



UTOPIA: Automatic Pivot Table Assistant

Whanhee Cho, Anna Fariha



Messy spreadsheet data causes unusable pivot tables

Movie Title	Genre	Gross
2001: A Space Odyssey	Technology Fiction	100 M
Interstellar	Science Fiction, drama, Adventure	500 M
The Matrix	action, sci-fi	200 M

Traditional pivot table tools struggle with messy data

Messy spreadsheet data is common [1]

Data **parsing** requires additional expertise

Resolving **synonymous variants** requires manual effort

Users don't prefer **switching environments**

Which movie genre achieves the most gross? Let's find out this through a Microsoft Excel pivot table with Genre and Gross.



Genre	Sum of Gross
Technology Fiction	100 M
Science Fiction, drama, Adventure	500 M
action, sci-fi	200 M

Multiple Values

No one wants to know the total gross of movies of **epic drama, superhero action, and crime genre!**
- **Multiple values** in the Genre cause this problem.



Genre	Sum of Gross
Technology Fiction	100 M
Science Fiction	500 M
Sci-Fi	200 M
...	...

Synonymous Variants

I split the data with Microsoft Power Query. But, some values share **the same semantic meaning**.



Genre	Sum of Gross
Action	900 M
Sci-Fi	800 M
Drama	920 M
Adventure	600 M

I finally got the desired result, but I had to find and replace variants about **240 times**. It took me **3 hours!**

UTOPIA provides ideal pivot tables without data preprocessing

Upload Data

Data Overview

Title	Year	Genre	Gross	Director	Casts
The Shawshank Redemption	1994	prison drama	28 M	Frank Darabont	Tim Robbins, Mo ...
The Dark Knight	2008	superhero action, crime, epic drama	535 M	Christopher Nolan	Christian Bale, H ...
The Matrix	1997	action, cyberpunk sci-fi	171 M	Lana Wachowski	Lilly Wachowski, ...

Select Attributes

☐ Title

☒ Year

☒ Genre

☒ Gross

☐ Director

☐ Casts

☐ Certificate

☐ Runtime

Pivot Table Fields

Column

Year

Row

Genre

Value and Aggregate

Gross

SUM AVG MIN MAX CNT

Pivot Table

Sum of Gross

	2010	2011	2012
crime	388 M	138 M	41 M
action	558 M	164 M	1416 M
action	8 M	4 M	304 M
superhero action	251 M	0 M	1071 M
team action	293 M	146 M	0 M
epic action	6 M	14 M	41 M
drama	621 M	886 M	673 M
drama	200 M	600 M	200 M
epic drama	231 M	764 M	393 M
prison drama	82 M	1 M	4 M

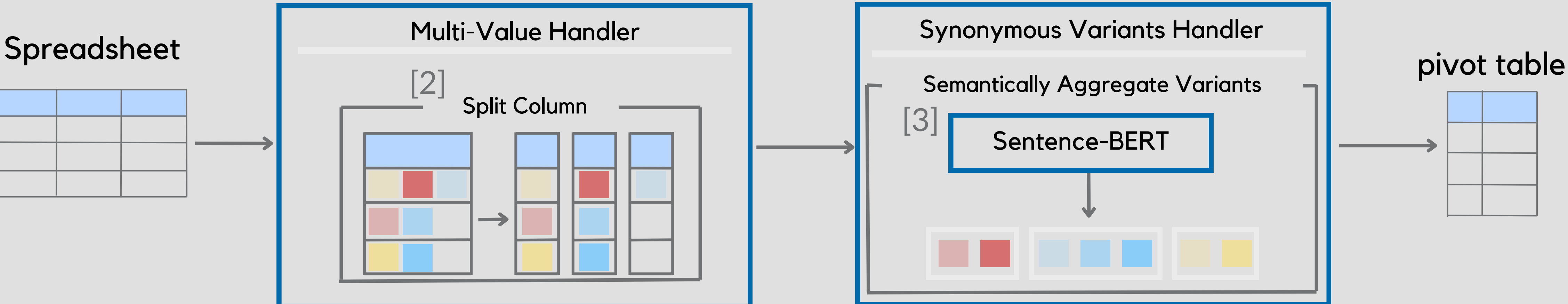
Synonymous variants are **grouped**

UTOPIA **semantically aggregates** synonymous variants.
Expand or collapse rows to view groups.

Each cell contains a **single value**

UTOPIA **parses** multi-valued attributes
even in the **absence of delimiters**.

UTOPIA uses SOTA parsing technology and embeddings



Future Work: Recommend k-budgeted pivot tables



What are the interesting pivot tables?

[1] P Li, Y He, C Yan, Y Wang, and S Chaudhuri. 2023. Auto-Tables: Synthesizing Multi-Step Transformations to Relationalize Tables without Using Examples. PVLDB

[2] M Raza and S Gulwani. 2017. Automated Data Extraction Using Predictive Program Synthesis. AAAI

[3] T Gao, X Yao, and D Chen. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. EMNLP