

A black and white photograph showing a man in a dark suit operating a large, complex mechanical device. The device has a long, narrow output tray extending from it. The man is positioned in the center-right of the frame, looking down at the device. To the left, there is a wall of numerous small, square drawers or compartments. The overall scene suggests a technical or industrial environment from the mid-20th century.

Extending Knowledge Bases from Human-readable Tables

Benno Kruit

Human-readable Tables

Rank	Name	Points	Team	Chart positions
1	Domenic	88,110	dcode	US US R&B US Dance
2	Sally	72,400	Students	US US R&B
3	Nick	52,300	dcode	US Dance

Web Tables

S.No	Billing No	PCP Name	Amount Received	Bill Amount	Signature
1	123456789012	Dr. John	1500	1600	
2	123456789013	Dr. Smith	1200	1300	
3	123456789014	Dr. Brown	1800	1900	
4	123456789015	Dr. White	1400	1500	
5	123456789016	Dr. Green	1600	1700	
6	123456789017	Dr. Black	1300	1400	
7	123456789018	Dr. Red	1700	1800	
8	123456789019	Dr. Blue	1500	1600	
9	123456789020	Dr. Purple	1400	1500	

Spreadsheets

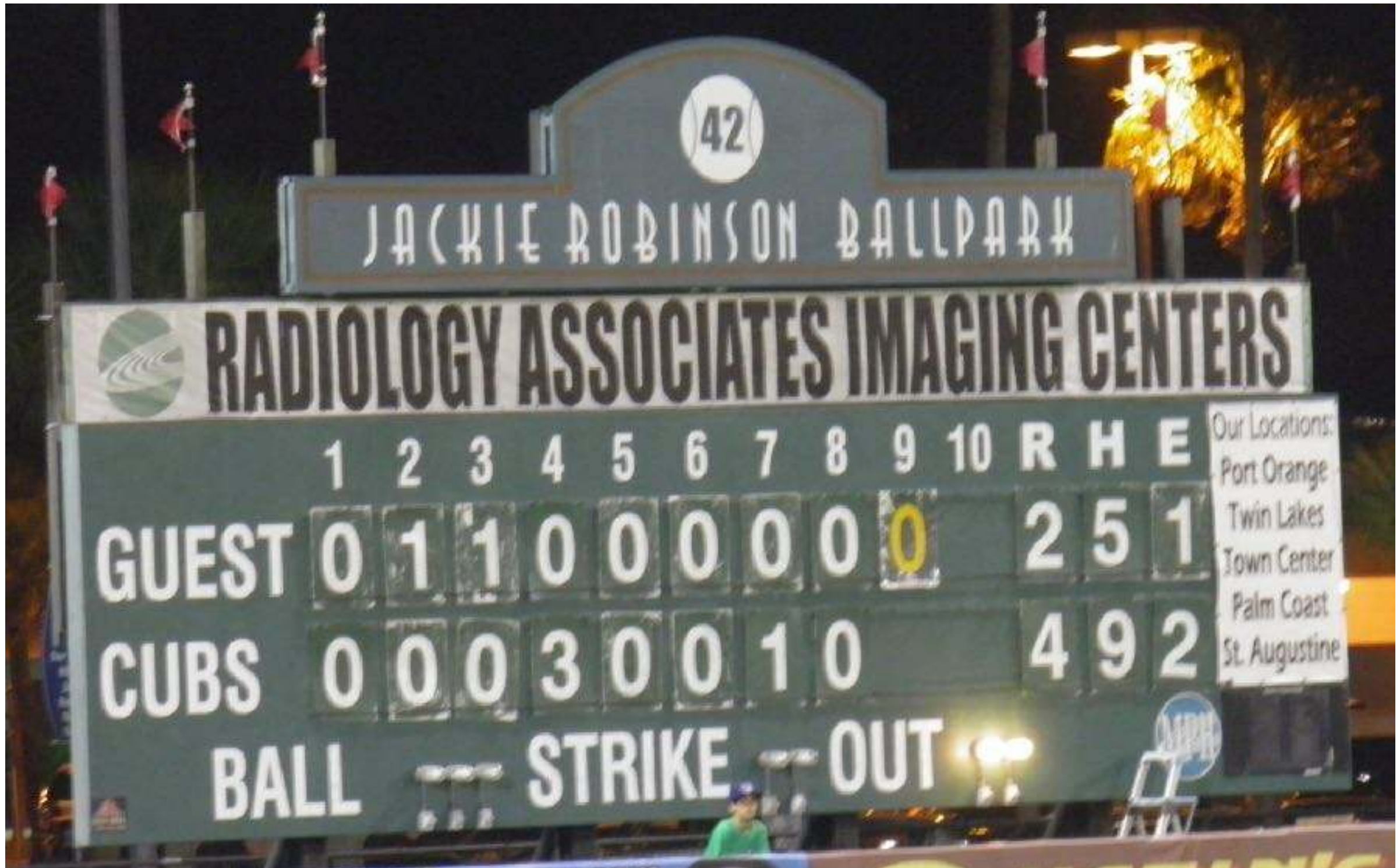
launched on the remaining 498K tables. Our approach is trivially parallelizable, and runs in 391 ms per table on average.

Ranking	Dataset	Prec@1	Prec@3
Only Label Index (TF-IDF score)	Wikitable	0.37	0.42
	T2D-v2	0.24	0.31
Labels + Embeddings (TransE)	Wikitable	0.61	0.72
	T2D-v2	0.62	0.74

Table 1: Precision of slot-filling with/out KG embeddings, calculated on redundant extractions.

Tables in Articles

Human-readable Tables



Background Knowledge for

Interpretation

**Baseball teams
(not places)**

Runs per inning

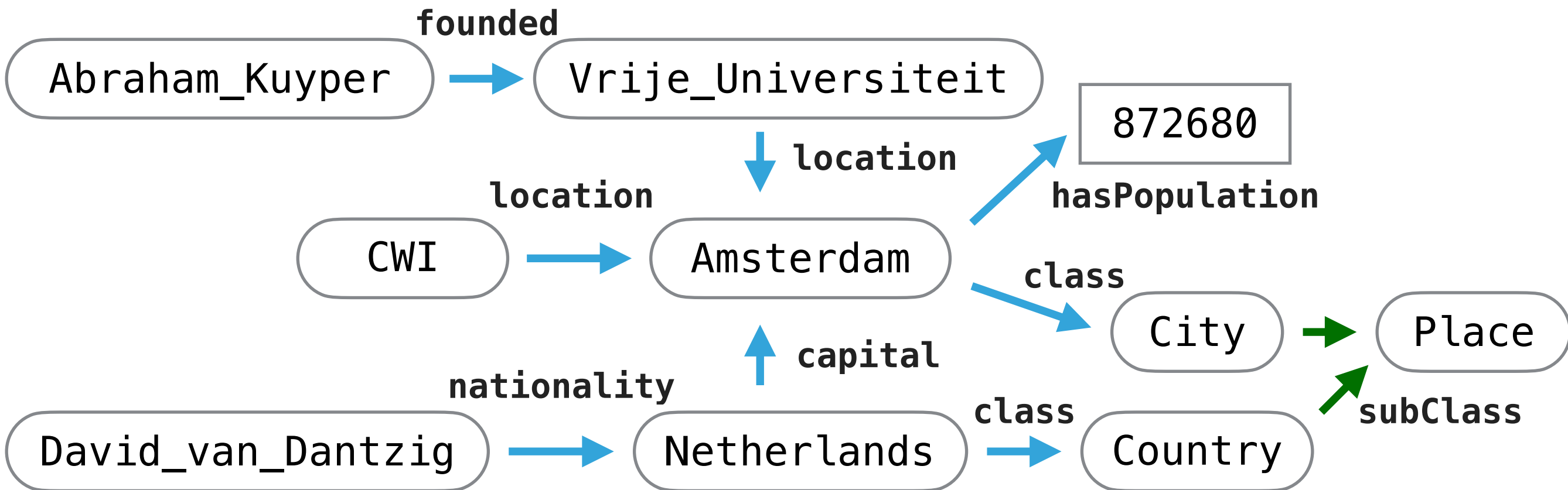
Runs

Hits

Errors

Team	1	2	3	4	5	6	7	8	9	R	H	E
Brooklyn	1	0	0	0	0	0	0	3	0	4	8	0
New York	0	0	0	0	0	0	1	0	4	5	8	0

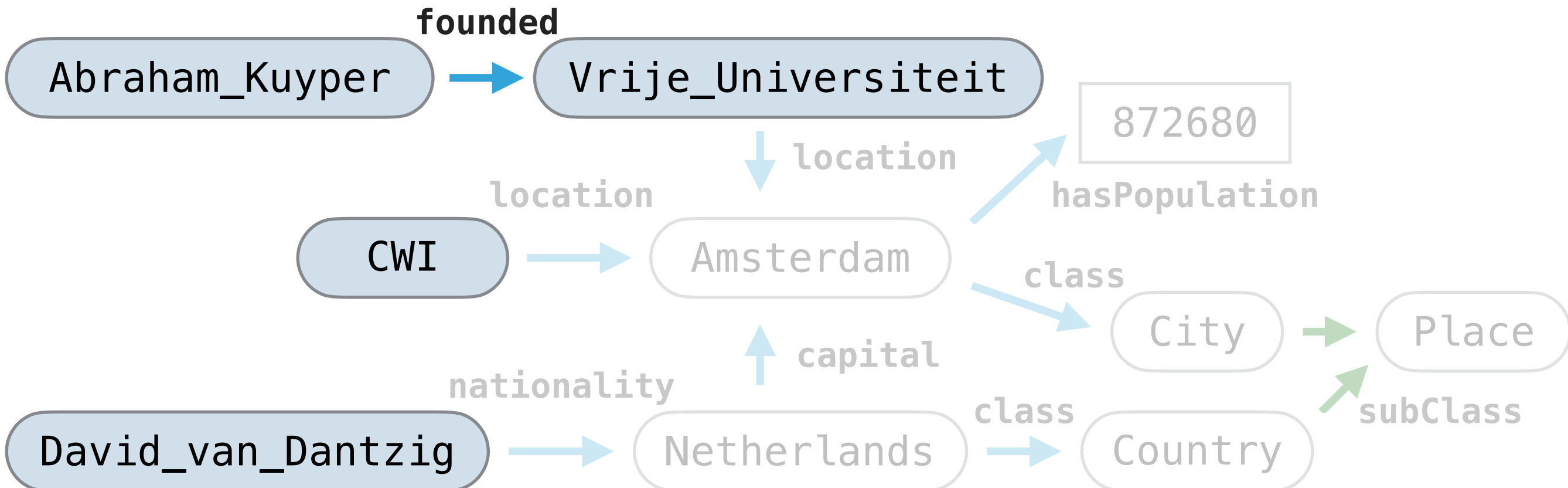
Knowledge Bases



Table

Interpretation

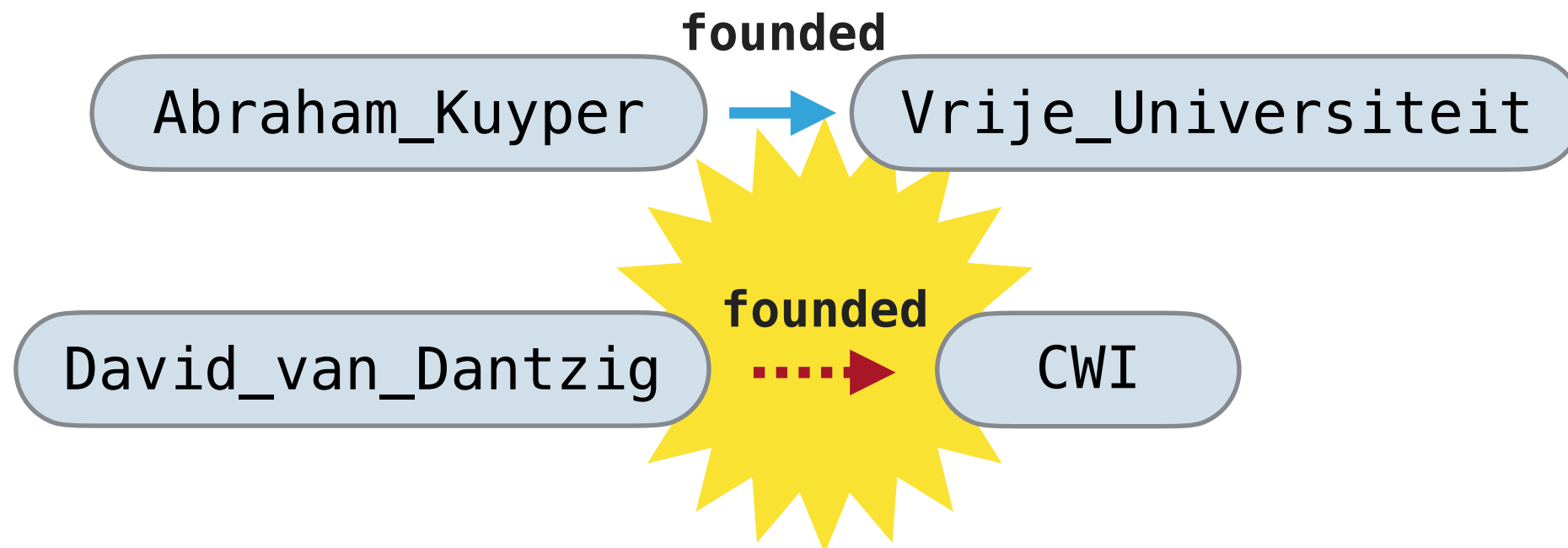
Name	Founder	Year
Vrije Universiteit	Abraham Kuijper	1880
Mathematisch Centrum	David van Dantzig	1946



Knowledge Base

Extension

Name	Founder	Year
Vrije Universiteit	Abraham Kuijper	1880
Mathematisch Centrum	David van Dantzig	1946



Challenges



Page Name	Year	Chart			Positions			
"Hou	Page Name	Single	Year	National Chart	Peak Position			
"Hou	Ray	Page Name	Titles	Year	Chart	Peak	Album	Certifications
"Hou	Ray	Ringo	Page Name	Year	Single			Chart positions
"Hou	Ray	Ringo	Commodores	1977	"Brick House"		US	5
"Hou	Ray	Ringo	Commodores	1977	"Brick House"		US R&B	4
"Hou	Ray	Ringo	Commodores	1977	"Brick House"		US Dance	34
"Hou	Ray	Ringo	Commodores	1977	"Easy"		US	4
"Hou	Ray	Ringo	Commodores	1977	"Easy"		US R&B	1
"Hou	Ray	Ringo	Commodores	1977	"Easy"		US Dance	—

1. New Facts

2. Diversity & Scale



3. New Domains

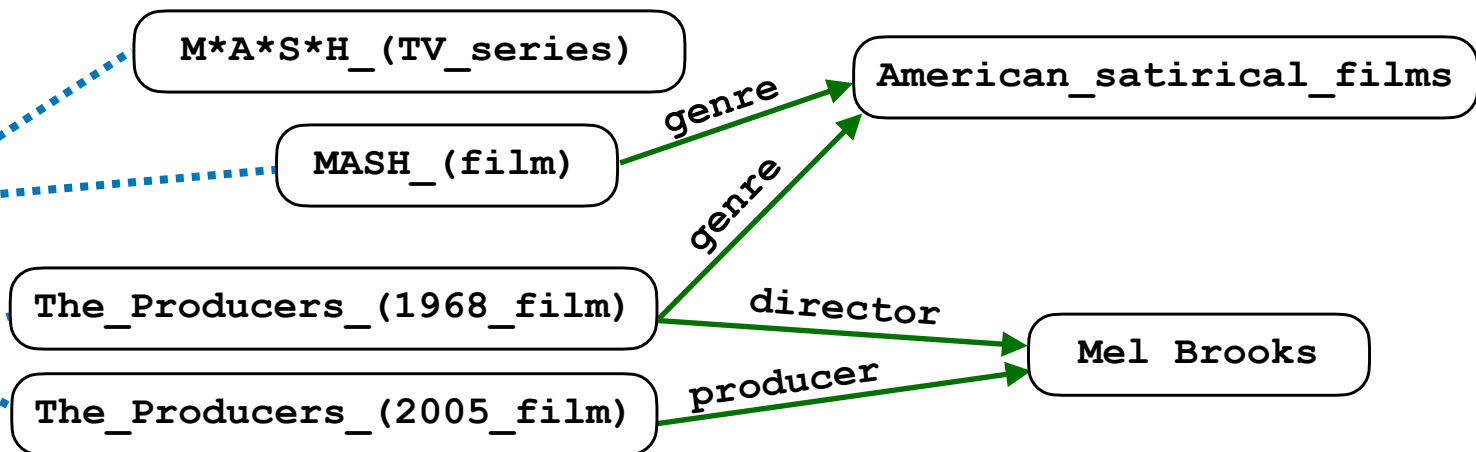
New Facts



**Goal: Interpret Tables
with New Facts**

New Facts

Director	Title
Fritz Lang	M
Robert Altman	M*A*S*H
Orson Welles	The Magnificent Ambersons
Mel Brooks	The Producers
Luc Besson	The Professional



Coherence

Matches

Background Knowledge



New Facts



+



Coherence

Likelihood

=

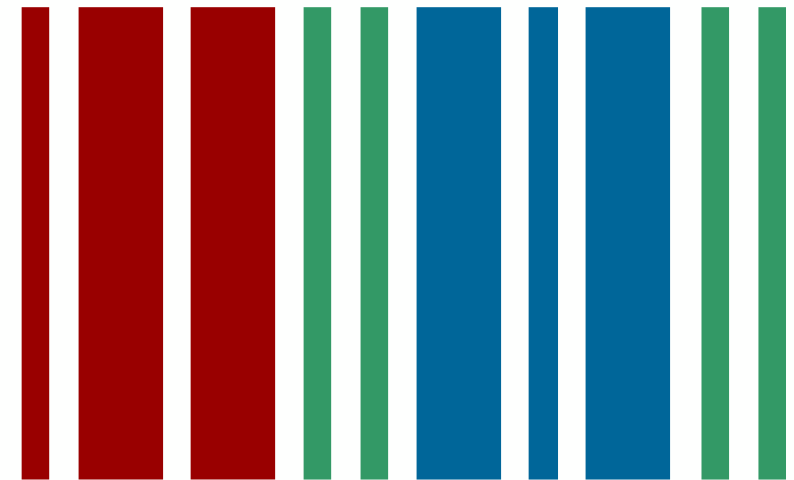
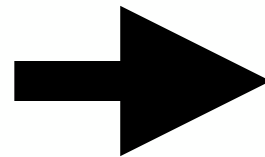


Diversity & Scale

Page Name	Year	Chart		Positions				
"Hou	Page Name	Single	Year	National Chart	Peak Position			
"Hou	Ray	Page Name	Titles	Year	Chart	Peak	Album	Certifications
"Hou	Ray	Ringo	Page Name	Year	Single			Chart positions
"Hou	Ray	Ringo						
"Hou	Ray	Ringo	Commodores	1977	"Brick House"	US		5
"Hou	Ray	Ringo	Commodores	1977	"Brick House"	US R&B		4
"Hou	Ray	Ringo	Commodores	1977	"Brick House"	US Dance		34
"Hou	Ray	Ringo	Commodores	1977	"Easy"	US		4
"Hou	Ray	Ringo	Commodores	1977	"Easy"	US R&B		1
"Hou	Ray	Ringo	Commodores	1977	"Easy"	US Dance		—



WIKIPEDIA
The Free Encyclopedia



WIKIDATA

**Goal: Extract Facts
from Real Tables**

Diversity & Scale

Step 1: Reshape



Commodores (album)

Year	Single	Chart positions		
		US	US R&B	US Dance
1977	"Brick House"	5	4	34
1977	"Easy"	4	1	—

Page Name	Year	Single		Chart positions
Commodores	1977	"Brick House"	US	5
Commodores	1977	"Brick House"	US R&B	4
Commodores	1977	"Brick House"	US Dance	34
Commodores	1977	"Easy"	US	4
Commodores	1977	"Easy"	US R&B	1
Commodores	1977	"Easy"	US Dance	—

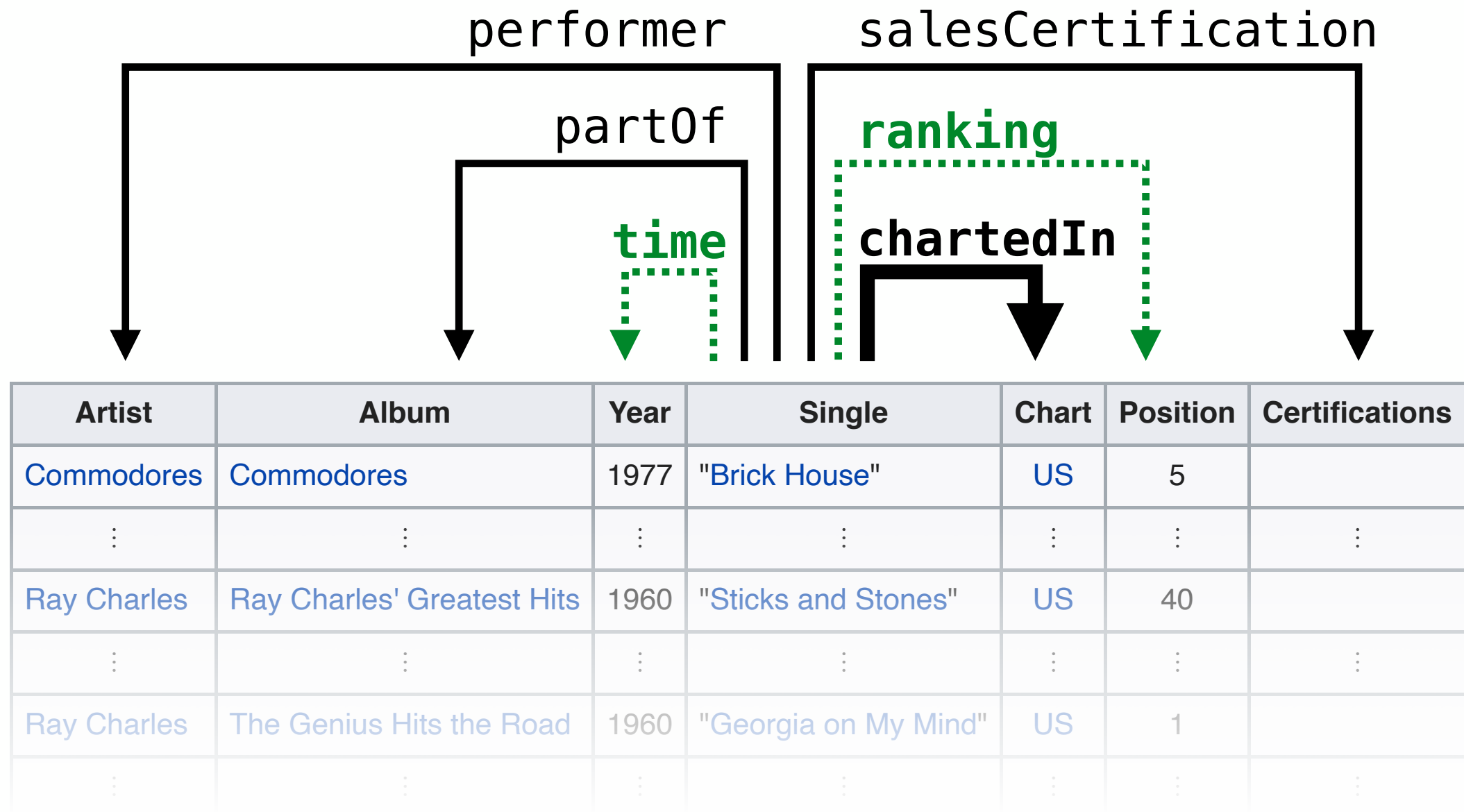
Diversity & Scale

Step 2: Cluster

Artist	Album	Year	Single	Chart	Position	Certifications
Commodores	Commodores	1977	"Brick House"	US	5	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Ray Charles	Ray Charles' Greatest Hits	1960	"Sticks and Stones"	US	40	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Ray Charles	The Genius Hits the Road	1960	"Georgia on My Mind"	US	1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Ringo Starr	Blast From Your Past	1971	"It Don't Come Easy"	UK	4	US: Gold
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Elvis Presley		1956	"Hound Dog"	US	8	
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Diversity & Scale

Step 3: Integrate

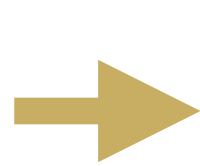


Diversity & Scale

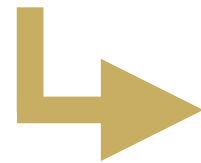
Page Name	Year	Chart	Positions					
*Hou	Page Name	Single	Year	National Chart	Peak Position			
*Hou	Ra	Page Name	Titles	Year	Chart	Peak	Album	Certifications
*Hou	Ra	Page Name	Year	Single	Chart positions			
*Hou	Ra	Commodores	1977	"Brick House"	US	5		
*Hou	Ra	Commodores	1977	"Brick House"	US R&B	4		
*Hou	Ra	Commodores	1977	"Brick House"	US Dance	34		
*Hou	Ra	Commodores	1977	"Easy"	US	4		
*Hou	Ra	Commodores	1977	"Easy"	US R&B	1		
*Hou	Ra	Commodores	1977	"Easy"	US Dance	—		



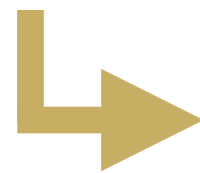
WIKIPEDIA
The Free Encyclopedia



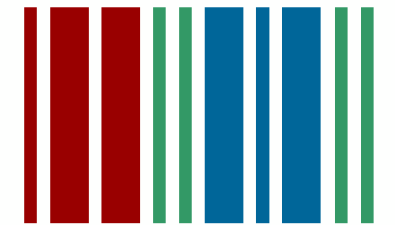
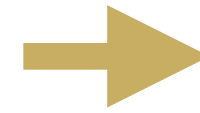
Reshape



Cluster



Integrate



WIKIDATA

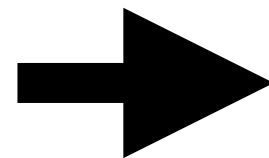
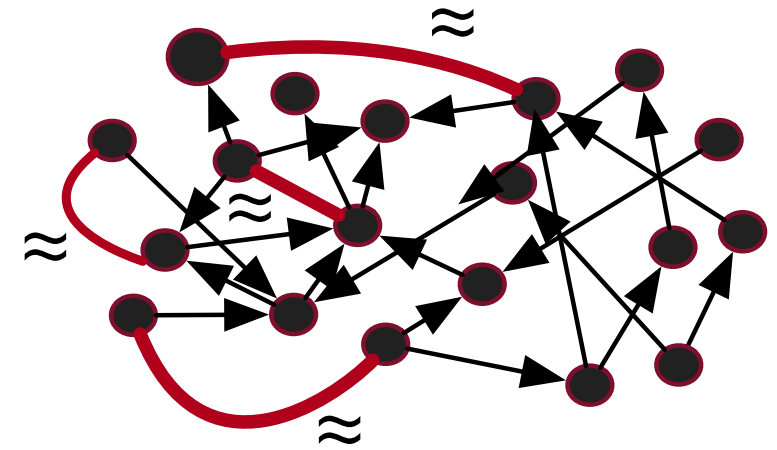
1.5 million tables → 21 million new extractions

New Domains

	l_1-l_n	#S	$\#l_1-W$	$\#l_n-W$	$\#l_1-V$	$\#l_2-V$
α_c	DP concentration parameter for each $c \in V$					50k
$P_0(e c)$	CFG base distribution					50k
x	Set of non-terminal nodes in the treebank					50k

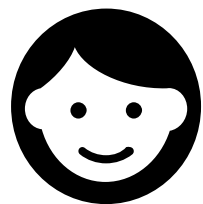
Models	Rerank size	Beam size	GMV	Latency	Words
miDNN	50	-	2.91%	9%	filthy, nasty,
miRNN	50	5	5.03%	58%	ple, terrible, aw-
miRNN+att.	50	5	5.82%	401%	idiotic, stupid,

m		dumb, ugly, etc.
$n = \{n_{c,e}\}$	Non-offensive	help, love, respect, believe,
$\Delta n^{S:m}$		congrats, hi, like, great,
		fun, nice, neat, happy,
		good, best, etc.



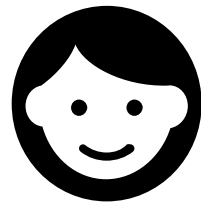
Tables from
Scientific Publications

New Knowledge Base

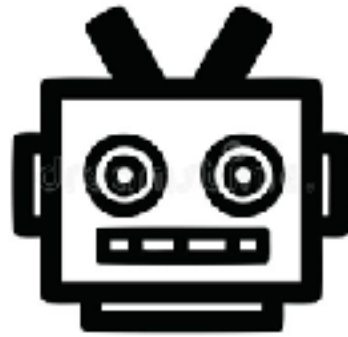


Goal: Minimize Manual Effort

New Domains



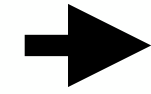
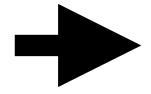
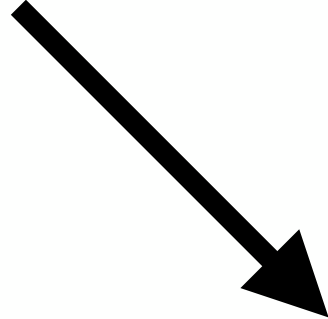
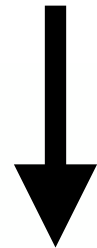
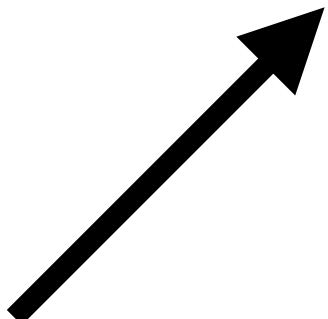
Definitions
Heuristics
Rules



Learning



Reasoning



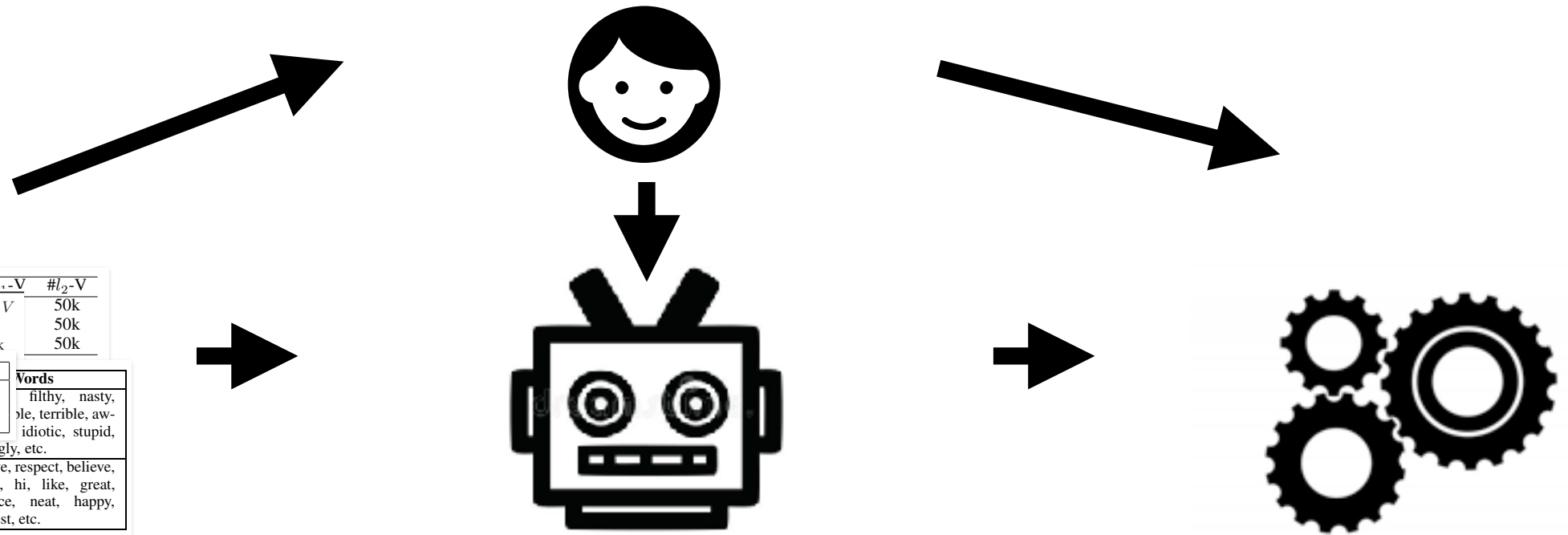
	l_1, l_2	#S	#l ₁ -W	#l ₂ -W	#l ₁ -V	#l ₂ -V
α_c	DP concentration parameter for each $c \in V$					50k
$P_0(e c)$	CFG base distribution					50k
x	Set of non-terminal nodes in the treebank					50k

Models	Rerank size	Beam size	GMV	Latency	Words
miDNN	50	-	2.91%	9%	filthy, nasty,
miRNN	50	5	5.03%	58%	le, terrible, aw-
miRNN+att.	50	5	5.82%	401%	idiotic, stupid,
					dumb, ugly, etc.

m	Non-offensive	help, love, respect, believe, congrats, hi, like, great, fun, nice, neat, happy, good, best, etc.
$n = \{n_{c,e}\}$		
$\Delta n^{S:m}$		

Tables

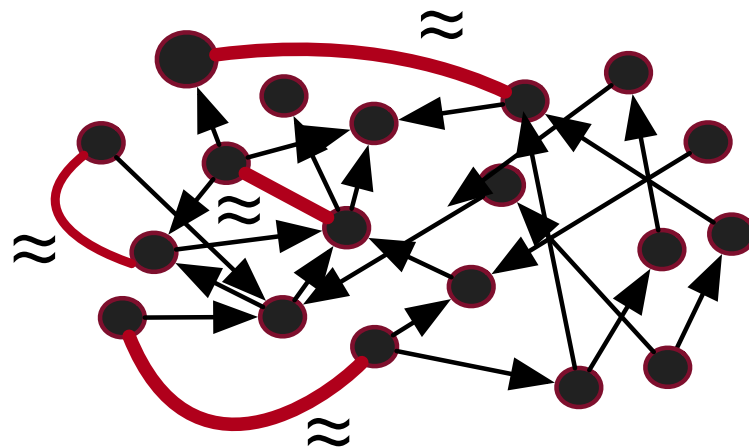
New Domains



	l_1, l_2	#S	# l_1 -W	# l_2 -W	# l_1 -V	# l_2 -V
α_c	DP concentration parameter for each $c \in V$					50k
$P_0(e c)$	CFG base distribution					50k
x	Set of non-terminal nodes in the treebank					50k

Models	Rerank size	Beam size	GMV	Latency	Words
miDNN	50	-	2.91%	9%	filthy, nasty,
miRNN	50	5	5.03%	58%	ole, terrible, aw-
miRNN+att.	50	5	5.82%	401%	idiotic, stupid,

m		
$n = \{n_{c,e}\}$	Non-offensive	dumb, ugly, etc.
$\Delta n^{S:m}$		help, love, respect, believe, congrats, hi, like, great, fun, nice, neat, happy, good, best, etc.



Conclusions



Page Name	Year	Chart		Positions				
"Hou	Page Name	Single	Year	National Chart	Peak Position			
"Hou	Ray	Page Name	Titles	Year	Chart	Peak	Album	Certifications
"Hou	Ray	Ringo						
"Hou	Ray	Ringo						
"Hou	Ray	Ringo	Commodores	1977	"Brick House"		US	5
"Hou	Ray	Ringo	Commodores	1977	"Brick House"		US R&B	4
"Hou	Ray	Ringo	Commodores	1977	"Brick House"		US Dance	34
"Hou	Ray	Ringo	Commodores	1977	"Easy"		US	4
"Hou	Ray	Ringo	Commodores	1977	"Easy"		US R&B	1
"Hou	Ray	Ringo	Commodores	1977	"Easy"		US Dance	—

