Hierarchical Event Modeling for Multimodal Narrative Comprehension

Portions of this work are based on a project currently under review for publication. The results and figures are preliminary and intended for discussion only.

Yi-Chun (Rimi) Chen
Ph.D. in Computer Science

Affiliation: Data-Mining Lab,
Department of Emergency Medicine
Lab Director: Dr. Samah Jarad-Fodeh

Motivation / Background

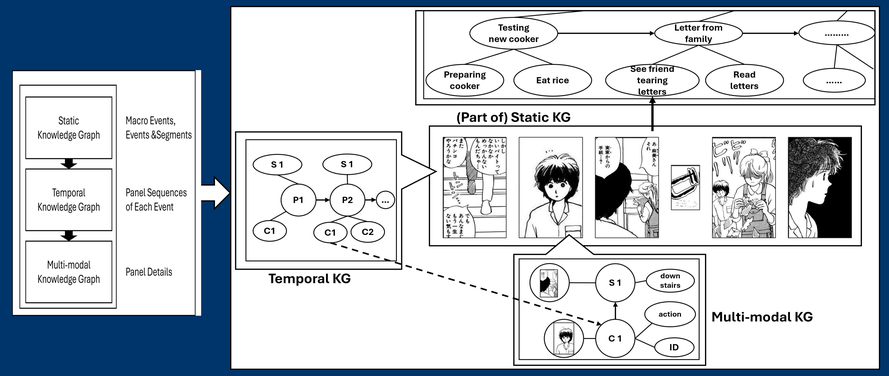
- Visual narrative (comics, image sequences) combine visuals, text, and layout to convey meaning
- Understanding these required linking visual and textual elements into coherent event structures

Methodological Approach

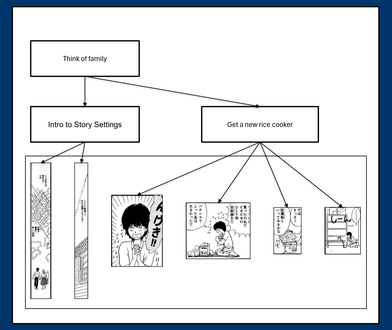
Hierarchical Narrative Graph Construction: three-level knowledge graph to model visual narratives using annotated image sequences

- **Panel-Level Multimodal Graphs**
  Each panel is encoded as a local multimodal graph linking characters, actions, objects, dialogue, and narration. Edges represent roles, co-occurrence, and speaker links.
- **Temporal Graph of Sequences**
  Panels are connected via directed edges (e.g., precedes_storytime) to model reading and narrative order, enabling timeline reconstruction across events.
- **Event-Level Semantic Graphs**
  Inspired by cognitive theories, we define a narrative hierarchy:
  - Segments – fine-grained actions or shifts
  - Events – coherent scenes or interactions
  - Macro-events – broader narrative arcs
- **Semantic Normalization**
  To handle label variation (e.g., hit, punch), we cluster similar terms using lexical and embedding-based similarity for consistent reasoning.
- **Graph Integration**
  All levels are linked via relations like instantiates and subevent-of, enabling symbolic reasoning across narrative layers.

Hierarchical Narrative Modeling



Hierarchical Framework Overview: Our system integrates visual and textual narrative information into a three-tiered knowledge graph. It models panel-level content, temporal relations across sequences, and abstract event structures to support symbolic reasoning.



Narrative Event Hierarchy: Individual panels are grouped into fine-grained segments, which compose mid-level events and aggregate into macro-events. This structure mirrors cognitive theories of event comprehension and supports multi-level graph construction.

Result Highlights

- Supports narrative segmentation, character interaction modeling, and event flow tracking
- Enables symbolic reasoning across narrative levels—from grounded visual-text elements to abstract intent structures
- Demonstrates multi-level graph inference in tasks such as:
  - Event segmentation: identifying coherent macro-events from panel sequences
  - Dialogue tracing: linking speech content and characters across event boundaries

Bridging to Broader Impact

This framework lays the groundwork for applying structured narrative reasoning to real-world domains. Potential applications include:

- **Health education:** guiding personalized visual explanations through narrative scaffolds
- **Goal-aligned patient communication:** using structured stories to support decision-making
- **Adaptive learning:** generating interactive sequences that reflect patient-specific intentions

Patient goal/ Communications → **Structured graphs** → **Visual Scene / Dialogue Snippet**

Envisioned application: visualizing patient-aligned goals through adaptive narrative scenes

Research Questions

- How can multimodal narratives be represented structurally for better understanding?
- Can hierarchical modeling support internet extraction and reasoning across domains(e.g., education, healthcare)?
- How might this model inform adaptive learning or visual goal communication?

Bridging to Lab Themes

- Patient-provider messages can also be understood as narrative-like sequence
- Visual storytelling principles may help model, generate, or interpret goal-driven health communication

Relevance to Lab Projects

Graph modeling of patient-provider communication

- Use similar hierarchical KG structure to model temporal-intent flow in medical dialogues

Patient goal aligned narrative generation

- Extend the generation framework from image sequences to interactive visual learning scenes guided by health goals

Visual Narrative assisted goal expression

- Explore structured storytelling as a way for patients to visually articulate care preferences or priorities

Future Directions

- Integrate medical ontologies with narrative modeling
- Use generated scenes for visual explanation or decision support
- Expand from static image sequence to interactive learning modules