Table Retrieval and Generation

Krisztian Balog
University of Stavanger
@krisztianbalog

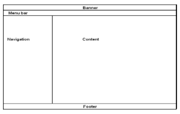
Information Access
& Interaction
http://iai.group

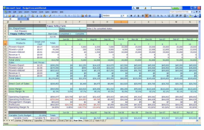
SIGIR'18 workshop on Data Search (DATA:SEARCH'18) | Ann Arbor, Michigan, USA, July 2018

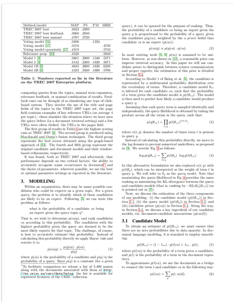
JOIN WORK WITH

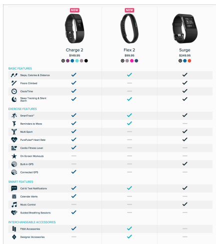
Shuo Zhang
@imsure318

TABLES ARE EVERYWHERE



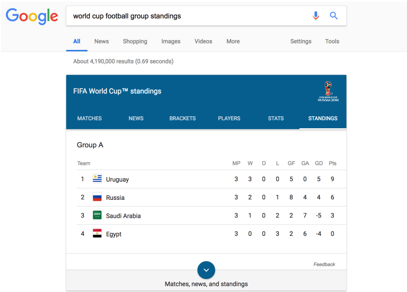






MOTIVATION

MOTIVATION



IN THIS TALK

- Three retrieval tasks, with tables as results
  - Ad hoc table retrieval
  - Query-by-table
  - On-the-fly table generation

THE ANATOMY OF A RELATIONAL (ENTITY-FOCUSED) TABLE

Formula 1 constructors' statistics 2016

Constructor	Engine	Country	Base	…
Ferrari	Ferrari	Italy	Italy	
Force India	Mercedes	India	UK	
Haas	Ferrari	US	US & UK	
Manor	Mercedes	UK	UK	
…				

THE ANATOMY OF A RELATIONAL (ENTITY-FOCUSED) TABLE

Table caption

Formula 1 constructors' statistics 2016

Constructor	Engine	Country	Base	…
Ferrari	Ferrari	Italy	Italy	
Force India	Mercedes	India	UK	
Haas	Ferrari	US	US & UK	
Manor	Mercedes	UK	UK	
…				

THE ANATOMY OF A RELATIONAL (ENTITY-FOCUSED) TABLE

Core column
(subject column)

Formula 1 constructors' statistics 2016

Constructor	Engine	Country	Base	…
Ferrari	Ferrari	Italy	Italy	
Force India	Mercedes	India	UK	
Haas	Ferrari	US	US & UK	
Manor	Mercedes	UK	UK	
…				

We assume that these entities are recognized and disambiguated, i.e., linked to a knowledge base

THE ANATOMY OF A RELATIONAL (ENTITY-FOCUSED) TABLE

Heading
column labels
(table schema)

Formula 1 constructors' statistics 2016

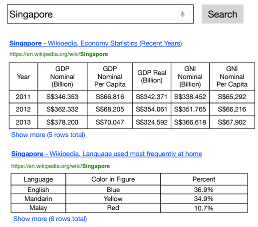
Constructor	Engine	Country	Base	…
Ferrari	Ferrari	Italy	Italy	
Force India	Mercedes	India	UK	
Haas	Ferrari	US	US & UK	
Manor	Mercedes	UK	UK	
…				

AD HOC TABLE RETRIEVAL

S. Zhang and K. Balog. **Ad Hoc Table Retrieval using Semantic Similarity.**
In: *The Web Conference 2018 (WWW '18)*

TASK

- *Ad hoc table retrieval*:
  - Given a keyword query as input, return a ranked list of tables from a table corpus


Singapore Search

[Singapore - Wikipedia, Economy Statistics (Recent Years)]
https://en.wikipedia.org/wiki/Singapore

Year	GDP Nominal (Billion)	GDP Nominal Per Capita	GDP Real (Billion)	GNI Nominal (Billion)	GNI Nominal Per Capita
2011	S$346.353	S$66,816	S$342.371	S$338.452	S$65,292
2012	S$362.332	S$68,205	S$354.061	S$351.765	S$66,216
2013	S$378.200	S$70,047	S$324.592	S$366.618	S$67,902

Show more (5 rows total)

[Singapore - Wikipedia, Language used most frequently at home]
https://en.wikipedia.org/wiki/Singapore

Language	Color in Figure	Percent
English	Blue	36.9%
Mandarin	Yellow	34.9%
Malay	Red	10.7%

Show more (6 rows total)

APPROACHES

- Unsupervised methods
  - Build a document-based representation for each table, then employ conventional document retrieval methods
- Supervised methods
  - Describe query-table pairs using a set of features, then employ supervised machine learning ("learning-to-rank")
  - **Contribution #1**: new state-of-the-art, using a rich set of features
  - **Contribution #2**: new set of semantic matching features

UNSUPERVISED METHODS

- Single-field document representation
  - All table content, no structure
- Multi-field document representation
  - Separate document fields for embedding document's title, section title, table caption, table body, and table headings

SUPERVISED METHODS

- Three groups of features
- **Query features**
  - #query terms, query IDF scores
- **Table features**
  - Table properties: #rows, #cols, #empty cells, etc.
  - Embedding document: link structure, number of tables, etc.
- **Query-table features**
  - Query terms found in different table elements, LM score, etc.
  - Our novel semantic matching features

FEATURES

Query features	
QLEN	Number of query terms
IDF_f	Sum of query IDF scores in field f
Table features	
#rows	The number of rows in the table
#cols	The number of columns in the table
#of NULLs in table	The number of empty table cells
PMI	The ACSDb-based schema coherency score
inLinks	Number of in-links to the page embedding the table
outLinks	Number of out-links from the page embedding the table
pageViews	Number of page views
tableImportance	Inverse of number of tables on the page
tablePageFraction	Ratio of table size to page size
Query-table features	
#hitsLC	Total query term frequency in the leftmost column cells
#hitsSLC	Total query term frequency in second-to-leftmost column cells
#hitsB	Total query term frequency in the table body
qInPgTitle	Ratio of the number of query tokens found in page title to total number of tokens
qInTableTitle	Ratio of the number of query tokens found in table title to total number of tokens
yRank	Rank of the table's Wikipedia page in Web search engine results for the query
MLM similarity	Language modeling score between query and multi-field document repr. of the table

SEMANTIC MATCHING

- Main objective: go beyond term-based matching
- Three components:
  1. Content extraction
  2. Semantic representations
  3. Similarity measures

SEMANTIC MATCHING 1. CONTENT EXTRACTION

- The "raw" content of a query/table is represented as a set of terms, which can be words or entities

SEMANTIC MATCHING 1. CONTENT EXTRACTION

- The "raw" content of a query/table is represented as a set of terms, which can be words or entities

Entity-based:

- Top-k ranked entities from a knowledge base

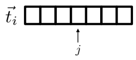
- Entities in the core table column
- Top-k ranked entities using the embedding document/section title as a query

SEMANTIC MATCHING 2. SEMANTIC REPRESENTATIONS

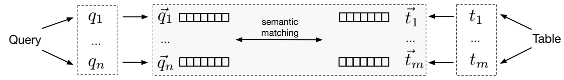
- Each of the raw terms is mapped to a semantic vector representation

SEMANTIC REPRESENTATIONS

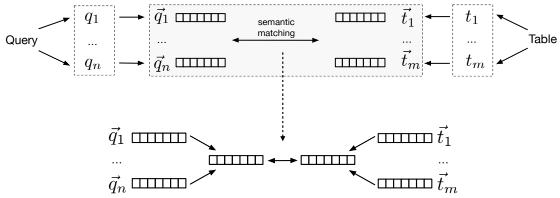
- Bag-of-concepts (sparse discrete vectors)
  - **Bag-of-entities**
    - Each vector element corresponds to an entity
    - $\vec{t}_i[j]$ is 1 if there exists a link between entities i and j in the KB
  - **Bag-of-categories**
    - Each vector element corresponds to a Wikipedia category
    - $\vec{t}_i[j]$ is 1 if entity i is assigned to Wikipedia category j
- Embeddings (dense continuous vectors)
  - **Word embeddings**
    - Word2Vec (300 dimensions, trained on Google news)
  - **Graph embeddings**
    - RDF2vec (200 dimensions, trained on DBpedia)



SEMANTIC MATCHING
3. SIMILARITY MEASURES

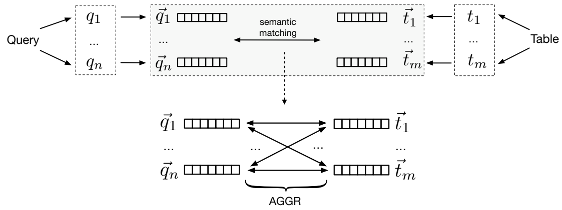


SEMANTIC MATCHING
EARLY FUSION MATCHING STRATEGY



Early: Take the centroid of semantic vectors and compute their cosine similarity

SEMANTIC MATCHING
LATE FUSION MATCHING STRATEGY



Late: Compute all pairwise similarities between the query and table semantic vectors, then aggregate those pairwise similarity scores (sum, avg, or max)

EXPERIMENTAL EVALUATION

EXPERIMENTAL SETUP

- Table corpus
  - WikiTables corpus[1]: 1.6M tables extracted from Wikipedia
- Knowledge base
  - DBpedia (2015-10): 4.6M entities with an English abstract
- Queries
  - Sampled from two sources[2,3]
- Rank-based evaluation
  - NDCG@5, 10, 15, 20

QS-1	QS-2
video games	asian countries currency
us cities	laptops cpu
kings of africa	food calories
economy gdp	guitars manufacturer

[1] Bhagavatula et al. **TabEL: Entity Linking in Web Tables.** In: *ISWC '15.*
[2] Cafarella et al. **Data Integration for the Relational Web.** *Proc. of VLDB Endow.* (2009)
[3] Venetis et al. **Recovering Semantics of Tables on the Web.** *Proc. of VLDB Endow.* (2011)

RELEVANCE ASSESSMENTS

- Collected via crowdsourcing
  - Pooling to depth 20, 3120 query-table pairs in total
- Assessors are presented with the following scenario
  - "Imagine that your task is to create a new table on the query topic"
  - A table is …
    - **Non-relevant** (0): if it is unclear what it is about or it about a different topic
    - **Relevant** (1): if some cells or values could be used from it
    - **Highly relevant** (2): if large blocks or several values could be used from it

RESEARCH QUESTIONS

- **RQ1**: Can semantic matching improve retrieval performance?
- **RQ2**: Which of the semantic representations is the most effective?
- **RQ3**: Which of the similarity measures performs best?

RESULTS: RQ1

	NDCG@10	NDCG@20
Single-field document ranking	0.4344	0.5254
Multi-field document ranking	0.4860	0.5473
WebTable[1]	0.2992	0.3726
WikiTable[2]	0.4766	0.5206
LTR baseline	0.5456	0.6031
STR (LTR + semantic matching)	**0.6293**	**0.6825**

- Can semantic matching improve retrieval performance?
- Yes. STR achieves substantial and significants improvements over LTR.

[1] Cafarella et al. **WebTables: Exploring the Power of Tables on the Web.** *Proc. of VLDB Endow.* (2008)
[2] Bhagavatula et al. **Methods for Exploring and Mining Tables on Wikipedia.** In: *IDEA '13.*

RESULTS: RQ2

Sem. Repr.	Early	Late-max	Late-sum	Late-avg	ALL
Bag-of-entities	0.6754 (+11.99%)	0.6407 (+6.23%)†	0.6697 (+11.04%)‡	0.6733 (+11.64%)‡	0.6696 (+11.03%)‡
Bag-of-categories	0.6287 (+4.19%)	0.6245 (+3.55%)	0.6315 (+4.71%)†	0.6240 (+3.47%)	0.6149 (+1.96%)
Word embeddings	0.6181 (+2.49%)	0.6328 (+4.92%)	0.6371 (+5.64%)†	0.6485 (+7.53%)†	0.6588 (+9.24%)†
Graph embeddings	0.6326 (+4.89%)	0.6142 (+1.84%)	0.6223 (+3.18%)	0.6316 (+4.73%)	0.6340 (+5.12%)
ALL	0.6736 (+11.69%)†	0.6631 (+9.95%)†	0.6831 (+13.26%)‡	0.6809 (+12.90%)‡	0.6825 (13.17%)‡

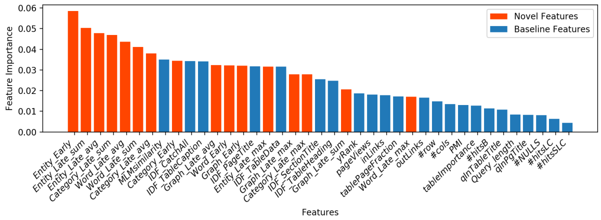
- Which of the semantic representations is the most effective?
- Bag-of-entities.

RESULTS: RQ3

Sem. Repr.	Early	Late-max	Late-sum	Late-avg	ALL
Bag-of-entities	0.6754 (+11.99%)	0.6407 (+6.23%)†	0.6697 (+11.04%)‡	0.6733 (+11.64%)‡	0.6696 (+11.03%)‡
Bag-of-categories	0.6287 (+4.19%)	0.6245 (+3.55%)	0.6315 (+4.71%)†	0.6240 (+3.47%)	0.6149 (+1.96%)
Word embeddings	0.6181 (+2.49%)	0.6328 (+4.92%)	0.6371 (+5.64%)†	0.6485 (+7.53%)†	0.6588 (+9.24%)†
Graph embeddings	0.6326 (+4.89%)	0.6142 (+1.84%)	0.6223 (+3.18%)	0.6316 (+4.73%)	0.6340 (+5.12%)
ALL	0.6736 (+11.69%)†	0.6631 (+9.95%)†	0.6831 (+13.26%)‡	0.6809 (+12.90%)‡	0.6825 (13.17%)‡

- Which of the similarity measures performs best?
- Late-sum and Late-avg (but it also depends on the representation)

FEATURE ANALYSIS



QUERY-BY-TABLE

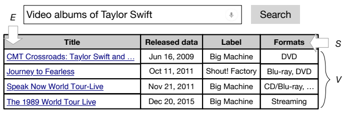
Currently under peer review

ON-THE-FLY TABLE GENERATION

S. Zhang and K. Balog. **On-the-fly Table Generation**.
In: *41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '18)*

TASK

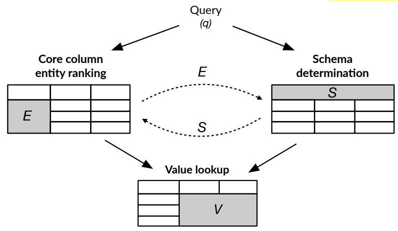
- *On-the-fly table generation*:
  - Answer a free text query with a relational table, where
    - the core column lists all relevant entities;
    - columns correspond to attributes of those entities;
    - cells contain the values of the corresponding entity attributes.



Title	Released data	Label	Formats
CMT Crossroads: Taylor Swift and …	Jun 16, 2009	Big Machine	DVD
Journey to Fearless	Oct 11, 2011	Shout! Factory	Blu-ray, DVD
Speak Now World Tour-Live	Nov 21, 2011	Big Machine	CD/Blu-ray, …
The 1989 World Tour Live	Dec 20, 2015	Big Machine	Streaming

APPROACH

Core column entity ranking and schema determination could potentially mutually reinforce each other.

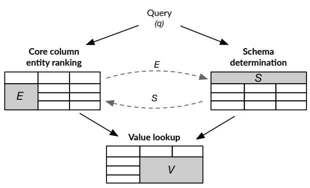


ALGORITHM

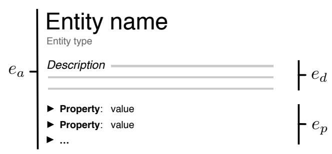
Algorithm 1: Iterative Table Generation

Data: q, a keyword query
Result: $T = (E, S, V)$, a result table

begin
  $E^0 \leftarrow rankEntites(q, \{\})$;
  $S^0 \leftarrow rankLabels(q, \{\})$;
  $t \leftarrow 0$;
  while $\neg terminate$ **do**
    $t \leftarrow t + 1$;
    $E^t \leftarrow rankEntites(q, S^{t-1})$;
    $S^t \leftarrow rankLabels(q, E^{t-1})$;
  end
  $V \leftarrow lookupValues(E^t, S^t)$;
  return (E^t, S^t, V)
end



KNOWLEDGE BASE ENTRY



CORE COLUMN ENTITY RANKING

$$score_t(e, q) = \sum_i w_i \phi_i(e, q, S^{t-1})$$

Feature	Iter. (t)
Term-based matching	
ϕ_1: $LM(q, e_a)$	≥ 0
Deep semantic matching	
ϕ_2: $DRRM\_TKS(q, e_d)$	≥ 0
ϕ_3: $DRRM\_TKS(q, e_p)$	≥ 0
ϕ_4: $DRRM\_TKS(s, e_d)$	≥ 1
ϕ_5: $DRRM\_TKS(s, e_p)$	≥ 1
ϕ_6: $DRRM\_TKS(q \oplus s, e_d \oplus e_p)$	≥ 1
Entity-schema compatibility	
ϕ_7: $ESC(S, e)$	≥ 1

CORE COLUMN ENTITY RANKING FEATURES

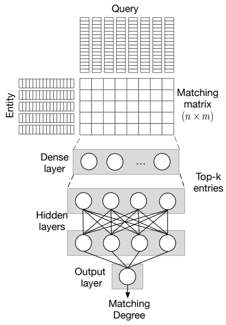
Feature	Iter. (t)
Term-based matching	
ϕ_1: $LM(q, e_a)$	≥ 0
Deep semantic matching	
ϕ_2: $DRRM\_TKS(q, e_d)$	≥ 0
ϕ_3: $DRRM\_TKS(q, e_p)$	≥ 0
ϕ_4: $DRRM\_TKS(s, e_d)$	≥ 1
ϕ_5: $DRRM\_TKS(s, e_p)$	≥ 1
ϕ_6: $DRRM\_TKS(q \oplus s, e_d \oplus e_p)$	≥ 1
Entity-schema compatibility	
ϕ_7: $ESC(S, e)$	≥ 1

Entity's relevance to the query computed using language modeling

CORE COLUMN ENTITY RANKING FEATURES

Feature	Iter. (t)
Term-based matching	
ϕ_1: $LM(q, e_a)$	≥ 0
Deep semantic matching	
ϕ_2: $DRRM\_TKS(q, e_d)$	≥ 0
ϕ_3: $DRRM\_TKS(q, e_p)$	≥ 0
ϕ_4: $DRRM\_TKS(s, e_d)$	≥ 1
ϕ_5: $DRRM\_TKS(s, e_p)$	≥ 1
ϕ_6: $DRRM\_TKS(q \oplus s, e_d \oplus e_p)$	≥ 1
Entity-schema compatibility	
ϕ_7: $ESC(S, e)$	≥ 1

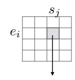
s is the concatenation of all schema labels in S
$\oplus$ is the string concatenation operator



CORE COLUMN ENTITY RANKING FEATURES

Feature	Iter. (t)
Term-based matching	
ϕ_1: $LM(q, e_a)$	≥ 0
Deep semantic matching	
ϕ_2: $DRRM\_TKS(q, e_d)$	≥ 0
ϕ_3: $DRRM\_TKS(q, e_p)$	≥ 0
ϕ_4: $DRRM\_TKS(s, e_d)$	≥ 1
ϕ_5: $DRRM\_TKS(s, e_p)$	≥ 1
ϕ_6: $DRRM\_TKS(q \oplus s, e_d \oplus e_p)$	≥ 1
Entity-schema compatibility	
ϕ_7: $ESC(S, e)$	≥ 1

Compatibility matrix:



$$C_{ij} = \begin{cases} 1, & \text{if } match_{KB}(e_i, s_j) \vee match_{TC}(e_i, s_j) \\ 0, & \text{otherwise .} \end{cases}$$

Entity-schema compatibility score:

$$ESC(S, e_i) = \frac{1}{|S|} \sum_j C_{ij}$$

SCHEMA DETERMINATION

$$score_t(s, q) = \sum_i w_i \phi_i(s, q, E^{t-1})$$

Feature		Iter. (t)
Column population	ϕ_1: $P(s\|q)$	≥ 0
	ϕ_2: $P(s\|q, E)$	≥ 1
Deep semantic matching	ϕ_3: $DRRM\_TKS(s, q)$	≥ 0
Attribute retrieval	ϕ_4: $AR(s, E)$	≥ 1
Entity-schema compatibility	ϕ_5: $ESC(s, E)$	≥ 1

SCHEMA DETERMINATION FEATURES

Feature		Iter. (t)
Column population	ϕ_1: $P(s\|q)$	≥ 0
	ϕ_2: $P(s\|q, E)$	≥ 1
Deep semantic matching	ϕ_3: $DRRM\_TKS(s, q)$	≥ 0
Attribute retrieval	ϕ_4: $AR(s, E)$	≥ 1
Entity-schema compatibility	ϕ_5: $ESC(s, E)$	≥ 1

$$P(s|q) = \sum_{T \in \mathcal{T}} \underbrace{P(s|T)}_{\text{Schema label likelihood}} \underbrace{P(T|q)}_{\text{Table's relevance to the query}}$$

Table's relevance to the query

Schema label likelihood

$$P(s|T) = \begin{cases} 1, & \max_{s' \in T_S} dist(s, s') \geq \gamma \\ 0, & \text{otherwise .} \end{cases}$$

SCHEMA DETERMINATION FEATURES

Feature		Iter. (t)
Column population	ϕ_1: $P(s\|q)$	≥ 0
	ϕ_2: $P(s\|q, E)$	≥ 1
Deep semantic matching	ϕ_3: $DRRM\_TKS(s, q)$	≥ 0
Attribute retrieval	ϕ_4: $AR(s, E)$	≥ 1
Entity-schema compatibility	ϕ_5: $ESC(s, E)$	≥ 1

$$P(s|q, E) = \sum_T \underbrace{P(s|T)}_{\text{Schema label likelihood}} \underbrace{P(T|q, E)}_{\text{Table's relevance to the query}}$$

Schema label likelihood

Table's relevance to the query

$$P(T|q, E) \propto P(T|E)P(T|q)$$

SCHEMA DETERMINATION FEATURES

Feature		Iter. (t)
Column population	ϕ_1: $P(s\|q)$	≥ 0
	ϕ_2: $P(s\|q, E)$	≥ 1
Deep semantic matching	ϕ_3: $DRRM\_TKS(s, q)$	≥ 0
Attribute retrieval	ϕ_4: $AR(s, E)$	≥ 1
Entity-schema compatibility	ϕ_5: $ESC(s, E)$	≥ 1

$$AR(s, E) = \frac{1}{|E|} \sum_{e \in E} \big(match(s, e, T) + drel(d, e) + sh(s, e) + kb(s, e)\big)$$

Similarity between entity e and schema label s with respect to T

Relevance of the document containing the table

#hits returned by a web search engine to the query "[s] of [e]" above threshold

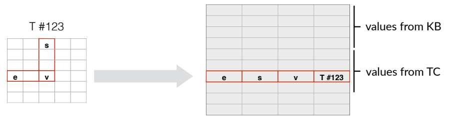
Whether s is a property of e in the KB

Kopliku et al. Towards a Framework for Attribute Retrieval. In: CIKM '11.

VALUE LOOKUP

- A catalog of possible entity attribute-value pairs
  - Entity, schema label, value, provenance quadruples

$\langle e, s, v, p \rangle$



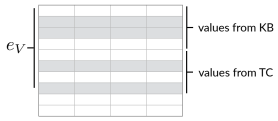
VALUE LOOKUP

- Finding a cell's value is a lookup in that catalog

$$score(v, e, s, q) = \max_{\substack{\langle s', v, p \rangle \in e_V \\ \boxed{match(s, s')}}} \boxed{conf(p, q)}$$

matching confidence
- KB takes priority over TC
- based on the corresponding table's relevance to the query

soft string matching
- "birthday" vs. "date of birth"
- "country" vs. "nationality"



EXPERIMENTAL EVALUATION

EXPERIMENTAL SETUP

- Table corpus
  - WikiTables corpus: 1.6M tables extracted from Wikipedia
- Knowledge base
  - DBpedia (2015-10): 4.6M entities with an English abstract
- Two query sets
- Rank-based metrics
  - NDCG for core column entity ranking and schema determination
  - MAP/MRR for value lookup

QUERY SET 1 (QS-1)

- List queries from the DBpedia-Entity v2 collection[1] (119)
  - *"all cars that are produced in Germany"*
  - *"permanent members of the UN Security Council"*
  - *"Airlines that currently use Boeing 747 planes"*
- Core column entity ranking
  - Highly relevant entities from the collection
- Schema determination
  - Crowdsourcing, 3-point relevance scale, 7k query-label pairs
- Value lookup
  - Crowdsourcing, 25 queries sample, 14k cell values

[1] Hasibi et al. **DBpedia-Entity v2: A Test Collection for Entity Search.** In: *SIGIR '17.*

QUERY SET 2 (QS-2)

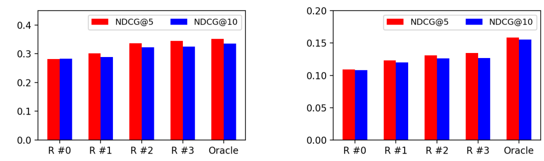
- Entity-relationship queries from the RELink Query Collection[1] (600)
  - Queries are answered by entity tuples (pairs or triplets)
  - That is, each query is answered by a table with 2 or 3 columns (including the core entity column)
  - Queries and relevance judgments are obtained automatically from Wikipedia lists that contain relational tables
  - Human annotators were asked to formulate the corresponding information need as a natural language query
    - *"find peaks above 6000m in the mountains of Peru"*
    - *"Which countries and cities have accredited armenian embassadors?"*
    - *"Which anti-aircraft guns were used in ships during war periods and what country produced them?"*

[1] Saleiro et al. **RELink: A Research Framework and Test Collection for Entity-Relationship Retrieval.** In: *SIGIR '17.*

CORE COLUMN ENTITY RANKING (QUERY-BASED)

	QS-1		QS-2	
	NDCG@5	NDCG@10	NDCG@5	NDCG@10
LM	0.2419	0.2591	0.0708	0.0823
DRRM_TKS (e_d)	0.2015	0.2028	0.0501	0.0540
DRRM_TKS (e_p)	0.1780	0.1808	**0.1089**	**0.1083**
Combined	**0.2821**	**0.2834**	0.0852	0.0920

CORE COLUMN ENTITY RANKING (SCHEMA-ASSISTED)

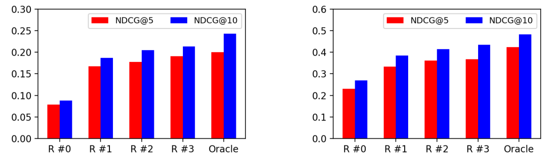


- **R #0:** without schema information (query only)
- **R #1-#3:** with automatic schema determination (top 10)
- **Oracle:** with ground truth schema

SCHEMA DETERMINATION (QUERY-BASED)

	QS-1		QS-2	
	NDCG@5	NDCG@10	NDCG@5	NDCG@10
CP	0.0561	0.0675	0.1770	0.2092
DRRM_TKS	0.0380	0.0427	0.0920	0.1415
Combined	**0.0786**	**0.0878**	**0.2310**	**0.2695**

SCHEMA DETERMINATION (ENTITY-ASSISTED)



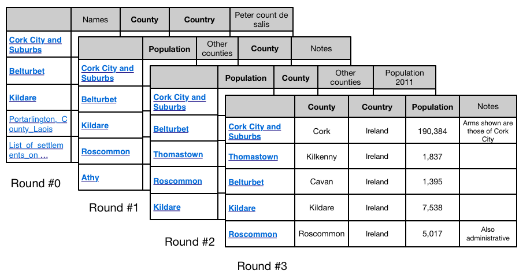
- **R #0:** without entity information (query only)
- **R #1-#3:** with automatic core column entity ranking (top 10)
- **Oracle:** with ground truth entities

RESULTS VALUE LOOKUP

	QS-1		QS-2	
	MAP	MRR	MAP	MRR
Knowledge base	0.7759	0.7990	0.0745	0.0745
Table corpus	0.1614	0.1746	**0.9564**	**0.9564**
Combined	**0.9270**	**0.9427**	**0.9564**	**0.9564**

EXAMPLE

"Towns in the Republic of Ireland in 2006 Census Records"



ANALYSIS

		QS-1			QS-2		
		↑	↓	-	↑	↓	-
QS-1	Round #0 vs. #1	43	38	38	52	7	60
	Round #0 vs. #2	50	30	39	61	5	53
	Round #0 vs. #3	49	26	44	59	2	58
QS-2	Round #0 vs. #1	166	82	346	386	56	158
	Round #0 vs. #2	173	74	347	388	86	126
	Round #0 vs. #3	173	72	349	403	103	94

SUMMARY

- Answering queries with relational tables, summarizing entities and their attributes
  - Retrieving existing tables from a table corpus
  - Generating a table on-the-fly
- Future work
  - Moving from homogeneous Wikipedia tables to other types of tables (scientific tables, Web tables)
  - Value lookup with conflicting values; verifying cell values
  - Result snippets for table search results
  - ...

QUESTIONS?

@krisztianbalog
krisztianbalog.com