NL-TO-VIS: A NATURAL LANGUAGE INTERFACE FOR GENERATING BUSINESS DATA VISUALIZATION

Le Tuan Anh - 230104001

Summary

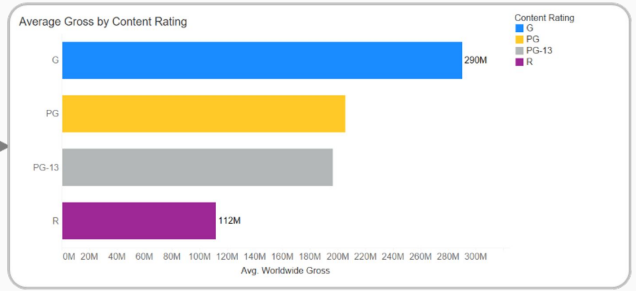
- Class: CS2205.MAR2024
- Github Link: https://github.com/mausLe/CS2205.MAR2024
- YouTube Link: https://youtu.be/fJ_MuK_SWR4
- Student: Le Tuan Anh - ID: 230104001

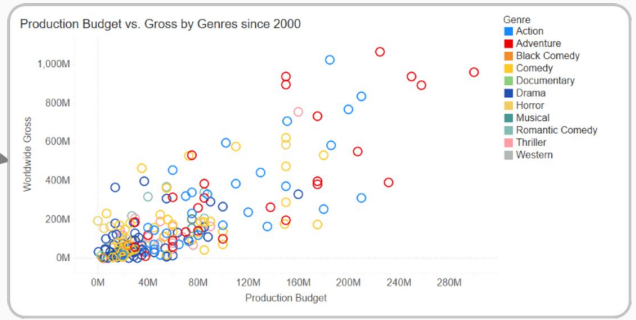
Introduction (1/2)

"A picture is worth a thousand words.

Movies-w-year.csv

Year	Title	Pro. Budget	...	Genre	Rating
1996	Fargo	7,000,000	...	Thriller	8.3
1999	The Matrix	65,000,000	...	Action	8.7
...	...	...	...	...	
2008	WALL-E	180,000,000	...	Comedy	8.5
2009	The Proposal	40,000,000	...	Romantic	5.4
2010	Inception	160,000,000	...	Thriller	9.1





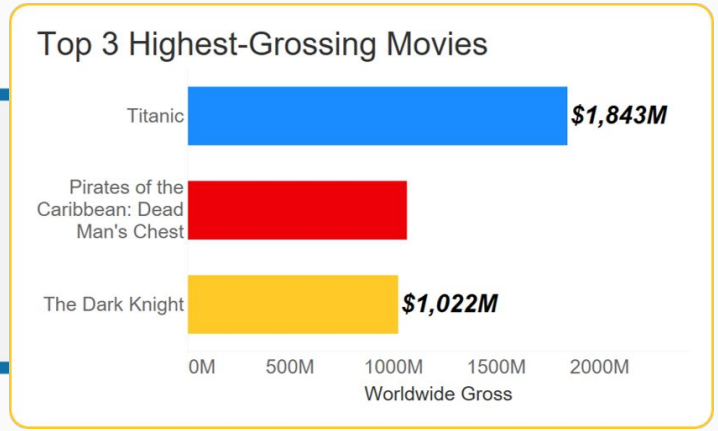
Introduction (2/2)

How to Bridge the Gap

between **DATA** and **INSIGHT**?

Year	Title	Pro. Budget	...	Genre	Rating
1996	Fargo	7,000,000	...	Thriller	8.3
1999	The Matrix	65,000,000	...	Action	8.7
...	...	...	...	...	
2008	WALL-E	180,000,000	...	Comedy	8.5
2009	The Proposal	40,000,000	...	Romantic	5.4
2010	Inception	160,000,000	...	Thriller	9.1

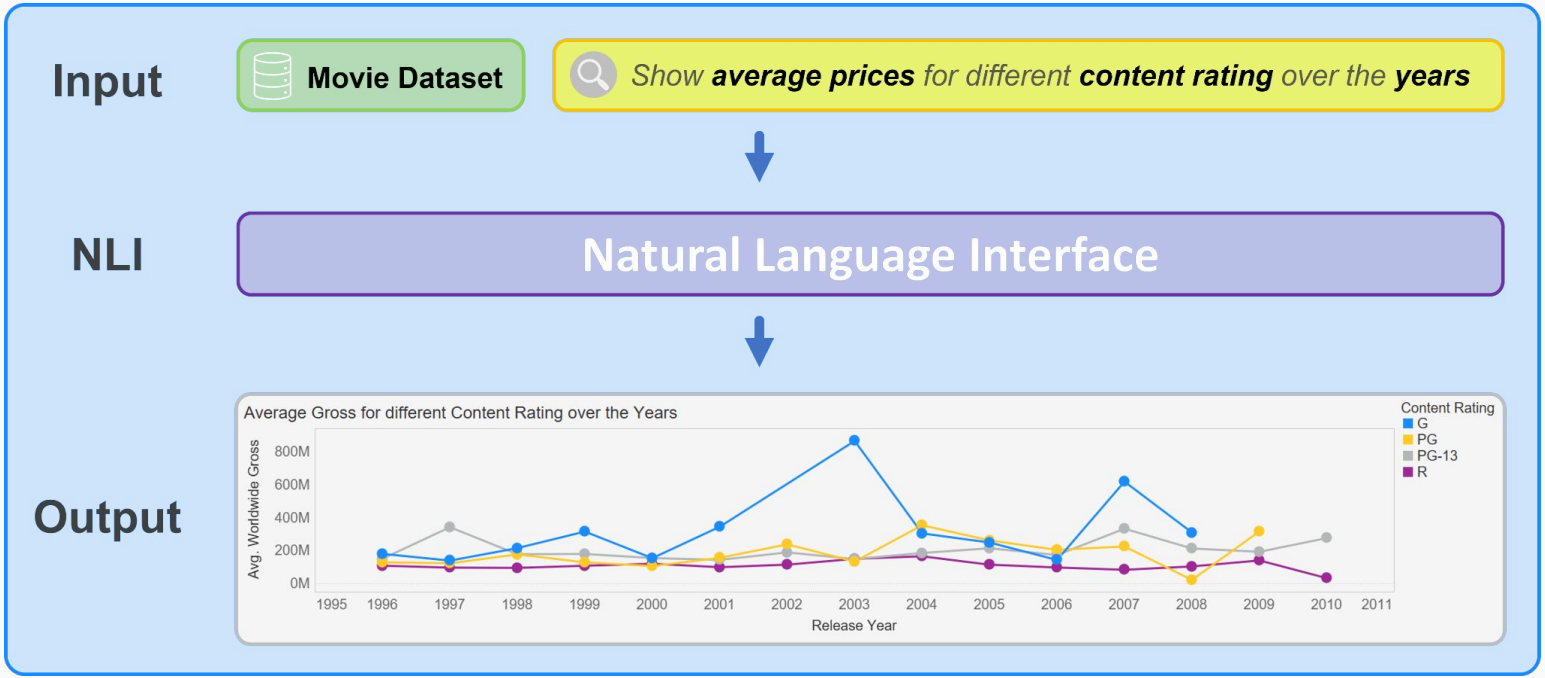
BRIDGE THE GAP



NL-to-Vis

NL-to-Vis

A Natural Language Interface for Generating Data Visualization



Key Challenges

Language Complexity with variations and ambiguities, requiring contextual understanding.

Under-specification when NLI needs more information about the aspects of data the user wants to visualize.

Evaluating NL-to-Vis is a resource-intensive process, due to:

- **Benchmarking:** Lack of sufficient data on NL-to-Vis.
- **User Studies:** Involving human evaluators to provide feedback.

Objectives

1. **Building the UITxNL-to-Vis -** a Business-Oriented English & Vietnamese Natural Language to Data Visualization dataset.

2. **Using NL4DV baseline** to explore advanced NLP techniques minimizing ambiguity, variation, and under-specification.

3. **Developing an NL-to-Vis Interface**, enabling users to explore data through natural language conversation.

Research Methods

1. Building the **UITxNL-to-Vis** by:
    - Scraping business dashboards from Tableau Public, PBI Gallery.
    - Annotating dashboards' user intents and visual specifications.

2. Exploring **Advanced NLP Techniques** for NL-to-Vis by:
    - Evaluating NL4DV baseline.
    - Exploring advanced NLP techniques like LLMs & RAG.

3. Developing an **NL-to-Vis Interface**, allowing users to interact with data like having a conversation.

Expected Outcomes

1. A valuable **Annotated Dataset** for NL-to-Vis research

2. **Improve NL-to-Vis** capabilities from NL4DV baseline by:
   - Minimizing ambiguity, variation, and under-specification.
   - Outperforming NL4DV baseline using techniques like LLMs, RAG.

3. A **Conversational** Data Exploration Interface.
   - Making data exploration more accessible and intuitive regardless of technical expertise in data analytics tools

References

[1] Arpit Narechania, Arjun Srinivasan, John T. Stasko:
NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. IEEE Trans. Vis. Comput. Graph. 27(2): 369-379 (2021)

[2] Rishab Mitra, Arpit Narechania, Alex Endert, John T. Stasko:
Facilitating Conversational Interaction in Natural Language Interfaces for Visualization. IEEE VIS (Short Papers) 2022: 6-10

[3] Yingchaojie Feng, Xingbo Wang, Bo Pan, Kamkwai Wong, Yi Ren, Shi Liu, Zihan Yan, Yuxin Ma, Huamin Qu, Wei Chen:
XNLI: Explaining and Diagnosing NLI-based Visual Data Analysis. CoRR abs/2301.10385 (2023)