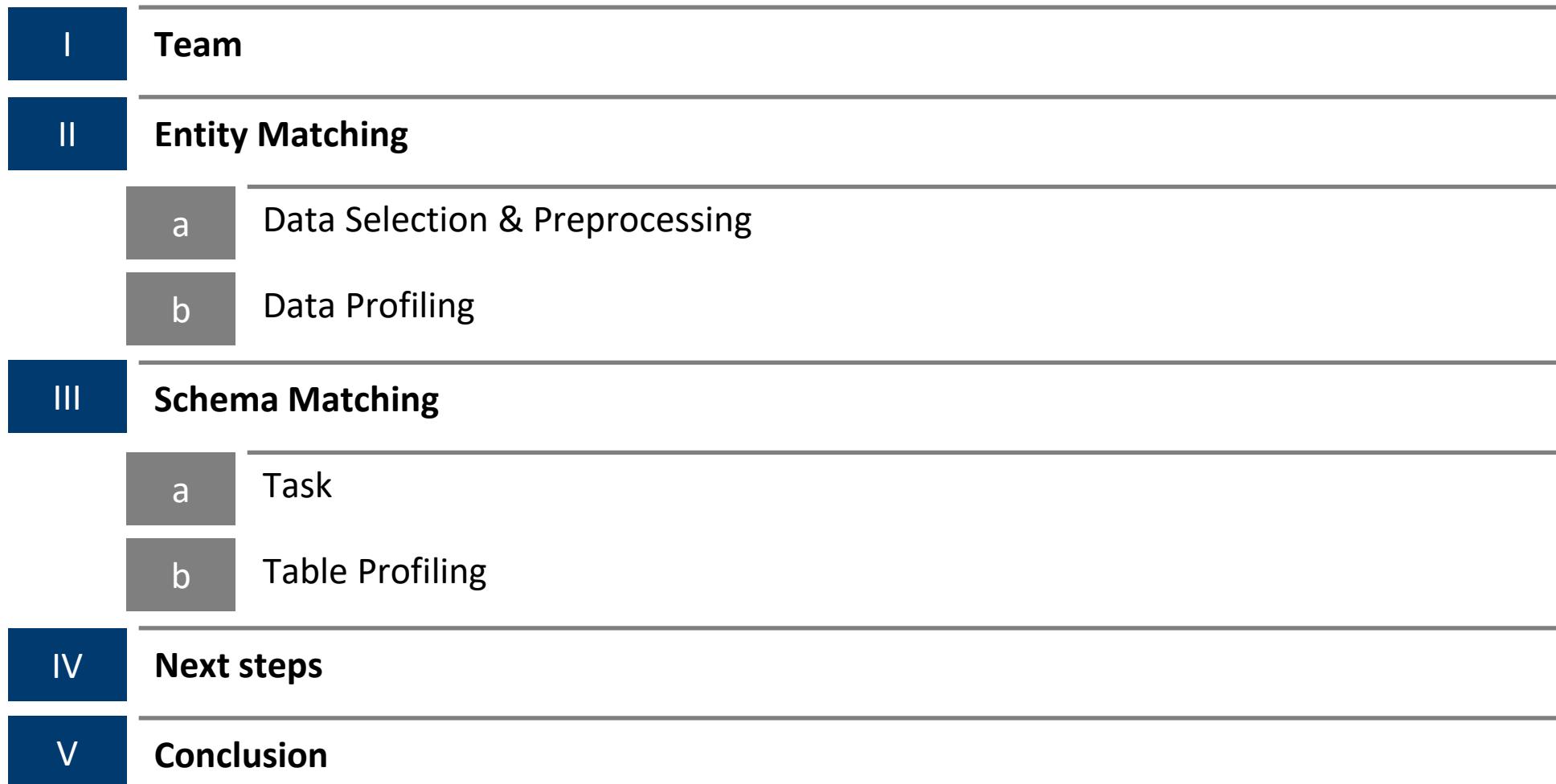




Phase 1a: Data Selection and Data Profiling I

Agenda



Team

Meet the Team

Entity Matching



Jannik

2nd Semester
Data Science



Kim

2nd Semester
Data Science



Cheng

3rd Semester
Data Science



Niklas

2nd Semester
Data Science

Schema Matching



Luisa

2nd Semester
Business Informatics



Jennifer

2nd Semester
Data Science



Marvin

4th Semester
Data Science



Estelle

3rd Semester
Data Science

Entity Matching

Data Selection

PRODUCT

- Biggest entity type
- Matchings via pre-clustered IDs already given in another corpus
- Annotation via these cluster_ids produces enough clusters with an interesting number of tables contained in the clusters

row_id	name	offers	description	url	brand	image	page_url
0	Fetish Fantasy Double Strap on	{"pricecurrency": "USD", "description": "Expert..."} Fantasy Miniature Knight Warrior With Battleax...	None	https://1-night-in-heaven.myshopify.com/product/125247588	Night in Heaven		https://1-night-in-heaven.myshopify.com/product/125247588
1	13037 神說 說了- 善約各 卷精華(卷 六)	{"price": "\$ 11.95", "pricecurrency": "USD"} ("name": "PHOTO BULLES VÉGÉTALES", "price": "6...")	Description: This fighter prepares to enter th...	https://adventures-and-hobbies.myshopify.com/product/125137138	None		https://adventures-and-hobbies.myshopify.com/c...
2	Colorful Luggage Tag Set	{"pricecurrency": "USD", "description": "Set o..."} 作者: 江 守 道 出版社 活出... 本書是作 者在美國 Richmond 的一系 列 講...	None	https://alleluia-bookstore.myshopify.com/product/109943438	Living Word		https://alleluia-bookstore.myshopify.com/product/109943438
3	PHOTO BULLES VÉGÉTALES	{"name": "PHOTO BULLES VÉGÉTALES", "price": "6..."} None	https://amis-fsh.myshopify.com/products/photo-bulles-v%C3%A9g%C3%A9tales	AMIS FSH		https://amis-fsh.myshopify.com/products/photo-bulles-v%C3%A9g%C3%A9tales	
4				https://be-the-boutique.myshopify.com/product/1010923180090	None		https://be-the-boutique.myshopify.com/c...

LOCALBUSINESS

- 3rd biggest entity type given
- Phone number as identifier

row_id	name	address	uri	page_url
0	Decent stone group Inc	{"postalcode": "L3R 2R9", "addresscountry": "C..."} {"addresslocality": "Marlton", "streetaddress": "..."} {"streetaddress": "1661 2nd Ave W, Suite 102", ...} {"addresslocality": "Surrey", "addressregion": "..."} {"streetaddress": "177, rue Guyon", "addresslo..."} {"name": "Decent stone group Inc"} {"name": "Carrero & Sons Concrete & Constru..."} {"name": "Losee Jonathan Ltd"} {"name": "Van Bros Forming Ltd"} {"name": "Centre Amour Et Mariage Inc"} https://homestars.com/companies/2936... https://homestars.com/companies/1938449-carre... https://homestars.com/companies/214554-losee-j... https://homestars.com/companies/228688-van-bro... https://homestars.com/companies/2570149-centre...	    	https://homestars.com/companies/2936... https://homestars.com/companies/1938449-carre... https://homestars.com/companies/214554-losee-j... https://homestars.com/companies/228688-van-bro... https://homestars.com/companies/2570149-centre...
1	Carrero & Sons Concrete & Construction			
2	Losee Jonathan Ltd			
3	Van Bros Forming Ltd			
4	Centre Amour Et Mariage Inc			

Preprocessing with Language Detection

APPROACH

Step 1

Cleaned the tables with TLD-based approach

- Use `*.com, *.net, *.org, *.uk`

Step 2

Used fastText Language Detection

- Every column containing a string or list checked for English language
- for Product: “name”, “description”, “page_url”; excluded “brand” column
- for LocalBusiness: “name”, “page_url”

STATISTICS

Product

from **100** down to
77 files

from **4,481,576** to
1,332,426 entries

28 tables with
cleaning ratio > 80%

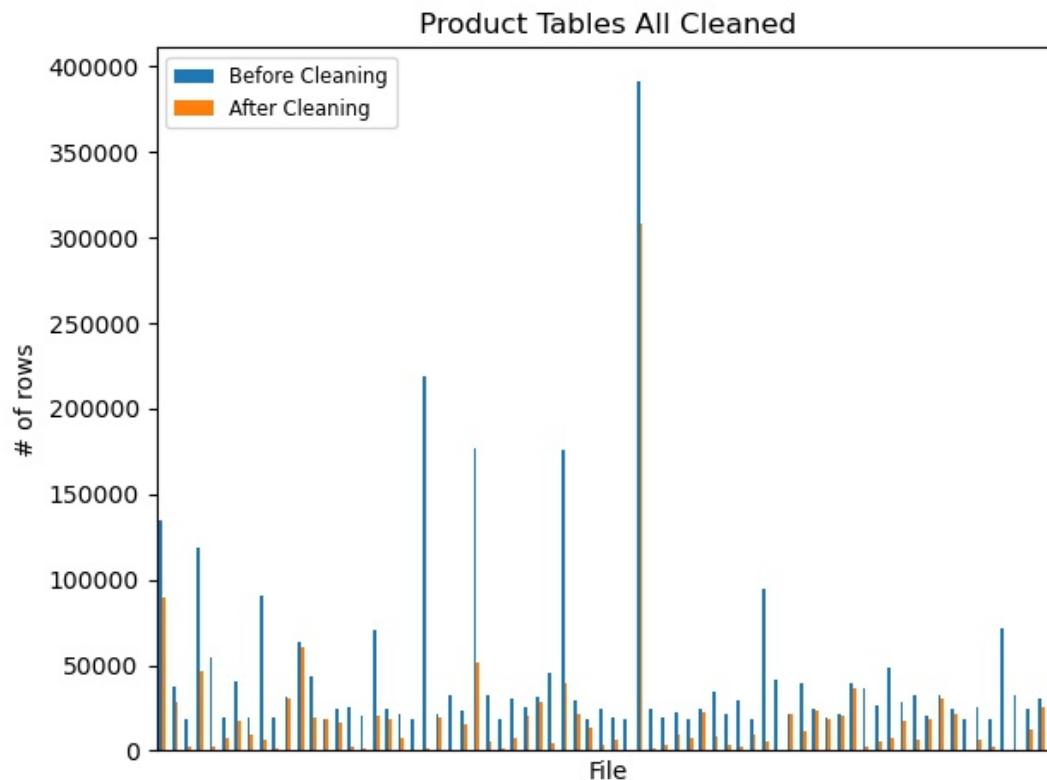
LocalBusiness

from **100** down to
55 files

from **1,114,508** to
215,069 entries

15 tables with
cleaning ratio > 80%

Preprocessing with Language Detection



Product Clustering

APPROACH

Use only cleaned Top 100 tables

- Match via the entries in the cluster_tables
- Problem: Maximal number of matches did not exceed 27, which were all in one table
- Technical difficulties with the uni-cluster and therefore limited computing power
- Difficulties to include Minimum3 tables with valuable information

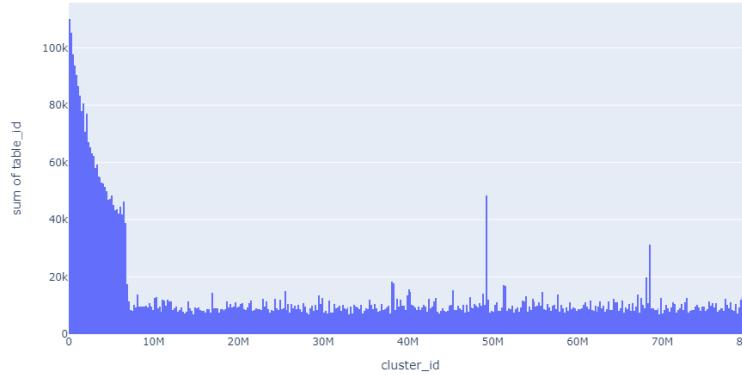
ALTERNATIVE

Clean the cluster tables

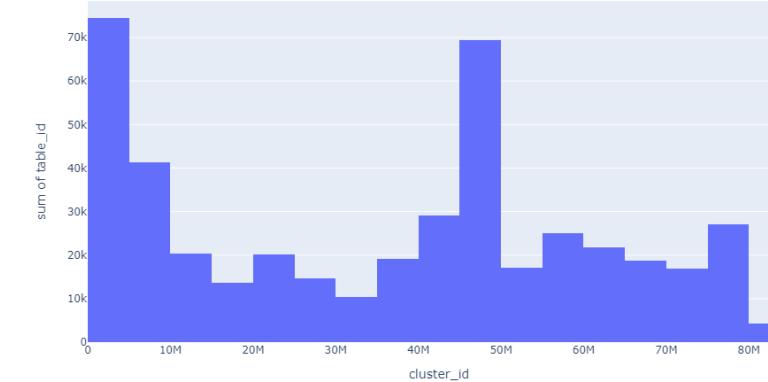
- Use `*.com, *.net, *.org, *.uk`
- Concatenate remaining information and use that to cluster the information via cluster_ids
- Get information about table distribution
- Next step: Bring together cluster information and table information

What is the distribution of tables per product cluster?

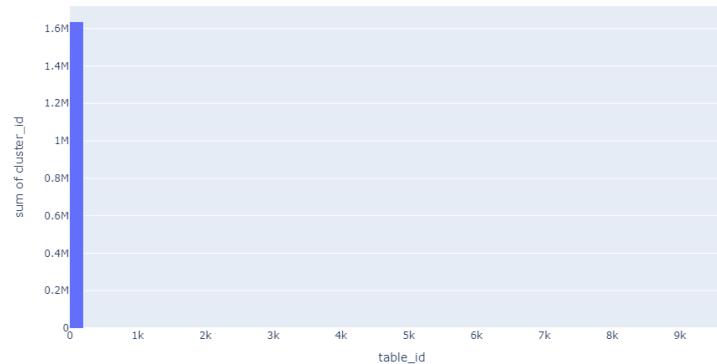
More than one table per cluster



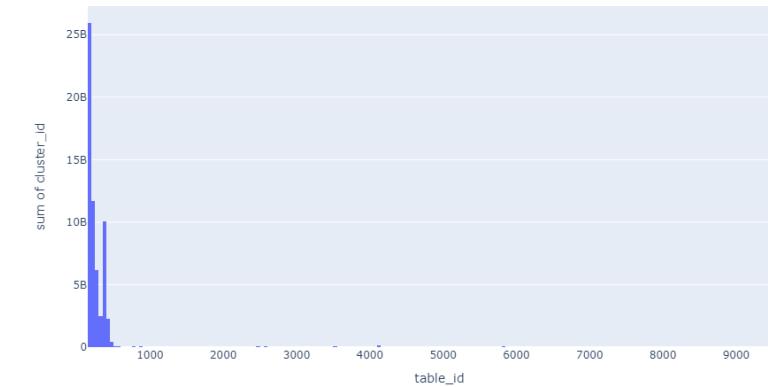
More than 150 tables per cluster



	count	mean	std	min	25%	50%	75%	max
table_id	1633965.0	4.0	17.0	2.0	2.0	2.0	3.0	9749.0

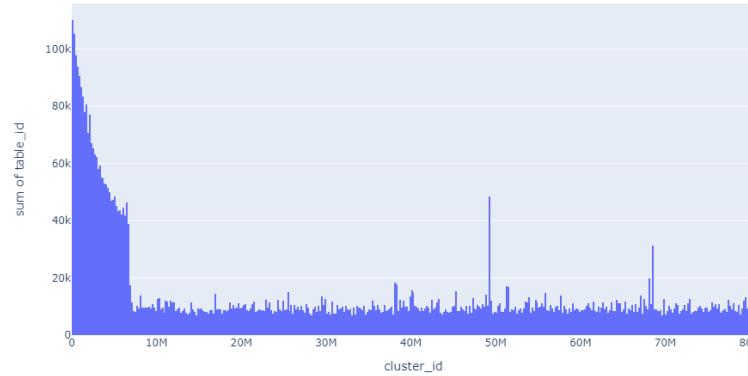


	count	mean	std	min	25%	50%	75%	max
table_id	1631.0	272.0	416.0	151.0	171.0	206.0	285.0	9749.0



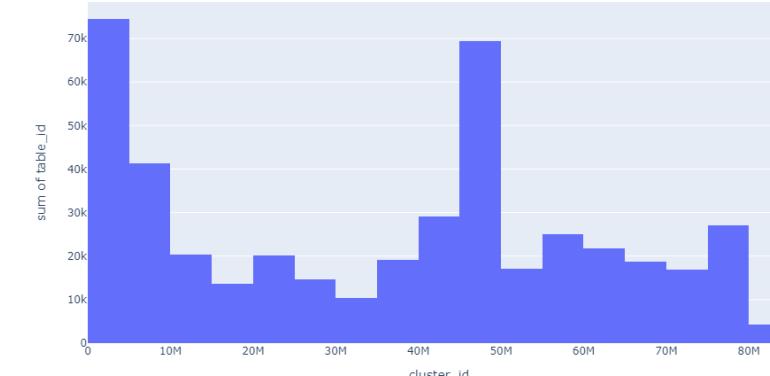
What is the distribution of tables per product cluster?

More than one table per cluster



	count	mean	std	min	25%	50%	75%	max
table_id	1633965.0	4.0	17.0	2.0	2.0	2.0	3.0	9749.0

More than 150 tables per cluster



	count	mean	std	min	25%	50%	75%	max
table_id	1631.0	272.0	416.0	151.0	171.0	206.0	285.0	9749.0

Problem:

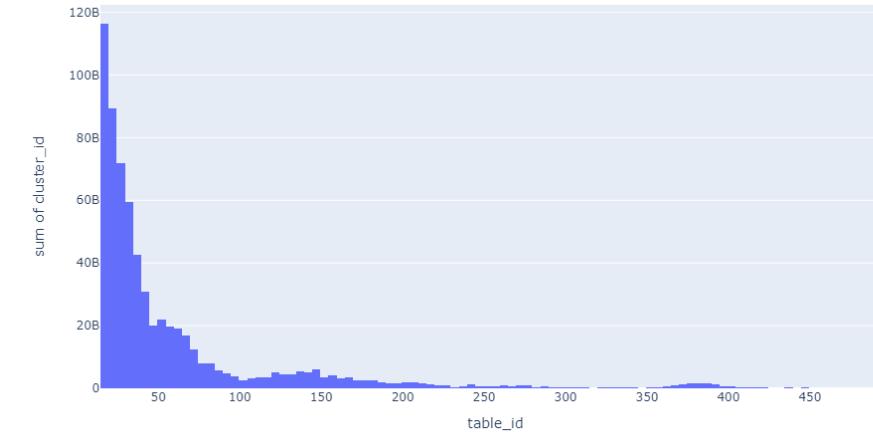
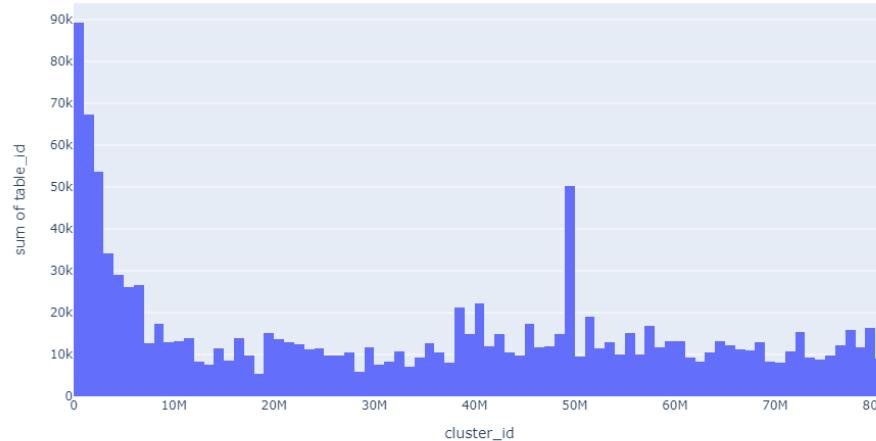
Too many small clusters with specific information especially in Top100

Problem:

Product focus especially flowers and bikes which seem easy to match

What is the distribution of tables per product cluster?

Distribution of clusters with at least 15 tables



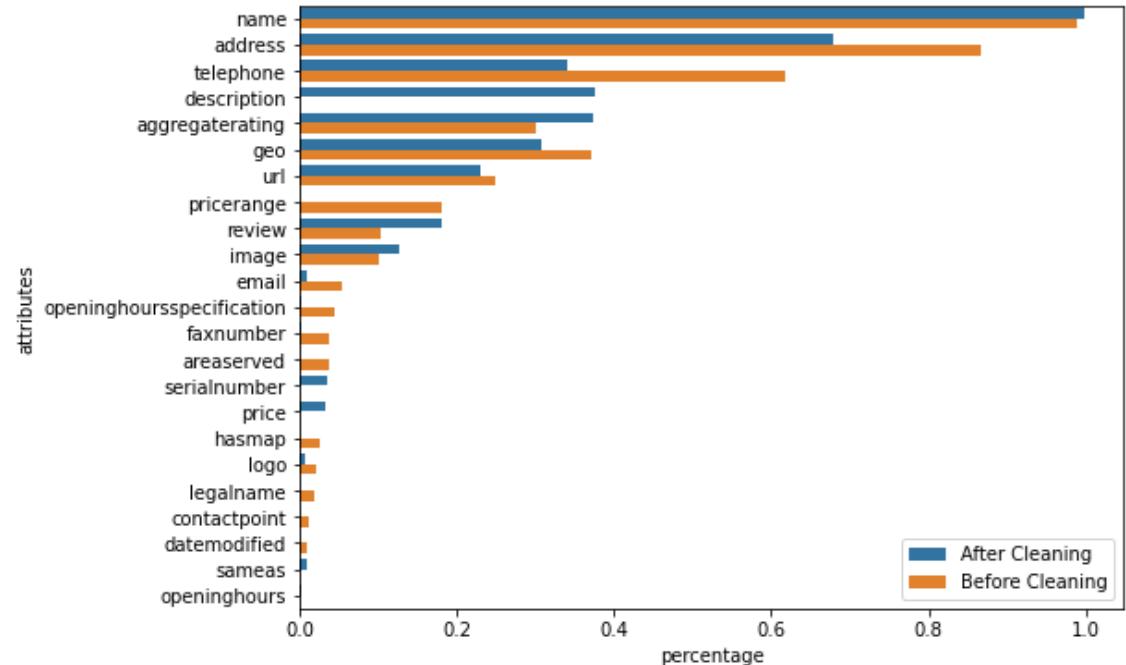
	count	mean	std	min	25%	50%	75%	max
table_id	21165.0	59.0	134.0	16.0	21.0	31.0	60.0	9749.0



A lot of interesting data in the area around 15-30 tables and a focus on the first 500.000 cluster_ids

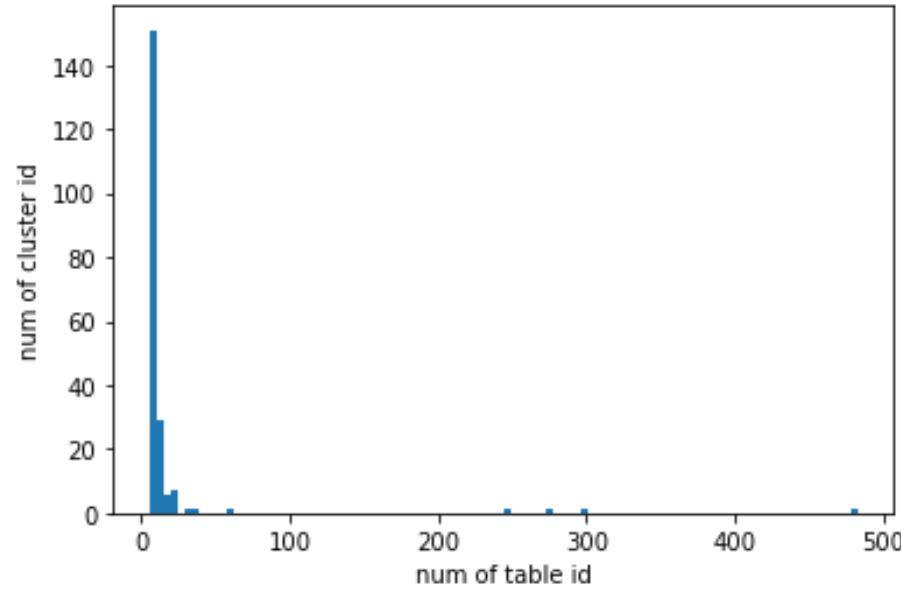
Entity Matching of LocalBusiness

- Attributes Statistics before and after language detection
- Attributes used as identifier
 - Telephone
 - Geo: longitude, latitude
 - Name
- Statistics:
 - Telephone: from 61.71% to 34.03%
 - Geo: from 37.03% to 30.67%
 - Name: from 98.785 to 99.82%

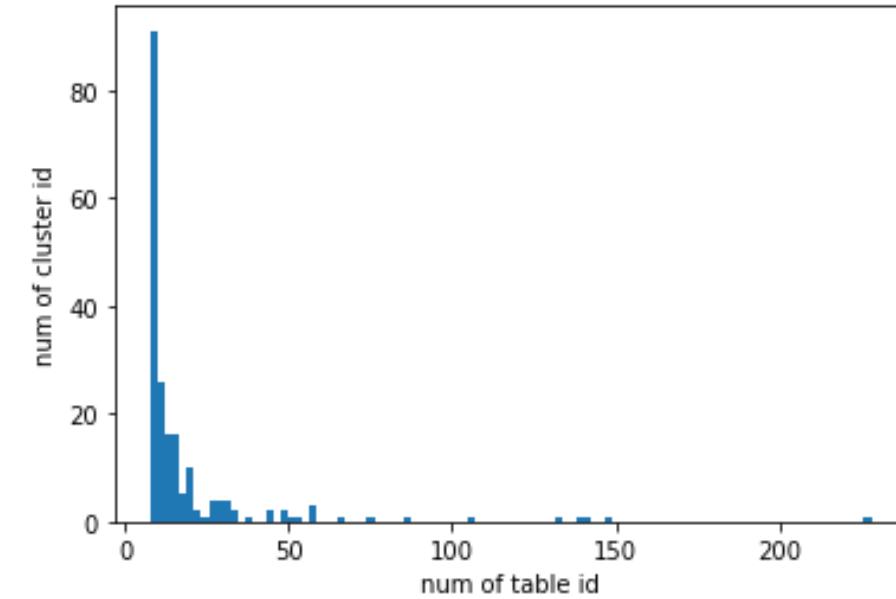


Entity Matching of LocalBusiness

Matching using Telephone (largest 200 clusters)



Matching using Geo (largest 200 clusters)



Entity Matching of LocalBusiness

- Problems of using telephone as identifier
 - Branches may share a same telephone number
 - Some entities have telephone number with 1 or 2 digits
- Problems of using geo as identifier
 - Geological locations refer to business headquarter
 - Different entities share the same building
- Alternative: hybrid matching strategy
 - Calculate token-based similarity score of entity names for further filtering: pairs with less than 0.6 Jaccard similarity score are dropped
 - Results: even worse 😞
 - Potential solution: trying to run algorithm using richer dataset by including Minimum3 tables

Hard Matches and Hard Non-Matches

- Merge entity tables ***Product*** and ***LocalBusiness***
- Use column **[‘name’]** to identify entities
- Example ***Product*** :

index		name	origin
0	0	S14/S12 series Aluminum Rear Wing Washer (GM)(2PC)	Product_3dcartstores.com_September2020.json.gz
1	1	Shim 10x12x0.2mm	Product_3dcartstores.com_September2020.json.gz
2	2	Vegetable Ivory Whistle. Set of 16	Product_3dcartstores.com_September2020.json.gz
3	3	Long Sleeve T-Shirt	Product_3dcartstores.com_September2020.json.gz
4	4	My Christmas Tree Table Runner Nutmeg-Barn Red	Product_3dcartstores.com_September2020.json.gz
...
618067	11388	I'll Be Your Huckleberry Serape Leather Drop Earrings	Product_shopgate.com_September2020.json.gz
618068	11389	Turquoise Squash Blossom Earrings	Product_shopgate.com_September2020.json.gz
618069	11390	Rocking that Buffalo Stone and Navajo Bead Necklace	Product_shopgate.com_September2020.json.gz
618070	11391	Anlasser Relais Starterrelais Bombardier Outlander/MAX/XT 400 2x2/4x4	Product_shopgate.com_September2020.json.gz
618071	11392	Comp LS Crew	Product_shopgate.com_September2020.json.gz
618072 rows × 3 columns			

Hard matches and hard non-matches

FURTHER PREPROCESSING

Using REGEX

- Strategy: Remove symbols but keep numbers
- Example: 'Iphone6' vs. 'Iphone7'
- Formula: [^\w\S]

VECTORIZATION

Using Doc2Vec

- implemented in gensim library
- Goal: get dense vector representation for each entity
- Training done on both large entity tables each

Hard Matches and Hard Non-Matches: Product Examples

HARD NON-MATCHES

Entity	Most Similar Entity	Cosine Metric
'Lifeproof Case Iphone 11'	'iPhone 11 Pro Max case'	0.9776
'Lego Star Wars The Complete Saga DS'	'Lego Star Wars: The Complete Saga - Wii Video Game'	0.9367
'10 2010 Audi A5 Quattro Fuel Injector 2.0L 4 Cyl Bosch High Pressure'	'18 2018 Audi A5 Quattro Fuel Injector 2.0L 4 Cyl Standard Motor Products'	0.9771

HARD MATCHES

Entity	Matching Entity	Cosine Metric
'iPhone 11 Pro Max case'	'For iphone 11 pro x xr xs max cell phone case cover with camera lens protection'	0.8514 (top 50)
'08MP-08FPS 90° Elbow Long Forged'	'08MP-08FPS 90° Elbow Forged'	0.9062 (15th place)
'Jasmine Dragon Pearls Green Tea'	'Jasmine Dragon Pearl Jasmine Green Tea'	0.9501 (top 5)

Hard Matches and Hard Non-Matches: LocalBusiness Examples

HARD NON-MATCHES

Entity	Most Similar Entity	Cosine Metric
'Robert A. Martin Law Office '	'Robert A. Gardner Law Office'	0.9849
'Hidden Creek Golf Club'	'Creek Golf Club'	0.9848
'Desert Dental Apple Valley'	'Apple Valley Dental Associates'	0.9786

HARD MATCHES

Entity	Matching Entity	Cosine Metric
'Pfizer Inc.'	'Pfizer Incorporation'	0.9898 (2nd best)
'Cohen & Cohen PA Personal Injury Attorneys'	Cohen & Cohen Personal Injury'	0.9477 (top 3)
'Bankruptcy Law Office Of Richard A Check S C'	'Bankruptcy Law Offices Of Richard A Check S C'	0.9717 (top 3)

Schema Matching

Task at a glance (Phase 1a)

DATA UNDERSTANDING

Become familiar with the structure of data

- Tables
- Statistics files
- Columns
- Technical characteristics

TABLE SELECTION

Find 20 entity types for which some schema looks similar

- Analyzed statistics & occurrence of columns for each class
- Compared and identified columns & properties across classes
- Identified 30 most frequent schema matches as well as ambiguous types

DATA PREPARATION

Data processing and profiling

- Statistical gathering of the chosen tables
- Language detection to remove non-English tables

Data Used:

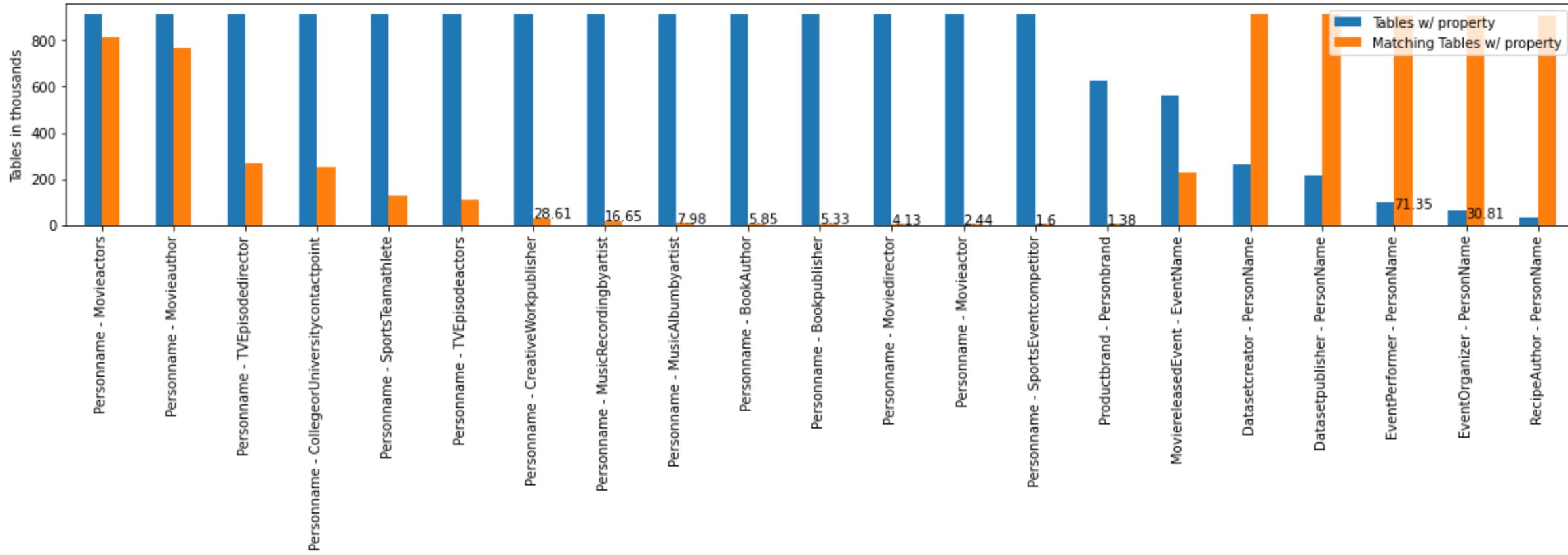
- Looked at Top 30 Schema.Org classes
- Small variance of statistics from Top100/Min3/Rest
- Also matched columns with big differences in table count
- Matched according to no. of tables, not rows

Results (all languages):

- 15 matches where both columns > 10.000 Tables
- 46 matches where one column < 10.000 Tables
- 14 matches where one column < 100 Tables
- Will focus on the matches with the most tables first

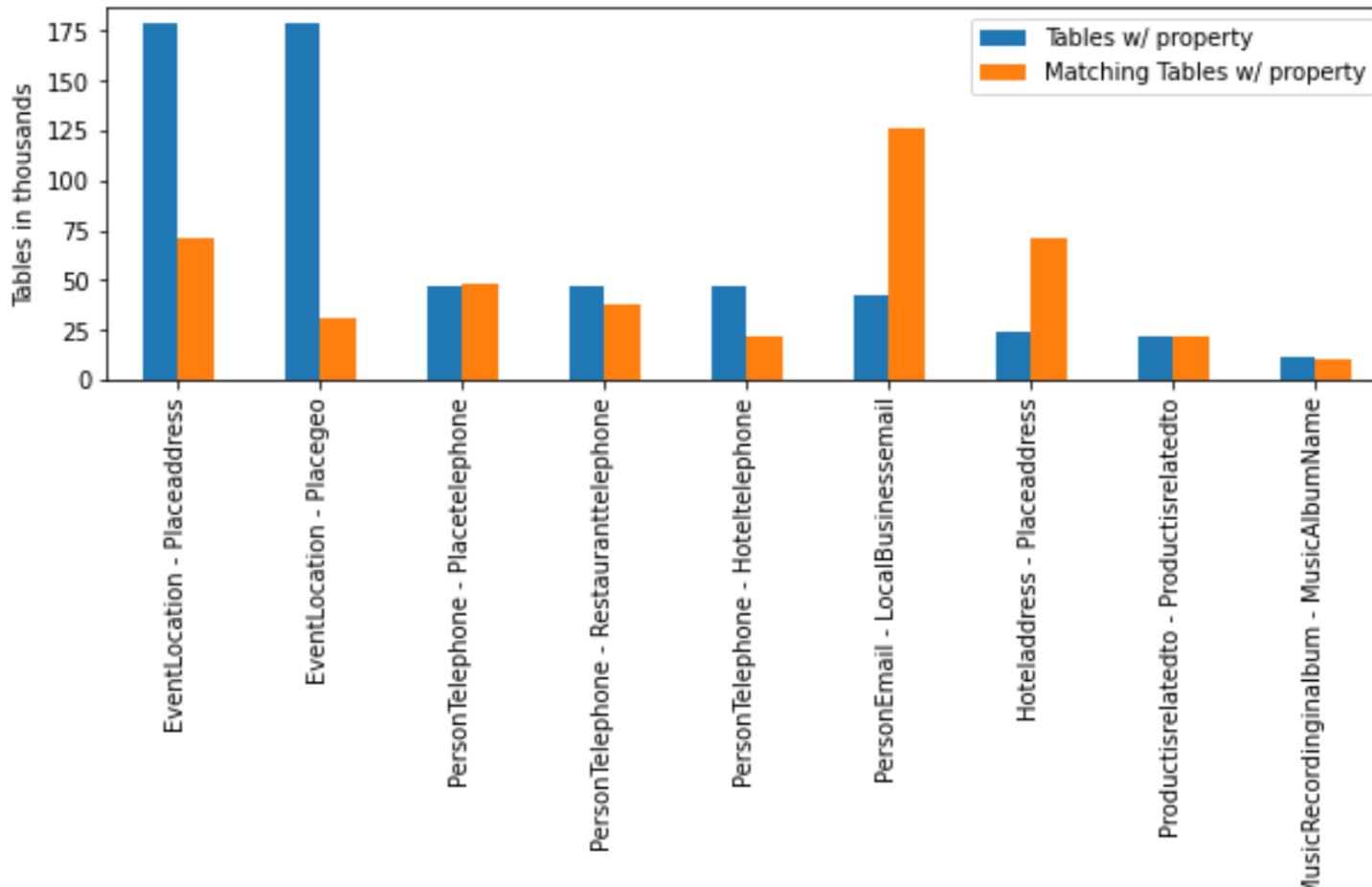
Entity & Column Type Distribution – Pre Language Elimination

Schema Matches with one high-count column

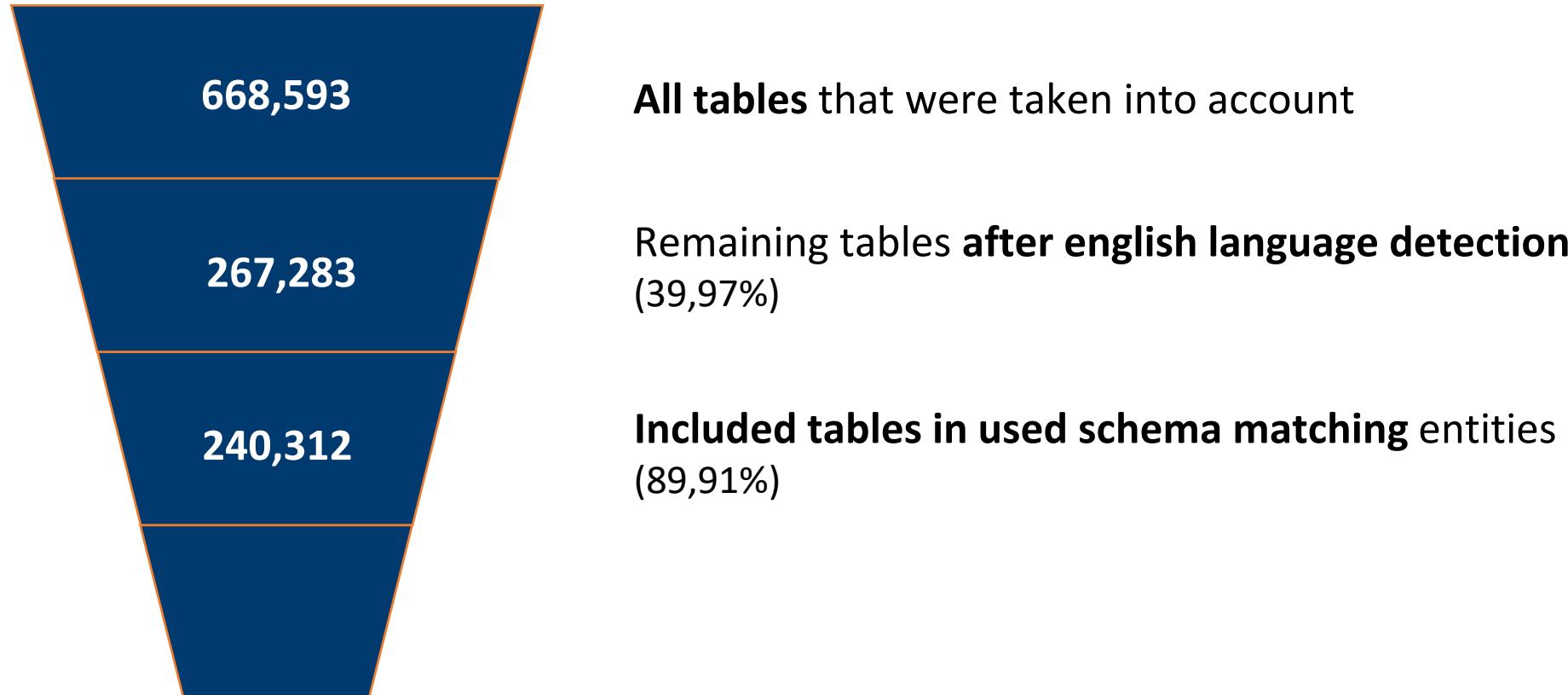


Entity & Column Type Distribution – Pre Language Elimination

Schema Matches medium-count columns

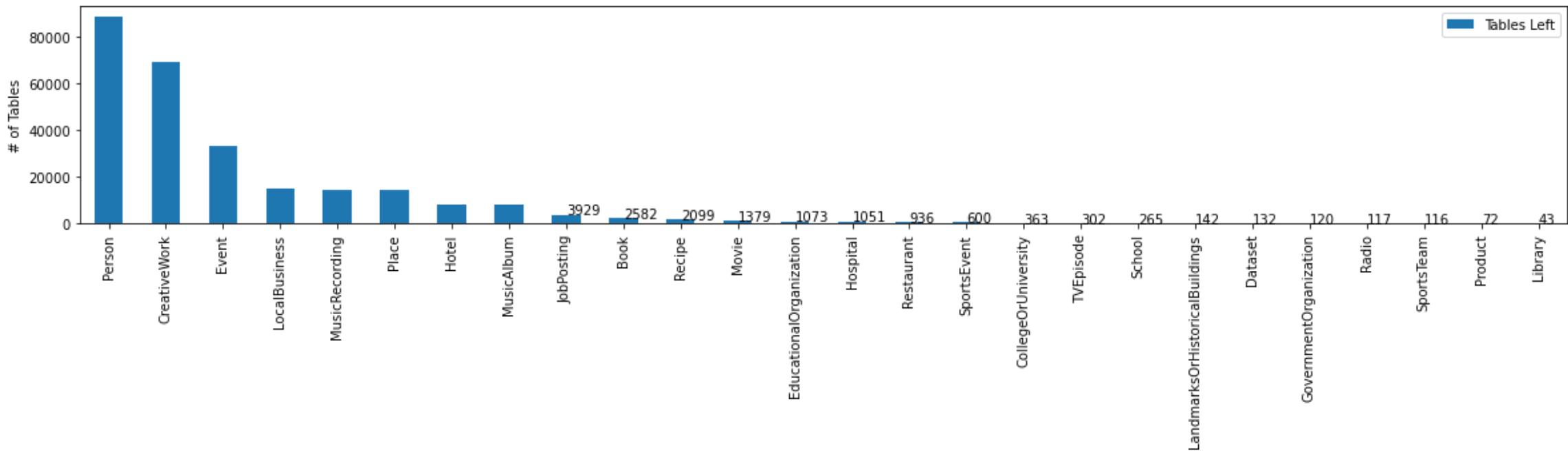


Evolution of Remaining Table Count



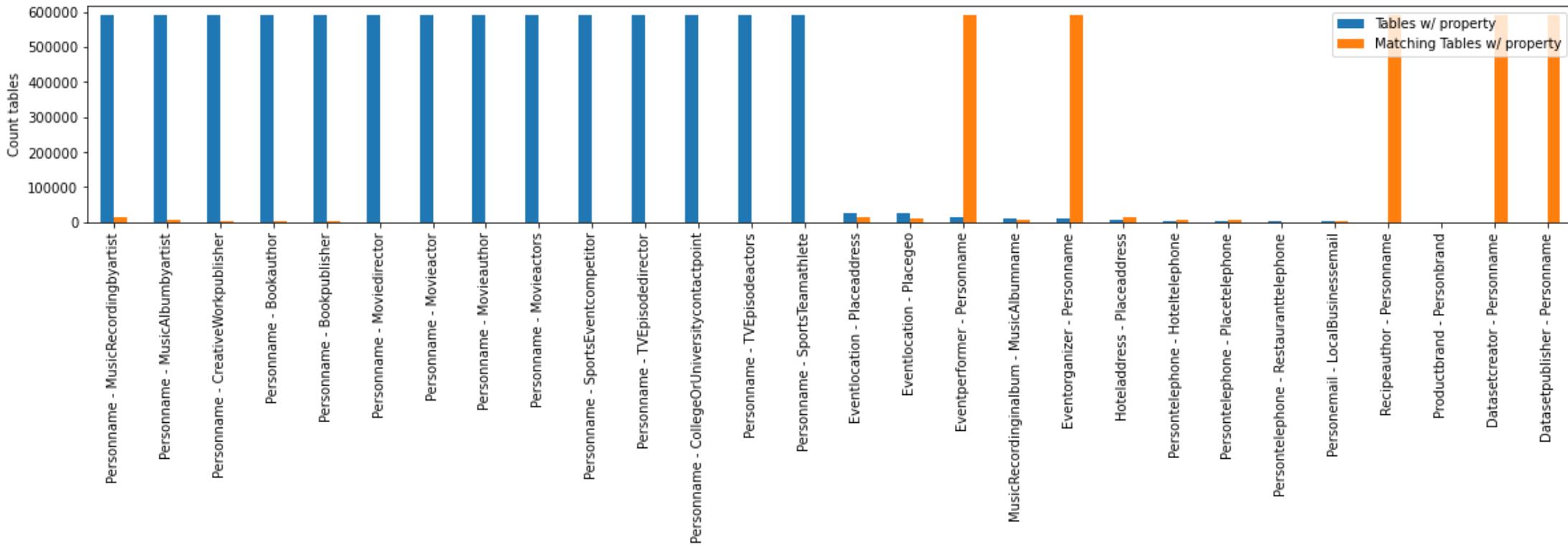
Distribution of Tables across Types

Distribution of remaining tables after english language detection



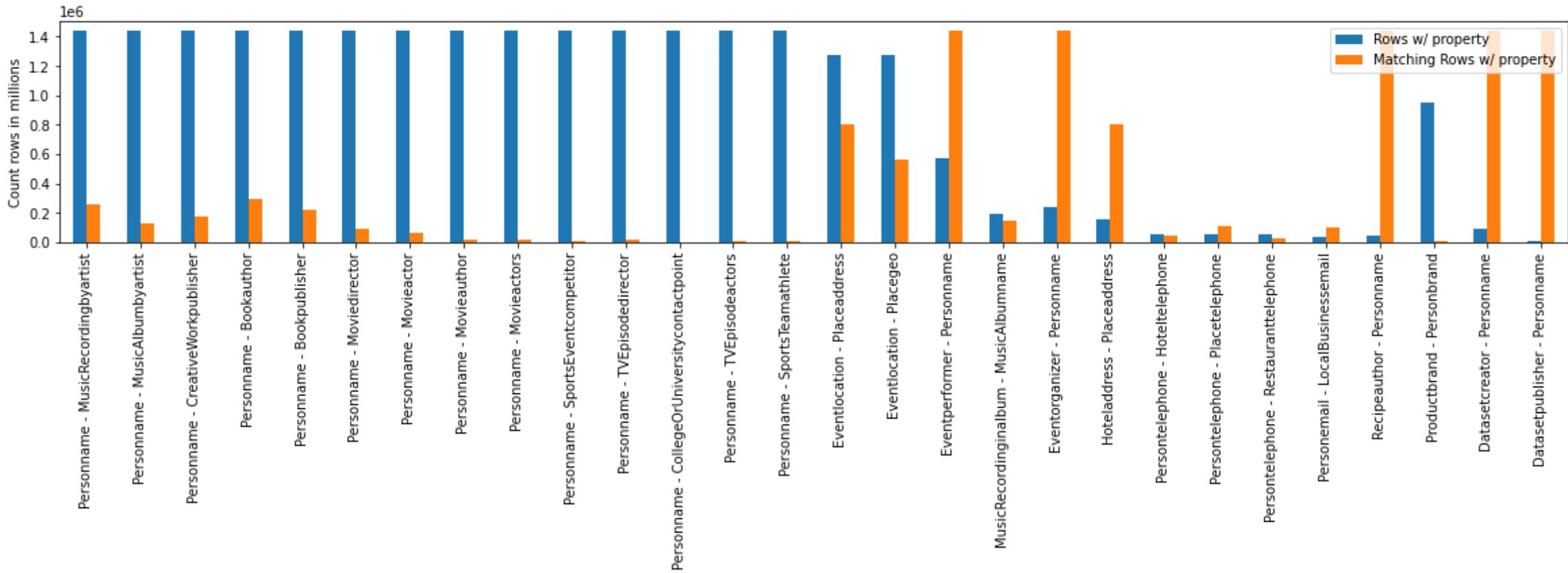
Entity & Column Type Distribution – Post Language Elimination

Top 30 Schema Matches (English only) - Count of tables per match



Entity & Column Type Distribution – Post Language Elimination

Top 30 Schema Matches (English only) - Count of rows per match



Example of 6 Ambiguous Column Types

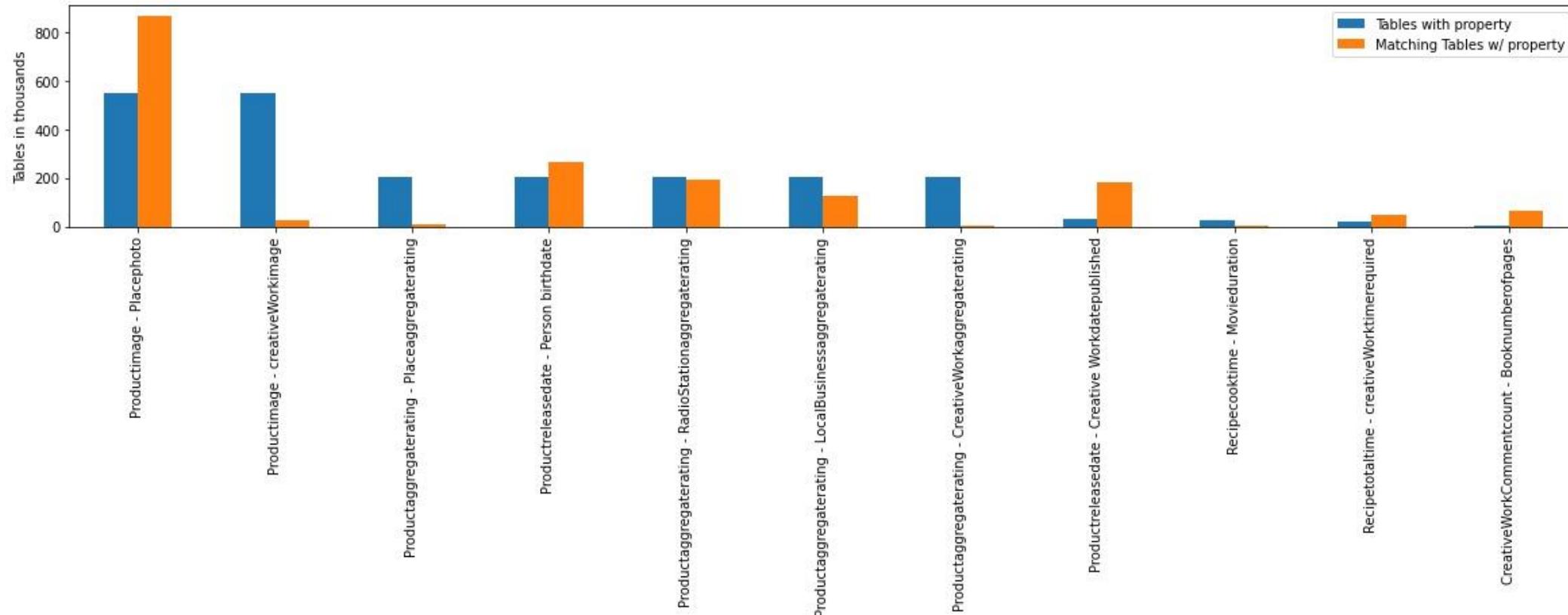
– Pre Language Elimination

Product: Image	Place: Photo
Product: Released Date	Person: Birthdate
Recipe: Cooktime	Movie: Duration
Product: Aggregate Rating	Local Business: Aggregate Rating
Recipe: Total Time	Creative Work: Time Required
Creative Work: Comment Count	Book: Number of Pages

Ambiguous Column Types across Entity Types

– Pre Language Elimination

Ambiguous Matches – Count of Tables



Next steps

Phase 1b – Overview and Questions



Table Selection

Training Set Building

Downsampling

Conclusion

Strengths:

- High motivation throughout whole team and time period
- Different backgrounds – smart distribution of work
- Frequent exchange with entity team
- Good communication within the team
- Frequent meetings

Weaknesses:

- No expert on topic yet – especially dipping into the data was exhausting
- Technical issues slowed process down

How to get better next time:

- Will become easier when we are more within the topic
- Eventually higher frequency of communication with Ralph to remove doubts

Attachment I

Entity Matching Overview (Schema Matching)

Class/Type	# Tables	Properties	Percentage	Tables w/ property	Match Class/Type	Match Property	Match Percentage	Tables	Matching Tables w/ proper	Average of tw
Event	229.980	Performer	42,63	98.040	Person	Name	98,27	921.777	905.830	501.935
Event	229.980	Organizer	26,3	60.485	Person	Name	98,27	921.777	905.830	483.157
Person	921.777	name	99,05	913.020	CreativeWork	publisher	11,35	252.106	28.614	470.817
Recipe	39.246	Author	76,46	30.007	Person	Name	98,27	921.777	905.830	467.919
Person	921.777	name	99,05	913.020	MusicRecording	byArtist	91,46	18.207	16.652	464.836
Event	229.980	Location	77,59	178.441	Place	address	93,05	76.679	71.350	124.896
Event	229.980	Location	77,59	178.441	Place	geo	40,18	76.679	30.810	104.626
Person	921.777	Email	4,54	41.849	LocalBusiness	email	27,04	465.816	125.957	83.903
Hotel	25.528	address	95,26	24.318	Place	address	93,05	76.679	71.350	47.834
Person	921.777	Telephone	5,07	46.734	Place	telephone	62,39	76.679	47.840	47.287
Person	921.777	Telephone	5,07	46.734	Restaurant	telephone	84,98	44.486	37.804	42.269
Person	921.777	Telephone	5,07	46.734	Hotel	telephone	83,92	25.528	21.423	34.079
Person	921.777	Telephone	5,07	46.734	Hotel	telephone	83,92	25.528	21.423	34.079
Product	2.028.974	isrelatedto	1,05	21.304	Product	isrelatedto	1,05	2.028.974	21.304	21.304
MusicRecording	18.207	inAlbum	63,94	11.642	MusicAlbum	Name	99,65	10.200	10.164	10.903
Person	921.777	name	99,05	913.020	MusicAlbum	byArtist	78,26	10.200	7.983	460.501
Person	921.777	name	99,05	913.020	Book	Author	51,04	11.462	5.850	459.435
Person	921.777	name	99,05	913.020	Book	publisher	46,54	11.462	5.334	459.177
Person	921.777	name	99,05	913.020	Movie	director	58,64	7.039	4.128	458.574
Person	921.777	name	99,05	913.020	Movie	actor	34,69	7.039	2.442	457.731
Person	921.777	name	99,05	913.020	SportsEvent	competitor	41,61	3.855	1.604	457.312
Person	921.777	name	99,05	913.020	Movie	actors	11,56	7.039	814	456.917
Person	921.777	name	99,05	913.020	Movie	author	10,87	7.039	765	456.893
Person	921.777	name	99,05	913.020	TVEpisode	director	27,74	959	266	456.643
Dataset	880	creator	29,89	263	Person	Name	99,05	921.777	913.020	456.642
Person	921.777	name	99,05	913.020	CollegeorUniversity	contactpoint	25,08	989	248	456.634
Dataset	880	publisher	24,43	215	Person	Name	99,05	921.777	913.020	456.618
Person	921.777	name	99,05	913.020	SportsTeam	athlete	25,56	489	125	456.573
Person	921.777	name	99,05	913.020	TVEpisode	actors	11,37	959	109	456.565
Product	2.028.974	brand	30,8	624.924	Person	brand	0,15	921.777	1.383	313.153

Attachment II

Language Detection Process – 2 step non-english elimination (Schema Matching)

	b:top_100	a:top_100	b:min3	a:min_3	b:rest	a:rest	Vorher/Nachher	Gesamt
Radio								
Step 1	100	57	146	100	180	108	426	
Step 2		20		48		49	117	0,27464789
LandmarksOrHistoricalBuildings								
Step 1	100	41	246	114	109	52	455	
Step 2		37		87		18	142	0,31208791
Library								
Step 1	100	47	52	23	310	130	462	
Step 2		16		11		16	43	0,09307359
SportsTeam								
Step 1	100	38	146	92	243	102	489	
Step 2		22		51		43	116	0,23721881
GovernmentOrganization								
Step 1	100	36	15	15	454	176	569	
Step 2		23		15		82	120	0,21089631
Dataset								
Step 1	100	40	532	201	248	126	880	
Step 2		30		72		30	132	0,15
TVEpisode								
Step 1	100	58	742	358	117	57	959	
Step 2		52		221		29	302	0,31491137
CollegeOrUn								
Step 1	100	60	174	87	715	321	989	
Step 2		54		46		263	363	0,36703741

School								
Step 1	100	69	168	122	836	504	1104	
Step 2		61		74		130	265	0,24003623
Hospital								
Step 1	100	62	358	269	1112	877	1570	
Step 2		40		162		849	1051	0,66942675
EducationalOrganization								
Step 1	100	58	485	294	3190	1906	3775	
Step 2		44		185		844	1073	0,28423841
SportsEvent								
Step 1	100	61	2576	1394	1179	692	3855	
Step 2		48		474		78	600	0,15564202
Movie								
Step 1	100	47	5688	2690	1251	746	7039	
Step 2		42		1217		120	1379	0,19590851
MusicAlbum								
Step 1	100	49	4088	3741	6012	5673	10200	
Step 2		44		3462		4785	8291	0,81284314
MusicRecording								
Step 1	100	74	13694	12810	4413	4206	18207	
Step 2		70		12198		2453	14721	0,80853518
Hotel								
Step 1	100	55	13014	8876	12414	7696	25528	
Step 2		46		3189		5062	8297	0,32501567

Attachment II

JobPosting						
Step 1	100	74	23497	13991		23597
Step 2		42		3887		3929 0,16650422
Recipe						
Step 1	100	65	30480	21550		30580
Step 2		27		2072		2099 0,06863963
Event						
Step 1	100	72	143898	91788		143998
Step 2		58		33297		33355 0,23163516
Book						
Step 1	100	63	7662	4256		7762
Step 2		52		2530		2582 0,33264623
CreativeWork						
Step 1	100	77	173855	121371		173955
Step 2		69		69245		69314 0,39845937
LocalBusiness						
Step 1	100	55	50506	31602		50606
Step 2		44		14943		14987 0,29615065
Person						
Step 1	100	67	127363	127363		127463
Step 2		65		88544		88609 0,69517429
Product						
Step 1	100	77	zu groß			100
Step 2		72				72 0,72

Place							
Step 1	100	78	27355	18440		27455	
Step 2		73		14315		14388 0,52405755	
Restaurant							
Step 1	100	65	6470	2887		6570	
Step 2		49		887		936 0,14246575	
						668593	
						267283 0,39976937	

Attachment III

Final tables and rows for schema matching (Schema Matching)

		top100	min3	rest		
Movie	director	36	710	63	809	Tables
		33915	54401	69	88385	Rows
	actor	24	366	32	422	Tables
		27075	34620	35	61730	Rows
	author	7	241	8	256	Tables
		731	14252	8	14991	Rows
	actors	6	214	11	231	Tables
		7538	7265	11	14814	Rows
MusicAlbum	byartist	36	2486	3985	6507	Tables
		102213	21978	5007	129198	Rows
	name	42	3455	4780	8277	Tables
		110892	32834	6005	149731	Rows
Hotel	telephone	16	2580	4297	6893	Tables
		22620	20739	5084	48443	Rows
	address	43	3051	4859	7953	Tables
		92870	55152	5770	153792	Rows
MusicRecord	inalbum	30	8866	1100	9996	Tables
		81411	109264	1354	192029	Rows
	byartist	45	11440	2199	13684	Tables
		124755	127307	2639	254701	Rows
Person	name	65	86521		86586	Tables
		293102	1143811		1436913	Rows
	telephone	6	3306		3312	Tables
		22403	34655		57058	Rows
	email	4	3029		3033	Tables
		4002	25001	20102	00102	Tables

Recipe	author	21	1099		1120	Tables
		6120	35405		41525	Rows
Event	location	92	27065		27157	Tables
		1004751	273636		1278387	Rows
Book	author	30	15898		15928	Tables
		430624	142132		572756	Rows
CreativeWork	publisher	20	9103		9123	Tables
		157629	85186		242815	Rows
LocalBusiness	email	41	1387		1428	Tables
		125884	170484		296368	Rows
Product	isrelatedto	33	1111		1144	Tables
		84696	136016		220712	Rows
Brand	brand	6	4198		4204	Tables
		24837	151293		176130	Rows
Product	isrelatedto	-	-			
		-	-			
Product	brand	48			48	Tables
		949568			949568	Rows

Attachment III

Place	address	66	13551		13617	Tables
		177231	627276		804507	Rows
geo		40	8911		8951	Tables
		146761	416186		562947	Rows
telephone		11	6402		6413	Tables
		27511	78758		106269	Rows
Restaurant	telephone	38	602		640	Tables
		18007	12377		30384	Rows
Dataset	creator	12	25	5	42	Tables
		87232	1019	6	88257	Rows
publisher		7	13	9	29	Tables
		6662	220	9	6891	Rows
TVEpisode	director	10	78	9	97	Tables
		10704	6227	11	16942	Rows
actors		5	59	3	67	Tables
		5484	5532	5	11021	Rows
SportsTeam	athlete	8	13	7	28	Tables
		3747	252	8	4007	Rows
CollegeorUni	contactpoint	1	4	88	93	Tables
		12	8	98	118	Rows
SportsEvent	competitor	9	98	6	113	Tables
		3552	696	7	4255	Rows

Attachment VI

English only schema matching entities (Schema Matching)

Class/Type	Properties	Tables w/ pr	Rows w/ pr	Match Class	Match Prop	Matching Ta	Matching Rows w/ property
Event	performer	15928	572756	Person	name	588684	1436913
Event	organizer	9123	242815	Person	name	588684	1436913
Person	name	588684	1436913	CreativeWork	publisher	4204	176130
Recipe	author	1120	41525	Person	name	588684	1436913
Person	name	588684	1436913	MusicRecord	byArtist	13684	254701
Event	location	27157	1278387	Place	address	13617	804507
Event	location	27157	1278387	Place	geo	8951	562947
Person	email	3033	39183	LocalBusiness	email	2028	98532
Hotel	address	7953	153792	Place	address	13617	804507
Person	telephone	3312	57058	Place	telephone	6413	106269
Person	telephone	3312	57058	Restaurant	telephone	640	30384
Person	telephone	3312	57058	Hotel	telephone	6893	48443
MusicRecord	inAlbum	9996	192029	MusicAlbum	name	8277	149731
Person	name	588684	1436913	MusicAlbum	byArtist	6507	129198
Person	name	588684	1436913	Book	author	1428	296368
Person	name	588684	1436913	Book	publisher	1144	220712
Person	name	588684	1436913	Movie	director	809	88385
Person	name	588684	1436913	Movie	actor	422	61730
Person	name	588684	1436913	SportsEvent	competitor	113	4255
Person	name	588684	1436913	Movie	actors	231	14814
Person	name	588684	1436913	Movie	author	256	14991
Person	name	588684	1436913	TVEpisode	director	97	16942
Dataset	creator	42	88257	Person	name	588684	1436913
Person	name	588684	1436913	CollegeOrUniversity	contactPoint	93	118
Dataset	publisher	29	6891	Person	name	588684	1436913
Person	name	588684	1436913	SportsTeam	athlete	28	4007
Person	name	588684	1436913	TVEpisode	actors	67	11021
Product	brand	48	949568	Person	brand	83	3432