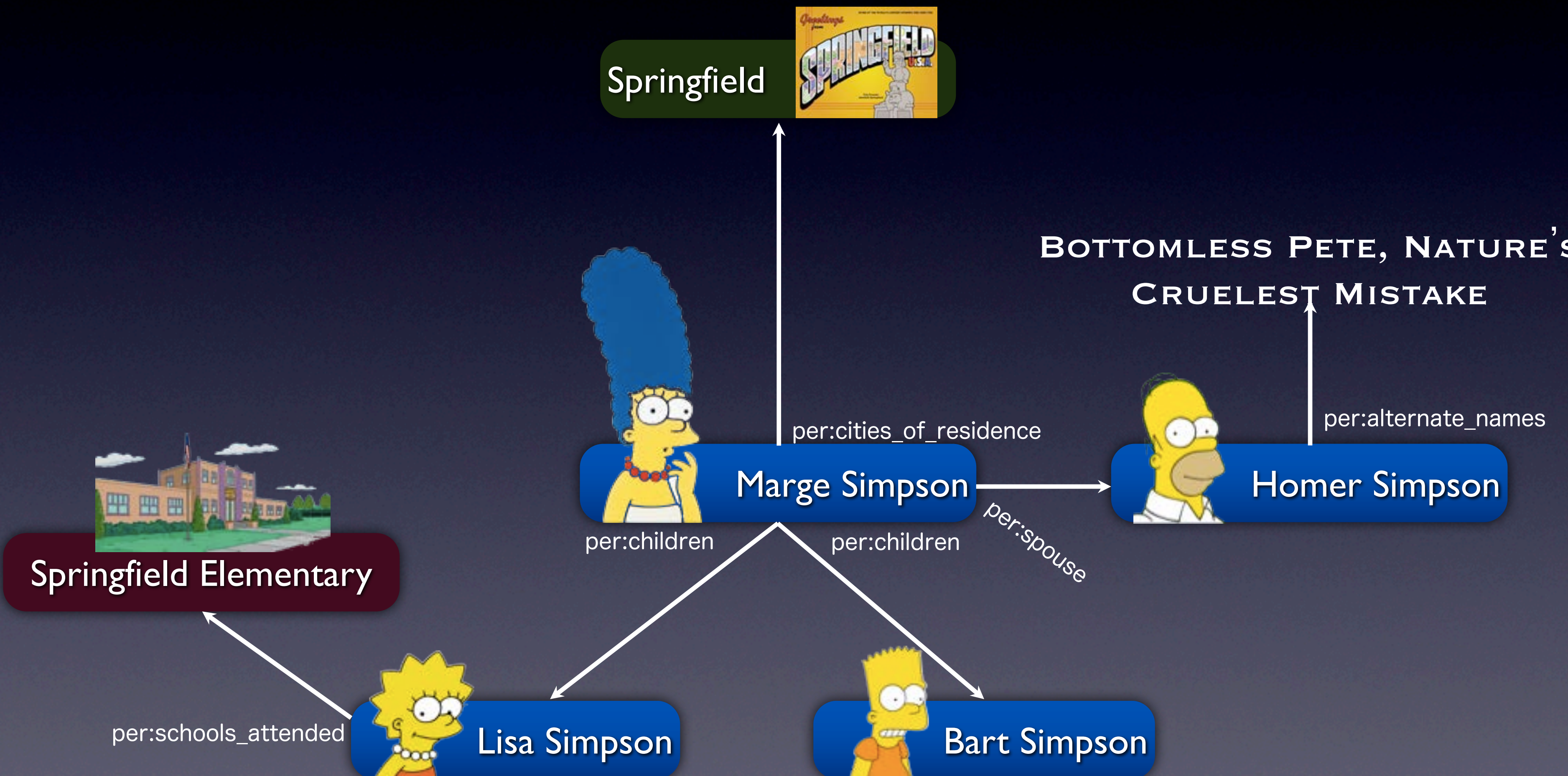
When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated european adult movies available on their cable channels.



How do you know that your KB is any good?



How do you know that your KB is any good? Align it to a ground truth KB

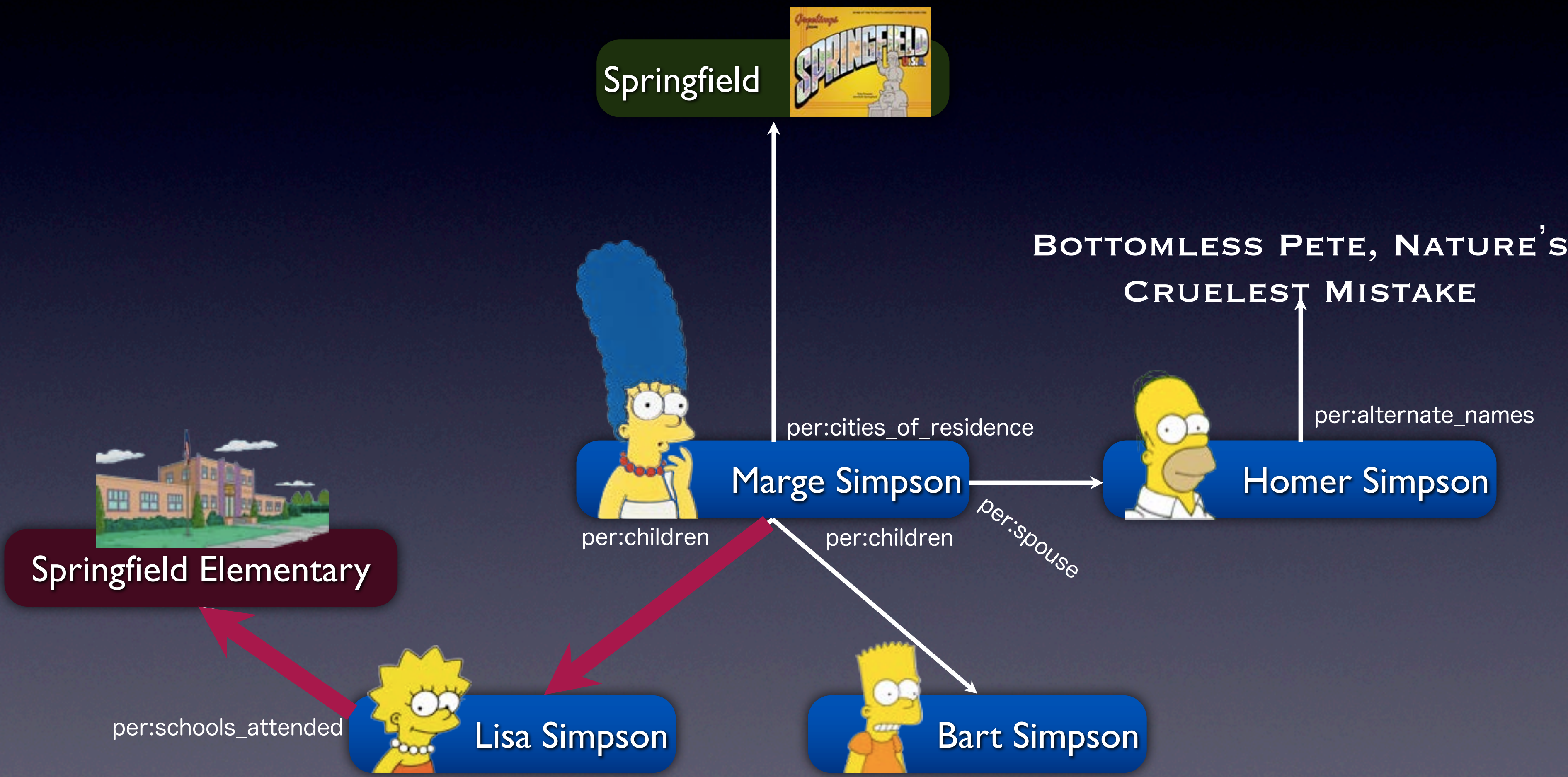


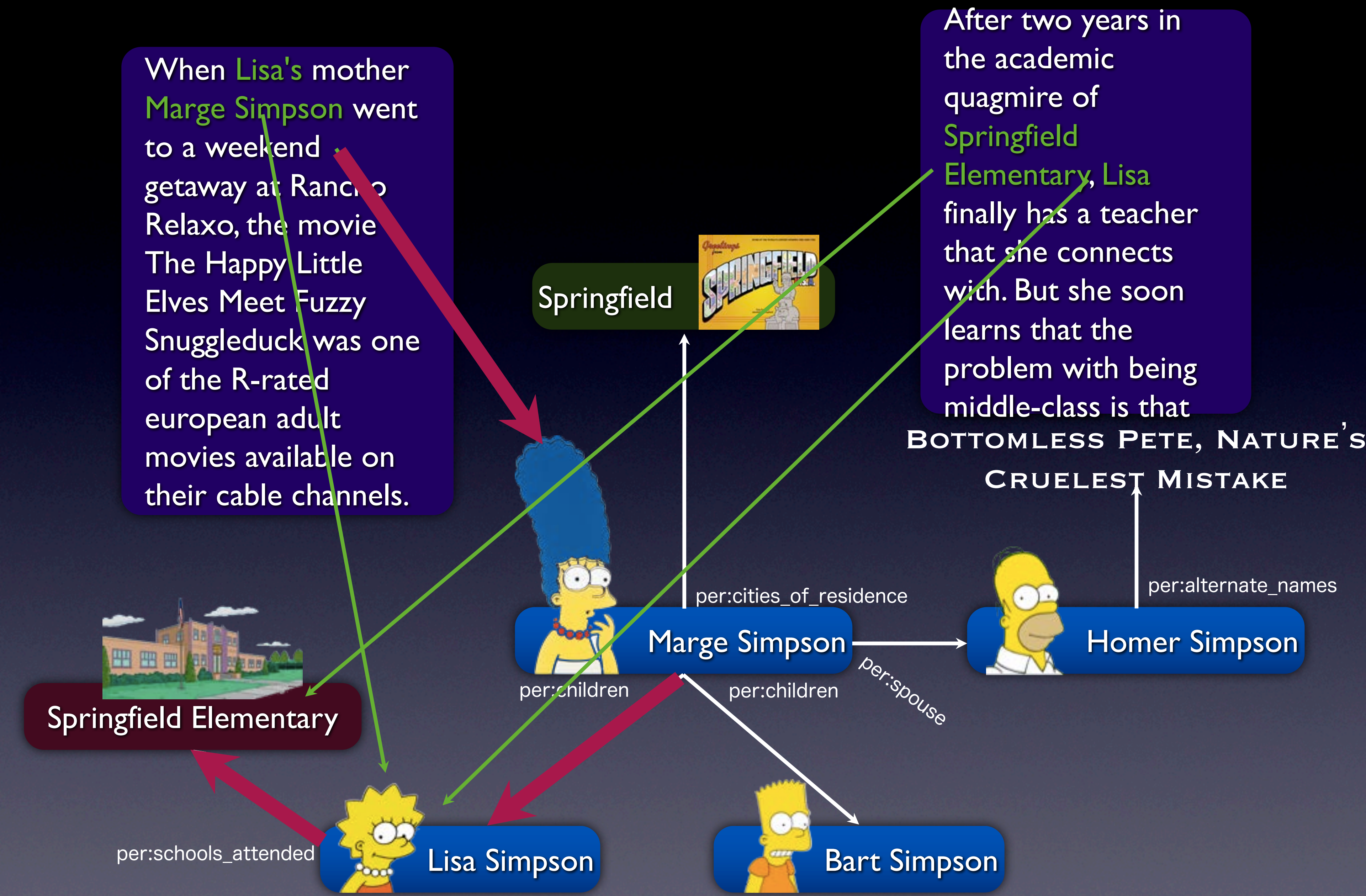
How do you know that your KB is any good?

Align it to a ground truth KB

Where did the children of Marge Simpson go to school?

per:children
per:schools_attended




When Lisa's mother Marge Simpson went to a weekend getaway at Rancho Relaxo, the movie The Happy Little Elves Meet Fuzzy Snuggleduck was one of the R-rated european adult movies available on their cable channels.
After two years in the academic quagmire of Springfield Elementary, Lisa finally has a teacher that she connects with. But she soon learns that the problem with being middle-class is that
Bottomless Pete, Nature's Cruelest Mistake
Springfield
per:cities_of_residence
per:alternate_names
Marge Simpson
Homer Simpson
per:spouse
per:children
per:children
Springfield Elementary
per:schools_attended
Lisa Simpson
Bart Simpson

Sample Evaluation Queries

Query Entity	First Relation	Second Relation
Adriana Petryna	per:title	
Blackstone Group	org:founded_by	
William Shore	per:organizations_founded	org:date_founded
Wistar Institute	org:employees	per:title
Andrew W. Mellon	per:children	per:organizations_founded
Lycee Alliance Israelite Universelle	org:employees	per:schools_attended
Tsitsi Jaji	per:schools_attended	org:students

2016 TAC Cold Start KBP

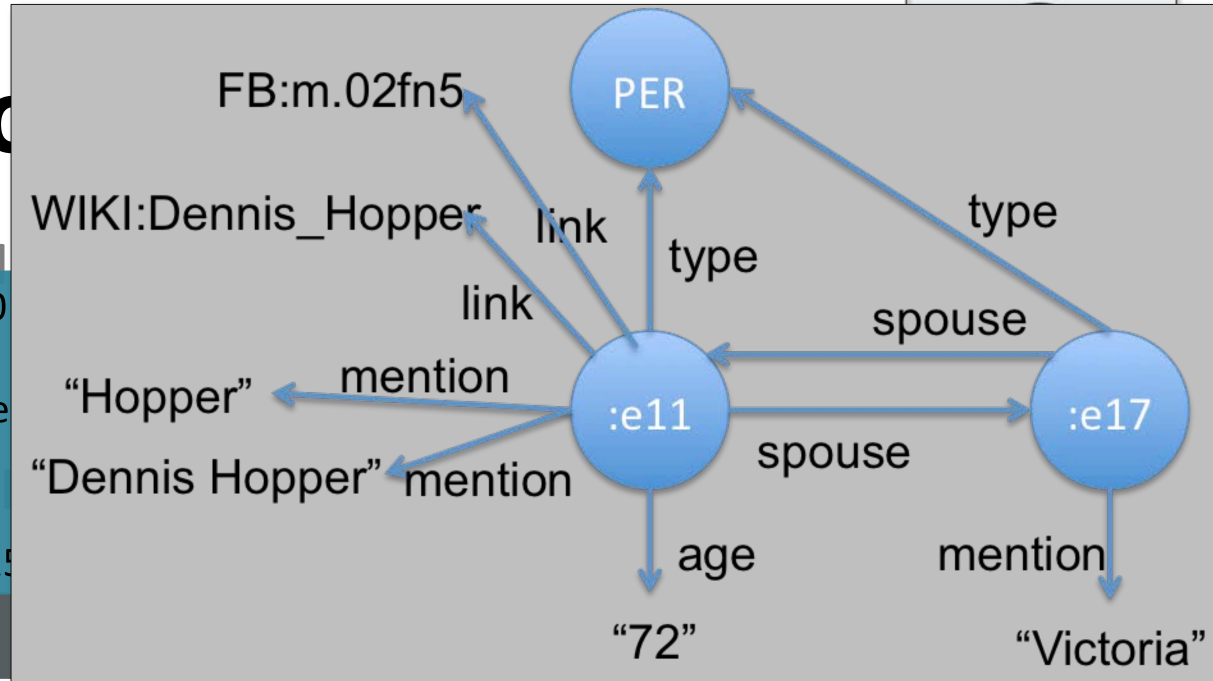
- Read 90K documents: newswire articles & social media posts in English, Chinese and Spanish
- Find entity mentions, types and relations
- Cluster entities within/across documents, link to reference KB when possible (*which George Bush*)
- Remove errors (*Obama born in Illinois*), draw sound inferences (*Malia and Sasha sisters*)
- Create knowledge graph with provenance data for entities, mentions and relations

2016 TAC Co

- Read 90K docu
  m
- Fi
- Cl

<DOC id="APW_ENG_20
<HEADLINE>
Divorce attorney says De
</HEADLINE>
<DATELINE>
LOS ANGELES 2010-03-25
</DATELINE>
<TEXT>
<P>
Dennis Hopper's divorce attorney says in a court filing that the actor is dying and can't undergo chemotherapy as he battles prostate cancer.
</P>
<P>
Attorney Joseph Mannis described the "Easy Rider" actor's condition in a declaration filed Wednesday in Los Angeles Superior Court.
</P>
<P>
Mannis and attorneys for Hopper's wife Victoria are fighting over when and whether to take the actor's deposition.
</P> ...

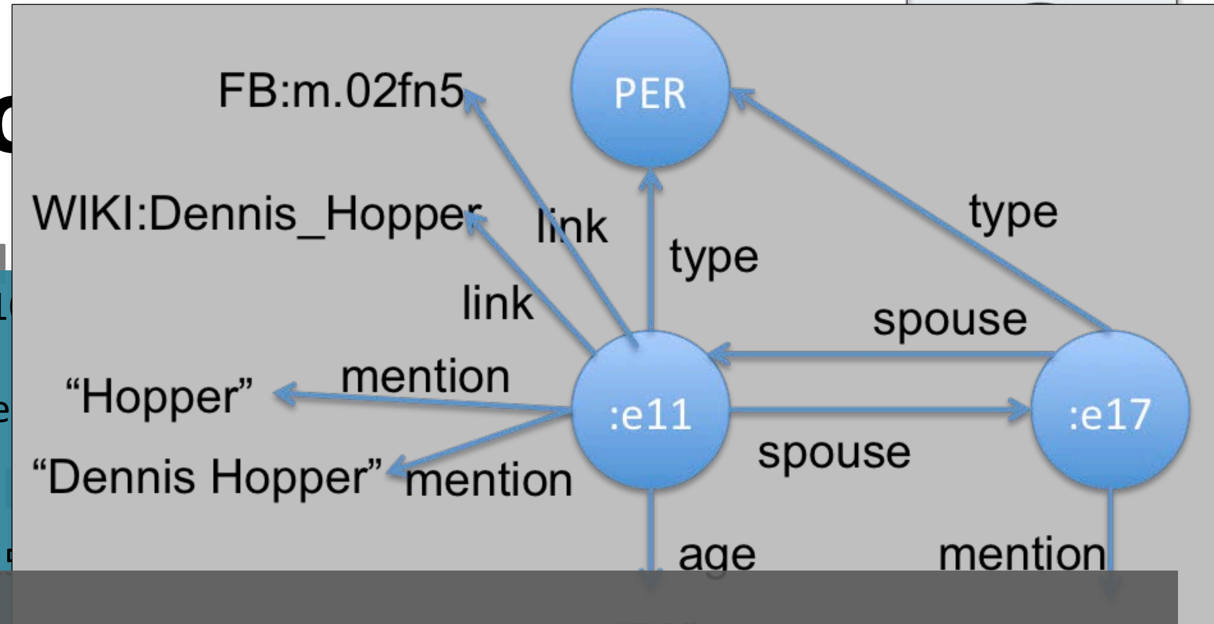
```
…
:e00211 type        PER
:e00211 link        FB:m.02fn5
:e00211 link        WIKI:Dennis_Hopper
:e00211 mention     "Dennis Hopper" APW_021:185-197
:e00211 mention     "Hopper"     APW_021:507-512
:e00211 mention     "Hopper"     APW_021:618-623
:e00211 mention     "丹尼斯·霍珀” CMN_011:930-936
:e00211 per:spouse :e00217      APW_021:521-528
:e00217 per:spouse :e00211      APW_021:521-528
:e00211 per:age     "72"         APW_021:521-528
…
```



2016 TAC Co

- Read 90K docu
  m
- Fi

<DOC id="APW_NG_201
<HEADLINE>
Divorce attorney says De
</HEADLINE>
<DATELINE>
LOS ANGELES 2010-03-2




```
:e00211 a kbp:per;
  kbp:mention "Hopper", "Dennis Hopper";
  kbp:spouse :e00217;
  kbp:age "72";
  kbp:link "m.02fn5"; ... .
[ ] a rdf:statement;
  rdf:subject :e00211;
  rdf:predicate "kbp:mention"
  rdf:object "Hopper";
  kbp:document "APW_021";
  kbp:provenance "APW_021:507-512", "APW_021:618-623".
```

…

KB Evaluation Methodology

- Evaluating KBs extracted from 90K documents is non-trivial
- TAC's approach is simplified by:
  - **Fixing the ontology** of entity types and relations
  - Specifying **a serialization** as triples + provenance
  - Sampling a KB using a set of **queries** grounded in an *entity mention* found in a document
- Given a KB, we can evaluate its **precision and recall** for a set of queries

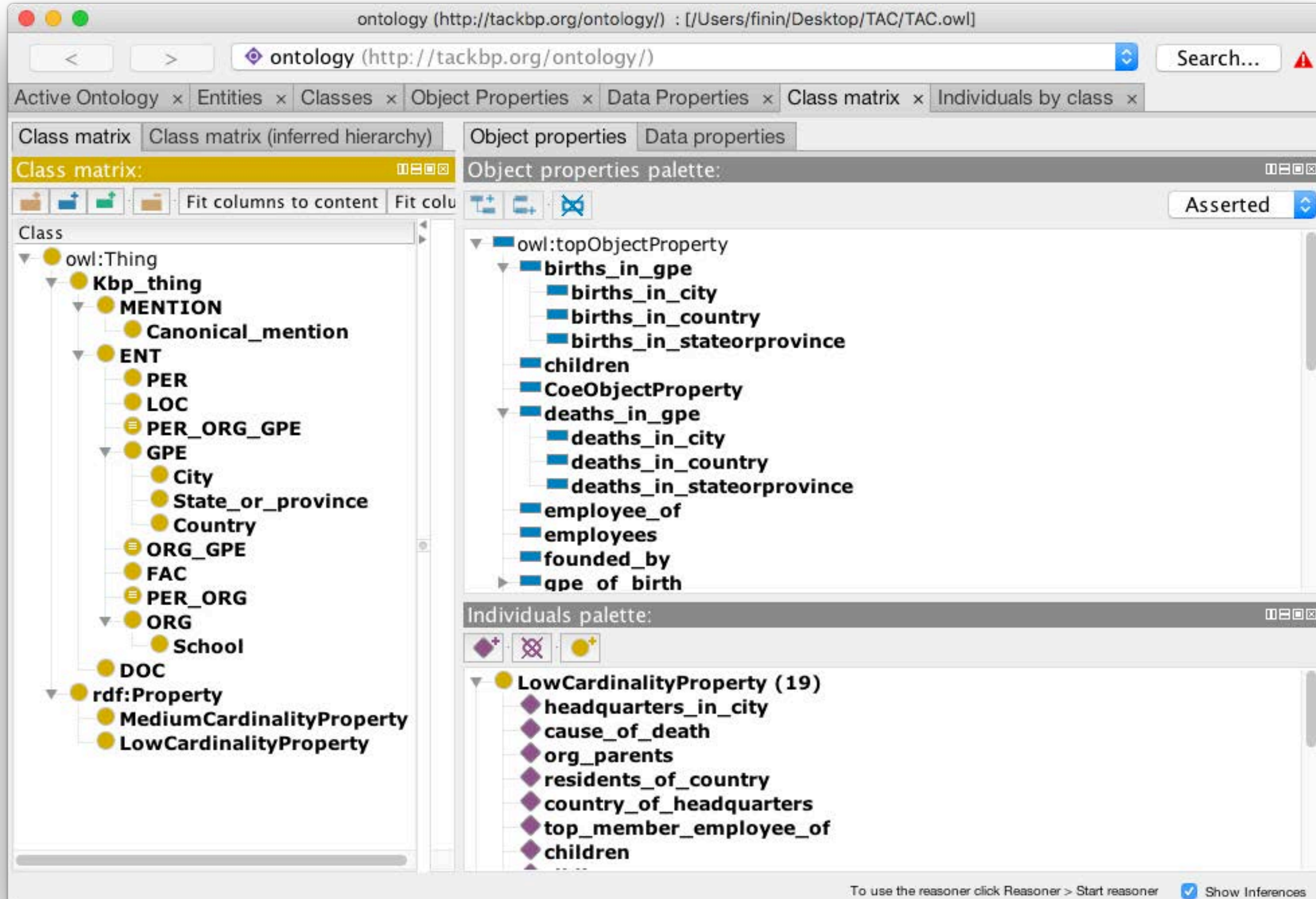
KB Evaluation Methodology

- **A query:** What are the names of schools attended by the children of the entity mentioned in document #45611 at characters 401-412
  - That mention is *George Bush* and the document context suggests it refers to the 41st U.S. president
  - Query given in structured form using TAC ontology
- **Assessors** determine good answers in corpus and check submitted results using their provenance
- **Answers**: entities for Yale, Harvard, Tulane, UT Austin, Univ. of Virginia, Boston College, ...

TAC Ontology

- Five basic entity types
  - PER: **people** (John Lennon) or groups (Americans)
  - ORG: **organizations** like IBM, MIT or US Senate
  - GPE: **geopolitical** entity like Boston, Belgium or Europe
  - LOC: **locations** like Lake Michigan or the Rockies
  - FAC: **facilities** like BWI or the Empire State Building
- Entity Mentions
  - **Strings** referencing entities by name (Barack Obama), description (the President) or pronoun (his)
- ~65 relations
  - Relations hold between two entities: parent_of, spouse, employer, founded_by, city_of_birth, ...
  - Or between an entity & string: age, website, title, cause_of_death, ...

TAC and COE Ontologies



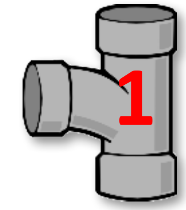
Our ontology has official TAC types/relations and many more we capture from tools and infer from the data

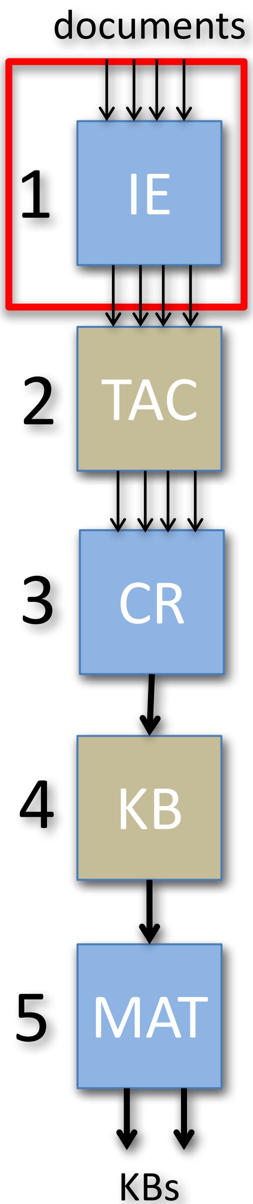
Monlingual
Kelvin

Kelvin

- **KELVIN**: **K**nowledge **E**xtraction, **L**inking, **V**alidation and **In**ference
- Developed at the *Human Language Technology Center of Excellence* at JHU and used in TAC KBP (2010-17), EDL (2015-17) and other projects
- Takes English, Chinese & Spanish documents and produce a knowledge graph in several formats
- We'll review its monolingual processing, look at the multi-lingual use case

1 Information Extraction

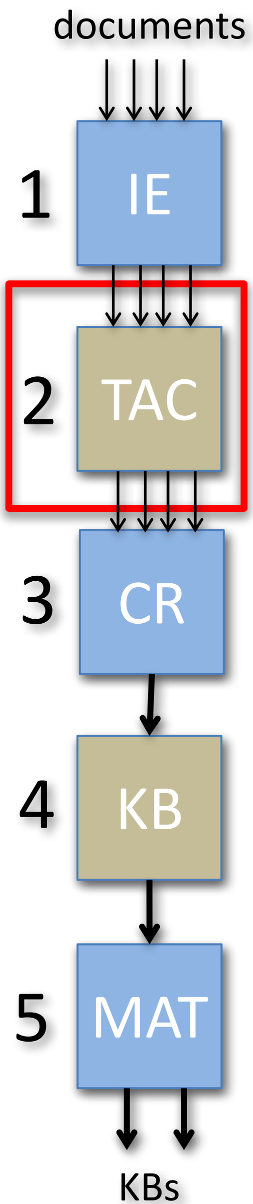




- Process documents in **parallel** on a grid, applying information extraction tools to find mentions, entities, relations and events

- Produce an **Apache Thrift** object for each document with text and relevant data produced by tools using a common **Concrete** schema for NLP data

2 Integrating NLP data

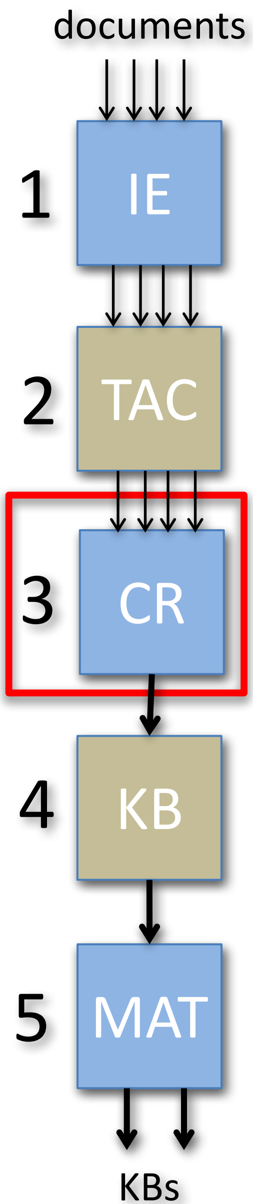


Process Concrete objects in parallel to:

- **Integrate** data from tools (e.g., Stanford, Serif)
- **Fix problems**, e.g., trim mentions, find missed mentions, deconflict tangled mention chains, ...
- Extract relations from **events** (life.born => date and place of birth)
- Map relations found by open IE systems to TAC ontology (*"is engineer at" => per:employee_of)*
- Map schema to extended **TAC ontology**

30K ENG: 430K entities; 1.8M relations

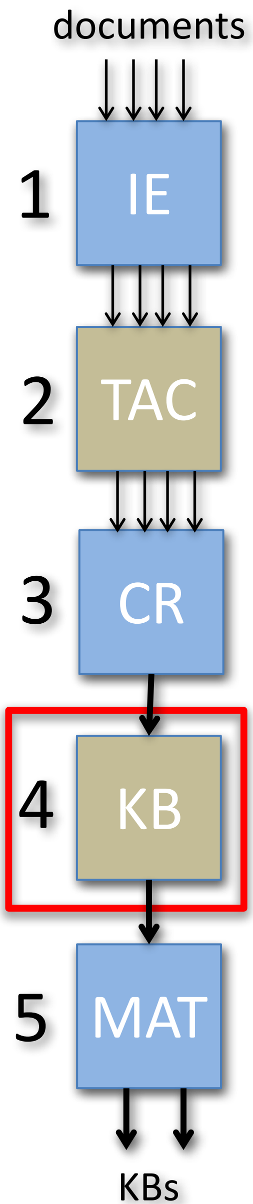
3 Kripke: Cross-Doc Coref



- Cross-document **co-reference** creates initial KB from a set of single-document KBs
  - Identify that *Barack Obama* entity in DOC32 is same individual as *Obama* in DOC342, etc.
- **Language agnostic**; works well for ENG, CMN, SPA document collections
- Uses entity **type** and **mention strings** and context of co-mentioned entities
- Untrained, agglomerative **clustering**

30K ENG: 210K entities; 1.2M relations

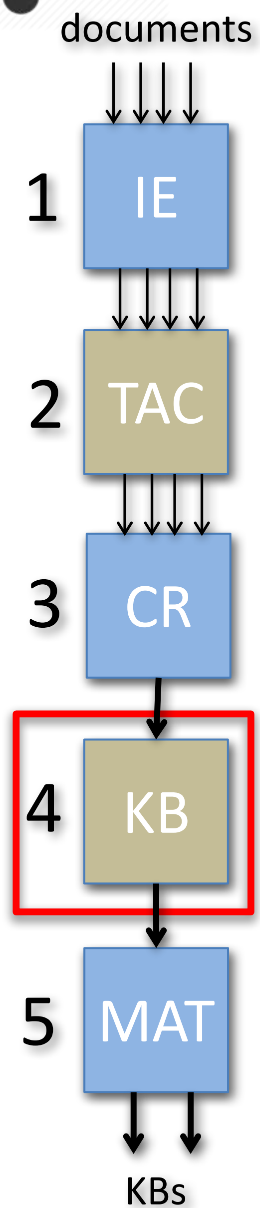
4 Inference and adjudication



Reasoning to

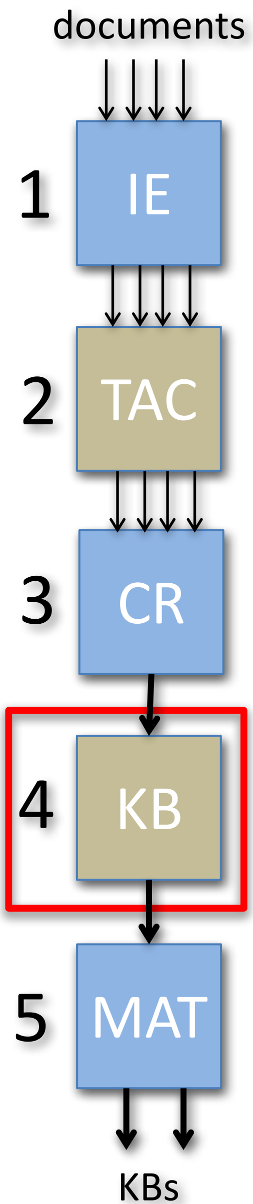
- Delete relations violating ontology constraints
  – *Person can't be born in an organization*
  – *Person can't be her own parent or spouse*
- Infer missing relations
  – *Two people sharing a parent are siblings*
  – *X born in place P_1, P_1 part of P_2 => X born in P_2*
  – *Person probably citizen of their country of birth*
  – *A CFO is a per:top_level_employee*

Entity Linking



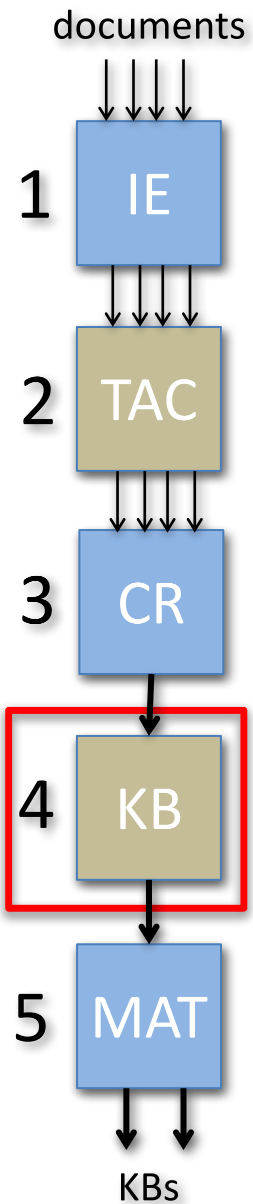
- Try to links entities to reference KB, a subset of Freebase with
  - ~4.5M entities and ~150M triples
  - Names and text in English, Spanish and Chinese
- Don't link if no matches, poor matches or ambiguous matches

KB-level merging rules



- Merge entities of same type linked to same KB entity
- Merge cities in same region with same name
- Highly discriminative relations give evidence of sameness
  - per:spouse is few to few
  - org:top_level_employee is few to few
- Merge PERs with similar names who were
  - Both married to the same person, or
  - Both CEOs of the same company, or …

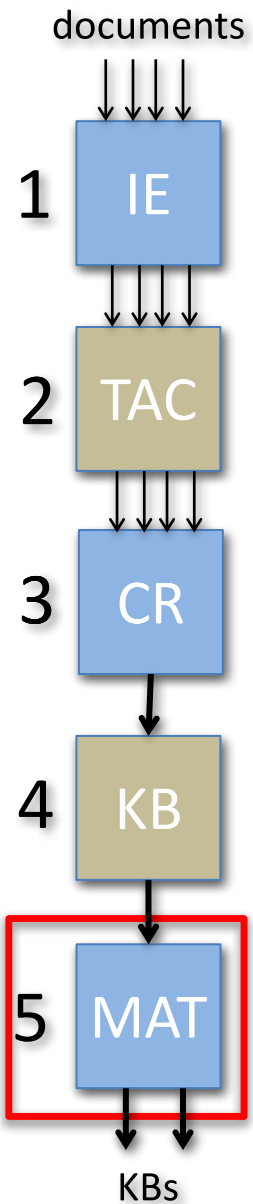
Slot Value Consolidation



- **Problem:** too many values for some slots, especially for 'popular' entities, e.g.,
  - An entity with four different *per:age* values
  - Obama had ~100 *per:employee_of* values
- **Strategy:** rank values and select best
  - Rank values by # of attesting docs and probability
  - Choose best N value depending on relation type

30K ENG: 183K entities; 2.1M relations

Materialize KB versions



- Encode KB in your favorite database or graph store

- We use the RDF/OWL Semantic Web technology stack

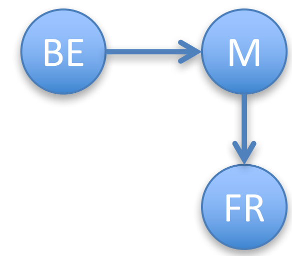
Multi-lingual Kelvin

Multilingual KBP

- Many examples where facts from different languages combine to answer queries or support inference

  Q: Who lives in the same city as *Bodo Elleke*?

  A: *Frank Ribery* aka *Franck Ribéry* aka *里贝里*



- Why we know both live in Munich:

  1. :e8 gpe:residents_of_city :e23 ENG_3:3217-3235

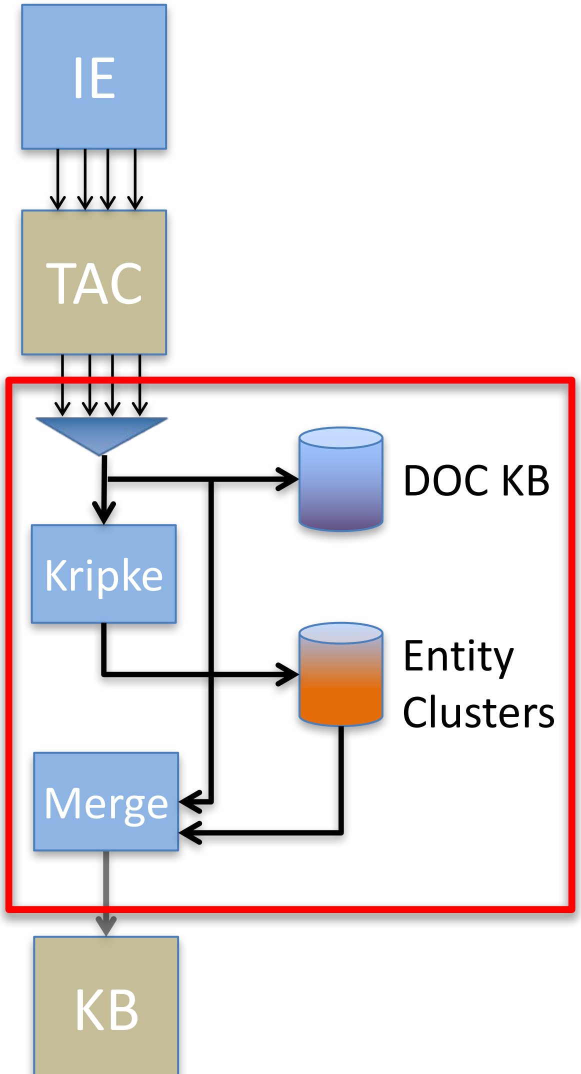
     ...said the younger **Bodo Elleke**, who was born in Schodack in 1930 and is now a retired architect **who lives in Munich**.

  2. :e8 gpe:residents_of_city :e25 CMN...0UTJ:292-361

     拉霍伊在接受西班牙国家电台的采访时肯定，今年的三位金球奖热门候选人中，梅西“度过了一个出色的赛季”，而拜仁**慕尼黑球员里贝里**则“赢得了一切”

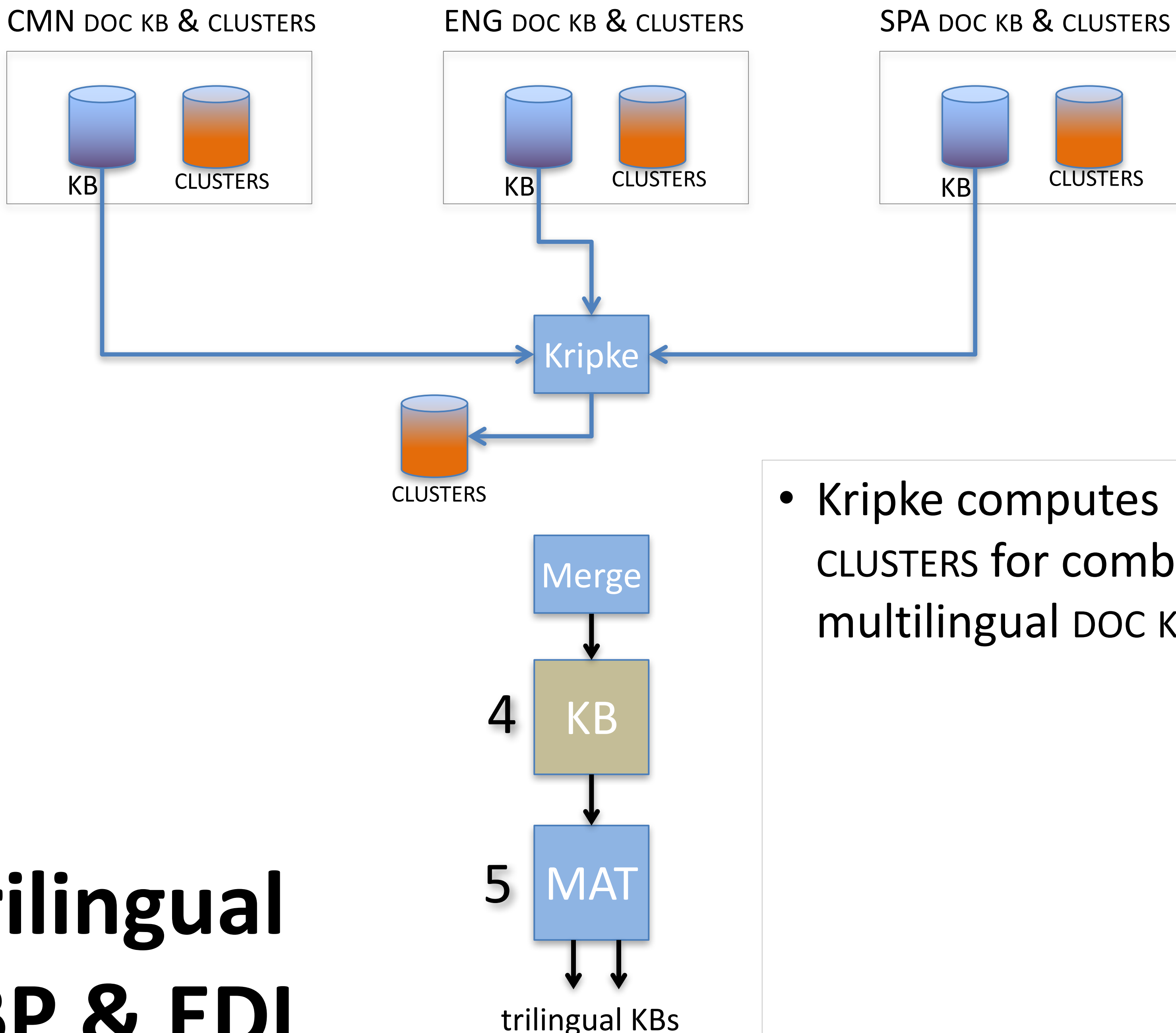
- Kripke merged entities with mentions *Frank Ribery*, *Franck Ribéry* & *里贝里*

Monolingual to Multilingual Kelvin



Zoom in on our cross-doc co-ref step

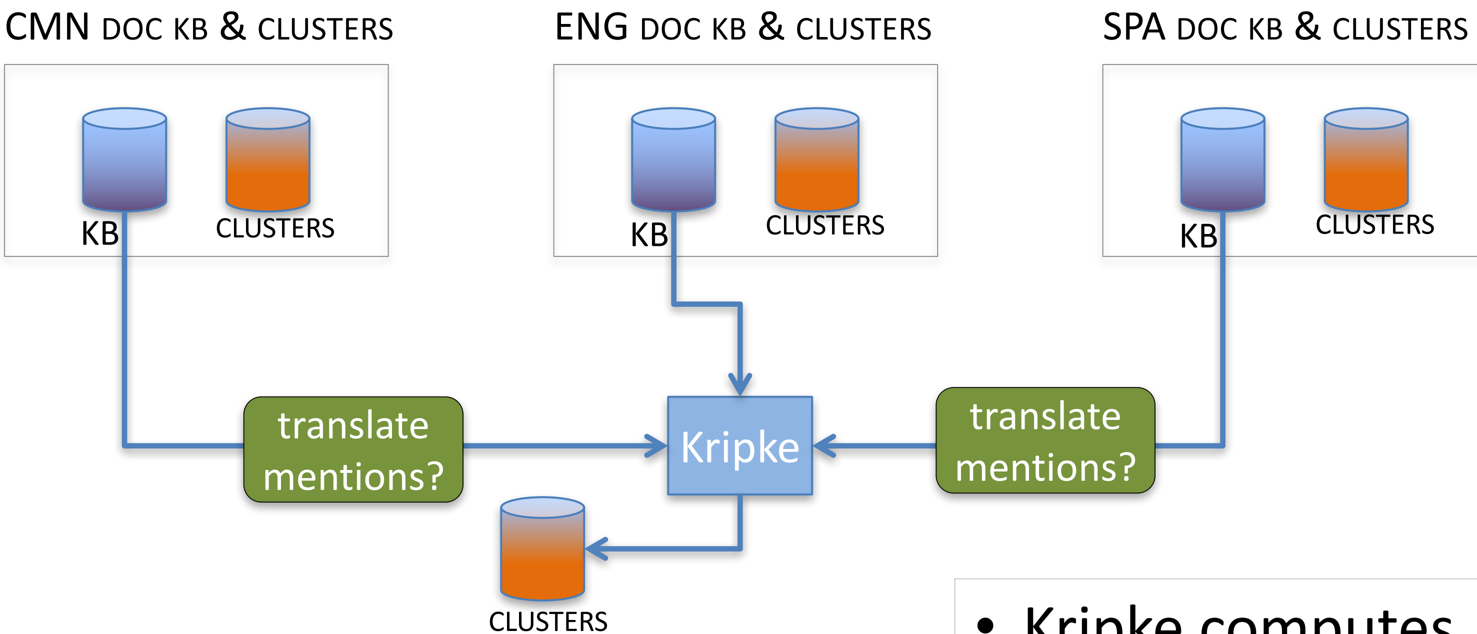
- Concatenate document-level KBs to form a **DOC KB** as input to Kripke
- Kripke outputs a set of **CLUSTERS** defining an equivalence relation
- Merger uses **CLUSTERS** to combine **DOC KB** entities, yielding the initial KB
- We use the **DOC KB** and **CLUSTERS** from each language to create an initial multilingual KB

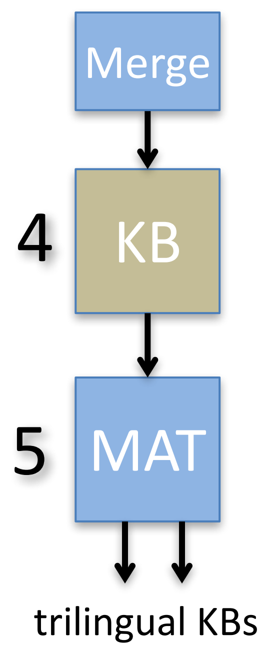


- Kripke computes CLUSTERS for combined multilingual DOC KBs

Trilingual KBP & EDL

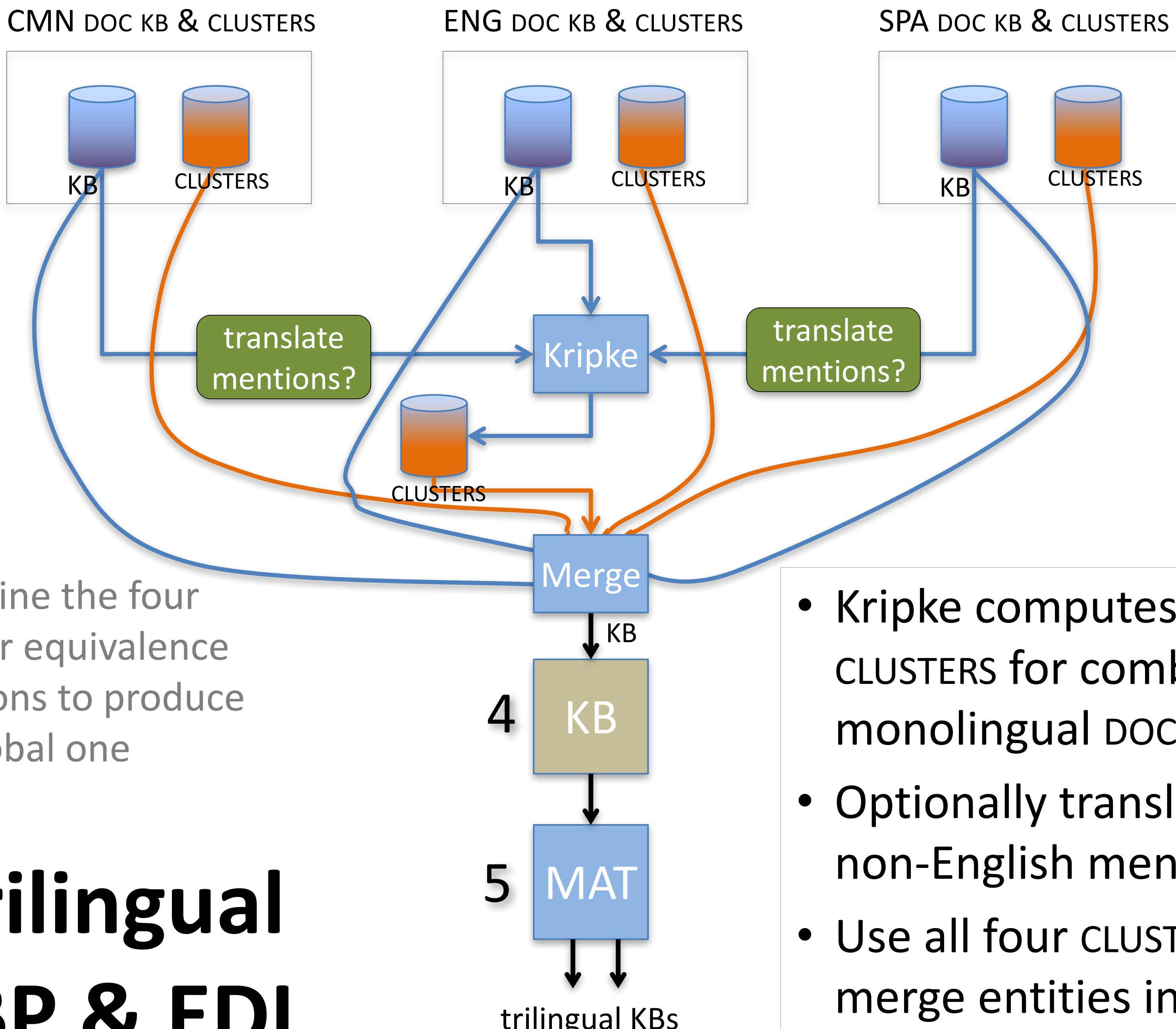
Trilingual KBP & EDL





Translating non-English mentions to English, when possible enhances clustering

- Kripke computes CLUSTERS for combined monolingual DOC KBS
- Optionally translate non-English mentions



Combine the four cluster equivalence relations to produce on global one

Trilingual KBP & EDL

- Kripke computes CLUSTERS for combined monolingual DOC KBS
- Optionally translate non-English mentions
- Use all four CLUSTERS to merge entities in the three DOC KBS

Results

2016 TAC KBP Results

- For the 7 KB and 11 SF submissions, depending on metric (macro/micro avg), we placed
  - 1st or 2nd of 5 on XLING and were the only team to do all three languages
  - 2nd or 4th of 18 on ENG depending on metric
  - 1st or 2nd of 4 on CMN depending on metric
  - We did poorly on SPA, finding few relations
- See workshop paper for details
- TAC EDL results are forthcoming

2016 TAC KBP Results

Run	0-hop				1-hop				All-hop			
	GT	R	W	D	GT	R	W	D	GT	R	W	D
E1	801	194	367	16	408	47	489	2	1209	241	856	18
E2	801	185	313	16	408	41	356	2	1209	226	669	18
E3	801	186	311	13	408	41	362	4	1209	227	673	17
E4	801	186	393	12	408	40	919	2	1209	226	1312	14
E5	801	195	366	15	408	47	489	2	1209	242	855	17
C1	751	141	125	16	230	57	134	10	981	198	259	26
C2	751	141	125	16	230	57	134	10	981	198	259	26
C3	751	141	125	16	230	57	134	10	981	198	259	26
C4	751	143	127	14	230	57	139	10	981	200	266	24
S1	332	4	9	2	213	0	0	0	545	4	9	2
S2	332	4	9	2	213	0	0	0	545	4	9	2
S3	332	3	3	1	213	0	0	0	545	3	3	1
S4	332	4	9	2	213	0	0	0	545	4	9	2
X1	4094	838	1177	96	1965	351	1956	52	6059	1189	3133	148
X2	4094	713	1107	110	1965	272	1196	62	6059	985	2303	172
X3	4094	697	936	116	1965	264	1021	65	6059	961	1957	181
X4	4094	321	385	33	1965	106	490	17	6059	427	875	50

Ground-truth, right, wrong & duplicate answers for 2016 KBP KB runs

2016 TAC KBP Results

	0-hop			1-hop			All-hop		
Run	P	R	F1	P	R	F1	P	R	F1
E1	0.3458	0.2422	0.2849	0.0877	**0.1152**	0.0996	0.2197	0.1993	0.2090
E2	0.3715	0.2310	0.2848	**0.1033**	0.1005	**0.1019**	**0.2525**	0.1869	0.2148
E3	**0.3742**	0.2322	**0.2866**	0.1017	0.1005	0.1011	0.2522	0.1878	**0.2153**
E4	0.3212	0.2322	0.2696	0.0417	0.0980	0.0585	0.1469	0.1869	0.1645
E5	0.3476	**0.2434**	0.2863	0.0877	**0.1152**	0.0996	0.2206	**0.2002**	0.2099
C1	**0.5301**	0.1877	0.2773	**0.2984**	**0.2478**	**0.2708**	**0.4333**	0.2018	0.2754
C2	0.5301	0.1877	0.2773	0.2984	0.2478	0.2708	0.4333	0.2018	0.2754
C3	0.5301	0.1877	0.2773	0.2984	0.2478	0.2708	0.4333	0.2018	0.2754
C4	0.5296	**0.1904**	**0.2801**	0.2908	0.2478	0.2676	0.4292	**0.2039**	**0.2764**
S1	0.3077	**0.0120**	**0.0232**	0.0000	0.0000	0.0000	0.3077	**0.0073**	**0.0143**
S2	0.3077	0.0120	0.0232	0.0000	0.0000	0.0000	0.3077	0.0073	0.0143
S3	**0.5000**	0.0090	0.0178	0.0000	0.0000	0.0000	**0.5000**	0.0055	0.0109
S4	0.3077	0.0120	0.0232	0.0000	0.0000	0.0000	0.3077	0.0073	0.0143
X1	0.4159	**0.2047**	**0.2743**	0.1521	**0.1786**	**0.1643**	0.2751	**0.1962**	**0.2291**
X2	0.3918	0.1742	0.2411	0.1853	0.1384	0.1585	0.2996	0.1626	0.2108
X3	0.4268	0.1702	0.2434	**0.2054**	0.1344	0.1625	**0.3293**	0.1586	0.2141
X4	**0.4547**	0.0784	0.1337	0.1779	0.0539	0.0828	0.3280	0.0705	0.1160

Micro precision, recall and F1 scores for 2016 KBP KB runs

2016 EDL XLING Results

	NER			Linking			Nil Linking			Clustering		
#	pre	rec	F1	pre	rec	F1	pre	rec	F1	pre	rec	F1
1	0.656	0.573	0.612	0.489	0.427	0.456	0.368	0.614	0.460	0.470	0.411	0.438
2	0.656	0.573	0.612	0.488	0.426	0.455	0.335	0.624	0.436	0.469	0.410	0.437
3	0.656	0.573	0.612	0.476	0.416	0.444	0.346	0.615	0.443	0.457	0.399	0.426
4	0.656	0.573	0.612	0.489	0.427	0.456	0.335	0.625	0.436	0.469	0.410	0.438
5	0.661	0.563	0.608	0.494	0.420	0.454	0.374	0.612	0.464	0.474	0.404	0.436

XLING run precision, recall and F1 measures for four key metrics: strong typed mention match (NER), strong all match (Linking), strong nil match (Nil), and mention ceaf plus (Clustering)

2016 Results Observations

- Overall XLING1 was best
- Variations for monolingual runs were similar
  - Using translated mentions for non-English helped
  - Using nominal mentions seemed to improve cross-doc co-ref slightly
- EDL scores (and maybe KBP) lowered by bug in our nominal mention trimming code; the nominal strings correctly identified but offsets were wrong ☹

Kelvin Docker Container

- **Problem:** Kelvin is a large and complex system that's difficult to port to a new Unix environment, let alone a different OS
- **Solution:** We use Docker to virtualize Kelvin as several containers that can be run on any system that supports Docker
  - e.g., most Unix systems, Mac OSX and Windows

Conclusion

Lessons Learned

- We always have to mind precision & recall
- Extracting information from text is inherently noisy; reading more text helps both
- Using machine learning at every level is important
- Making more use of probabilities will help
- Extracting information about a events is hard
- Recognizing the temporal extent of relations is important, but still a challenge

Conclusion

- KBs help in extracting information from text
- The information extracted can update the KBs
- The KBs provide support for new tasks, such as question answering and speech interfaces
- We'll see this approach grow and evolve in the future
- New machine learning frameworks will result in better accuracy