



Amsterdam
Data Science



BENNO KRUIT, PETER BONCZ, JACOPO URBANI

TAKCO:

A PLATFORM FOR EXTRACTING NOVEL FACTS FROM TABLES

WEB TABLES

- ▶ Useful for extending Knowledge Bases
- ▶ Diverse



TAKCO: A PLATFORM FOR EXTRACTING NOVEL FACTS FROM TABLES

No. overall	Title	Original airdate
17	1 "A Slight Case of Reincarnation"	31 December 1966
Adam is determined to free an African leader from the spell of the Face. *		
18	2 "Black Echo"	7 January 1967
Adam visits an exiled Russian Grand Duchess, who is requesting him because he is familiar with a pearl necklace that she once owned, and is now trying to reclaim it. However, upon arrival at the Duchess's home, Adam finds two horrifying secrets from his past are about to return...		
19	3 "Conspiracy of Death"	14 January 1967
Adam investigates the murder of an old wartime friend.		

Catholic, Orthodox, Protestant, and most Oriental Orthodox	Original language (Koine Greek)
<i>Canonical Gospels</i>	
Matthew	Greek (majority view: see note) ^{[N 2][34][35][36]}
Mark	Greek
Luke	Greek
John	Greek
<i>Apostolic History</i>	
Acts	Greek

Sled	Athletes	Event	Run 1		Run 2		Run 3	
			Time	Rank	Time	Rank	Time	Rank
LAT-1	Sandis Prūsis Mārcis Rullis	Two-man	48.10	13	48.06	12	47.92	8
LAT-2	Intars Dīcmanis Gatis Gūts	Two-man	48.07	12	48.22	18	48.14	13

Year	Population	Year	Population
1749	200	1885	453
1763	273	1907	344
1780	327	1925	330
1800	273	1945	351
1818	329	1955	295
1869	402		

Pos.	Rider	Points	POL	ITA	GER	SWE	GBR	DEN
1	(5) Billy Hamill	113	16	20	9	25	18	25
2	(1) Hans Nielsen	111	18	25	25	9	20	14
3	(4) Greg Hancock	88	12	13	13	16	16	18
4	(2) Tony Rickardsson	86	20	18	16	14	7	11
5	(8) Henrik Gustafsson	80	14	4	18	20	11	13

FACT EXTRACTION FROM WEB TABLES

- ▶ Many steps → need **debugging** tools
- ▶ Varied domains → need **patching**
- ▶ Different goals →
must understand **trade-offs**



- ▶ Python library
 - ▶ Modular components
 - ▶ User Interface
 - ▶ Explore configurations
 - ▶ Analyze extractions

TAKCO COMPONENTS

- ▶ Pre-processing
- ▶ Clustering
- ▶ Entity Linking
- ▶ Property Linking
- ▶ Novelty Analysis

PRE-PROCESSING

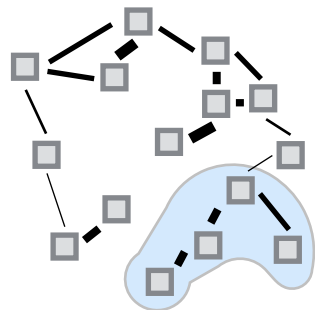
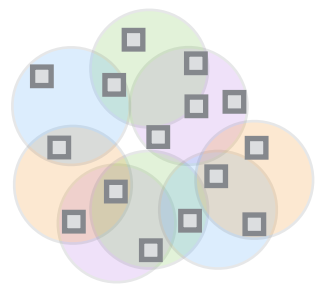
Commodores (album)



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	—

TABLE STITCHING



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	-	



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	
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Block

Match

Cluster

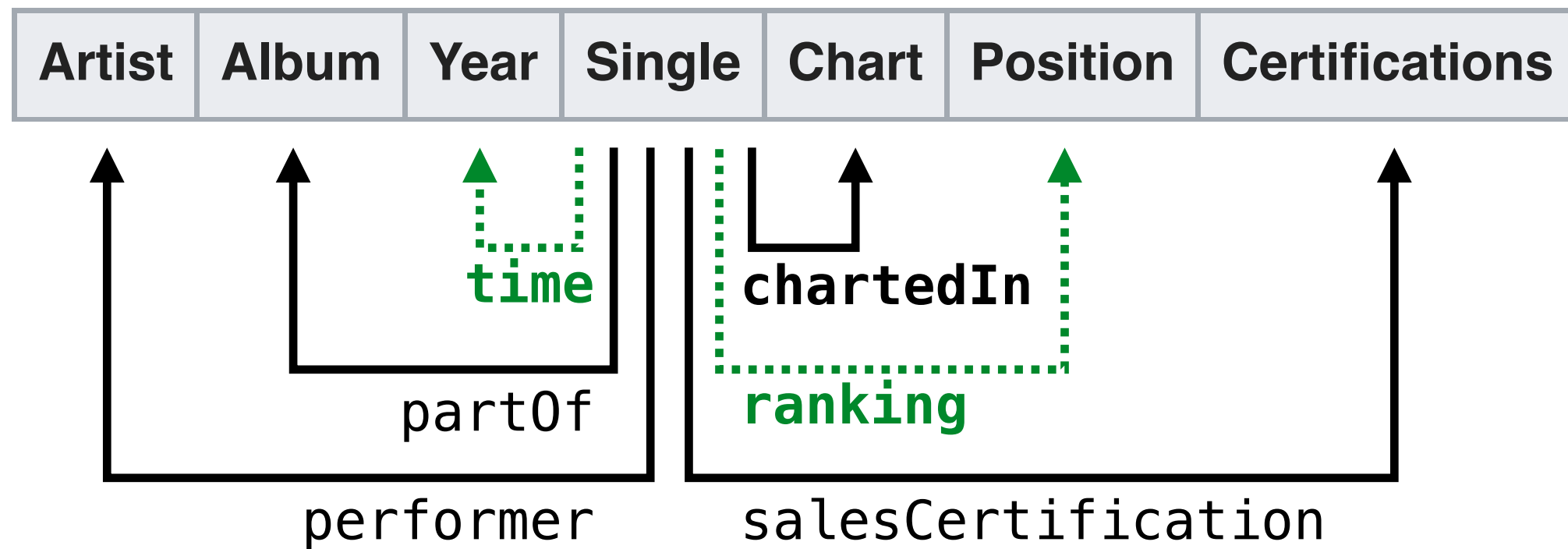
Unify

ENTITY LINKING



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	
⋮	⋮	⋮	⋮	⋮	⋮	⋮

PROPERTY LINKING



NOVELTY ANALYSIS

