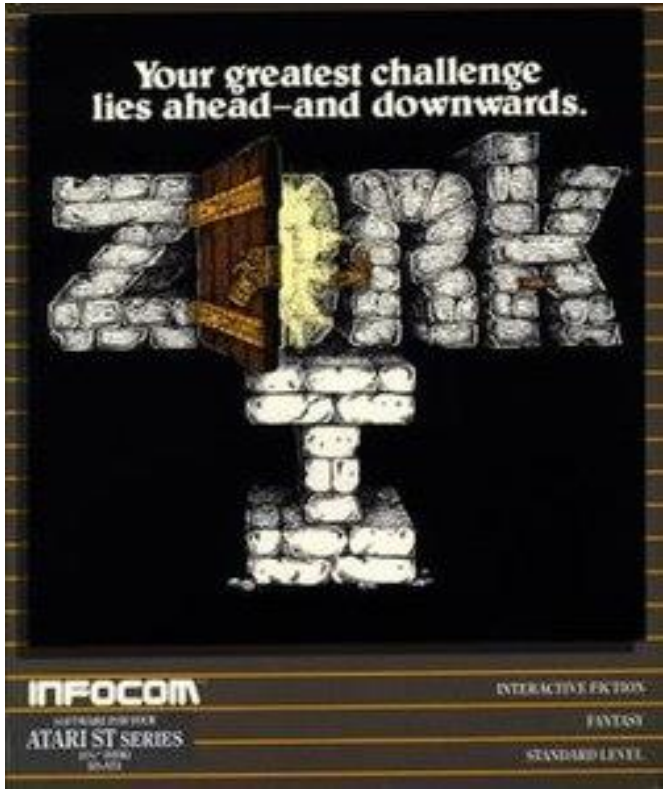
BERALL: Towards Generating Retrieval-augmented State-based Interactive Fiction Games

Rachel Chambers, Naomi Tack, Eliot Pearson, Lara J. Martin, Francis Ferraro (UMBC)

Presented by:
Dylan Lang

The Core Problem

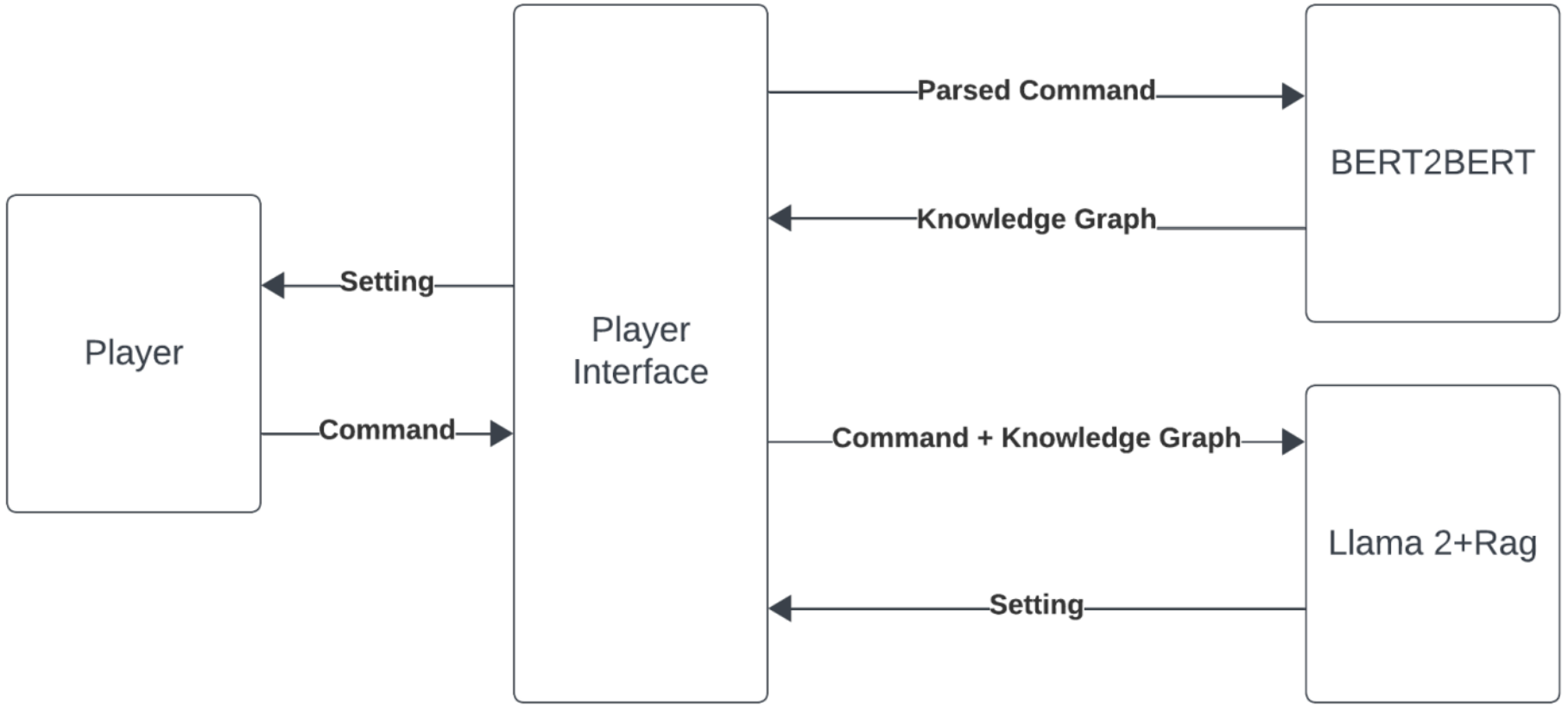
- **LLMs Generate Creative Text**
- **However, Lose Coherence Quickly**
- TAGs need consistent state tracking
  - (TAGs: Text Adventure Games)



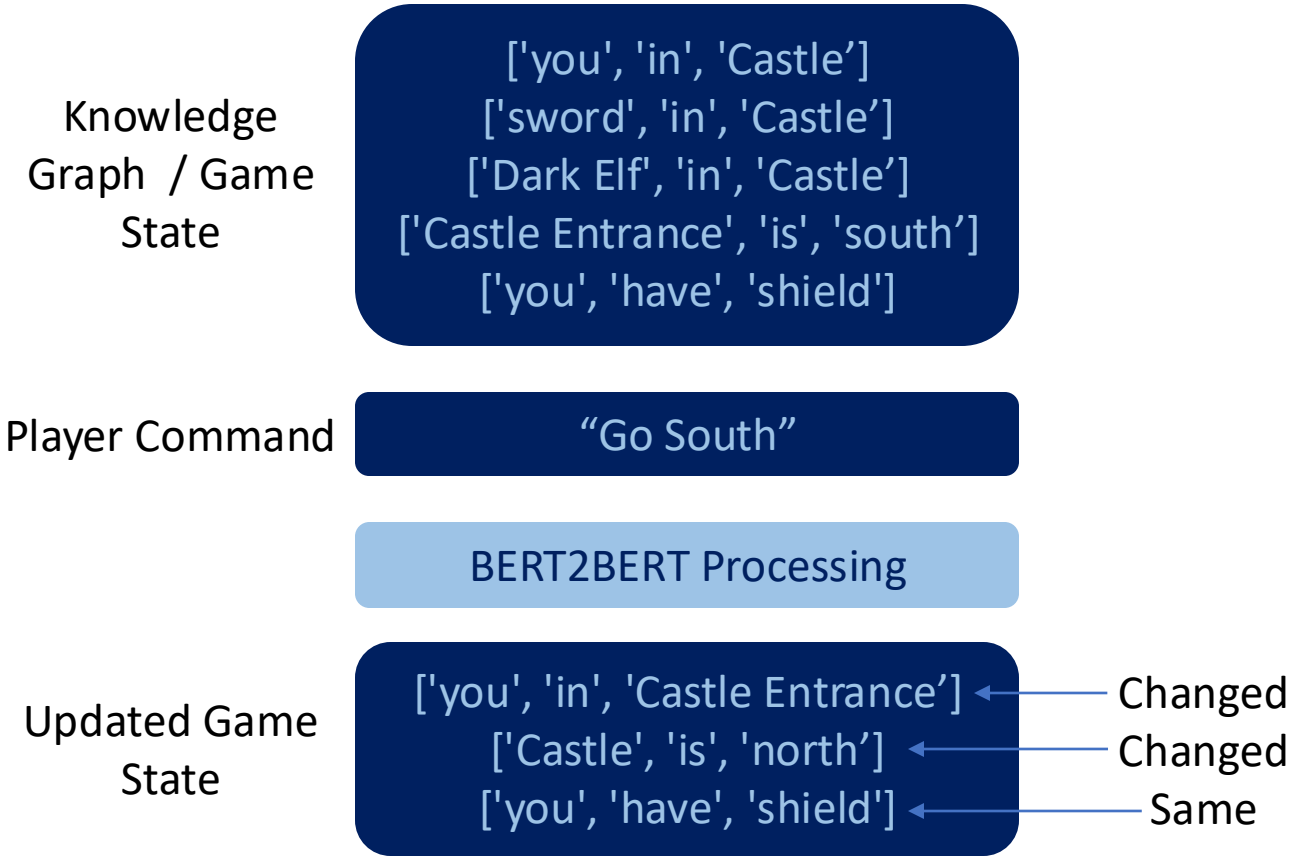
Zork 1 cover art

- Must Remember: Locations, inventory, world state, etc.

System Architecture

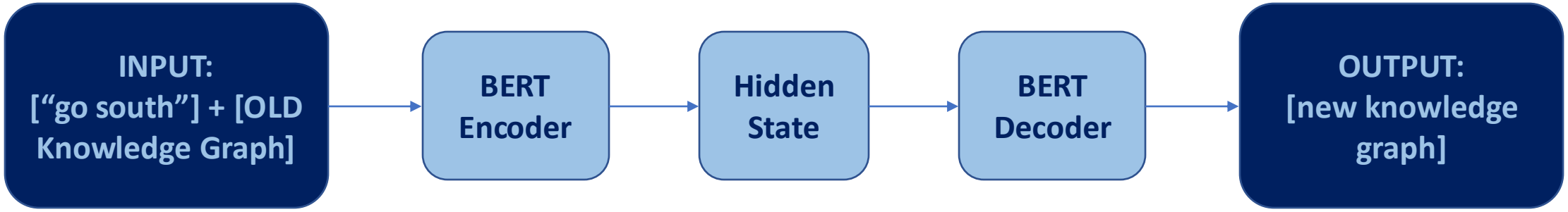


Knowledge Graph: Representing Game State

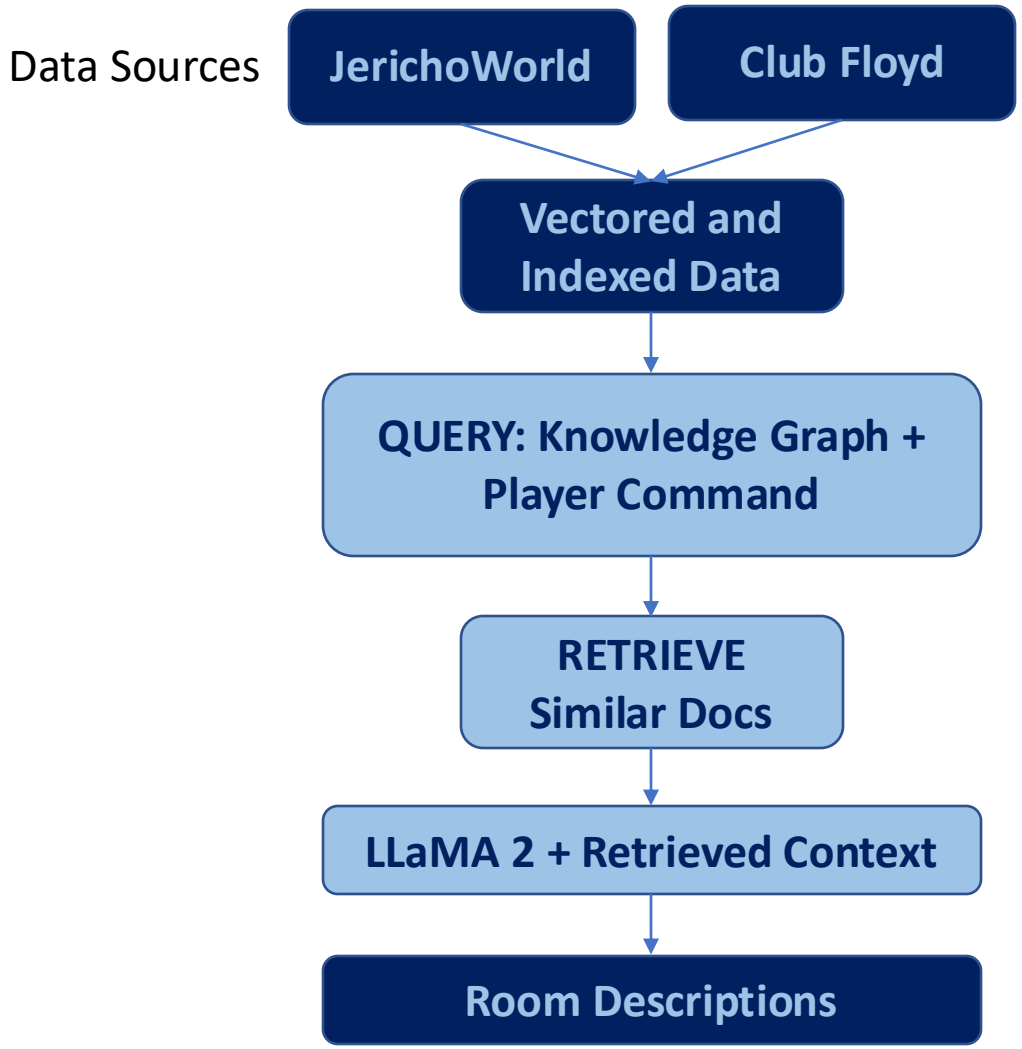


BERT2BERT: State Tracker

- Model: Encoder-decoder architecture
- Input: Previous graph + command
- Output: Updated Graph
- Details: Warm-start from BERT, Beam search (width=4)
- Sizes Tested: Tiny (~4.4M), Medium (~41M), Base-Uncased (~110M)



LLaMA 2 + RAG: Description Generator



Results

Model	ROUGE-P	ROGUE-R	ROGUE-F1
BERT-Tiny	6.4	13.1	7.8
BERT-Medium	**12.2**	**18.2**	**13.7**
BERT-Base	10.3	**18.2**	11.6

* without weight sharing *

Natural, coherent descriptions

Appropriate responses to commands

State doesn't always update correctly

Occasionally generates unwanted options

Strengths

- Neurosymbolic Design
  - Interpretable, Debuggable
- RAG Approach
  - LLM Generated in similar style
- Modular Architecture
  - Components are independent
- Captures TAG style
  - Natural language with similar tone

Weaknesses

- No Puzzle / Special Action / Win Conditions in Generation
  - Core element to TAGs is missing
- Poor State Tracking
  - F1 = 13.7
- Insufficient Evaluation
  - No User Studies or Baselines
- Lacks Causal Understanding
  - Doesn’t understand command effects
- Limited Dataset
  - ~500 transcriptions (Club Floyd) may be insufficient

Broader Context, Future Work

- This work concludes:
  - First neurosymbolic approach to TAG generation
  - **RAG can provide stylistic grounding**
  - **State consistency remains unsolved**
- Future Directions:
  - Improved prompt engineering
  - Dataset diversity
  - Incorporate planning techniques