STORYSCAPE: Spatial Reconstruction of Stories

Cagri Zaman
zaman@mit.edu

Manushaqe Muco
manjola@mit.edu

12/16/2014

Abstract

Because stories (narrative or everyday life events) and spatial perception are closely connected, in order to develop systems of human-level intelligence, we need to bridge story understanding with perception. For this purpose, we introduce STORYSCAPE. The system uses a set of story elements to construct spatial relations between actors, events and places and produce a graphical representation of the story. It also discovers spatial relations between story elements and makes inferences about the places where the events in the story take place.

1 Vision

Space and spatial relationships play an important role in how we, as humans, perceive the world. We facilitate our perceptual systems to infer implicit components that take part in stories such as, containment, proximity and directionality. Let us use the statement *"I went to meet professor X."* as an illustrative example. When hearing this statement, we build certain assumptions. It happened on planet Earth. It probably happened in the city where the university X works for is located. It probably happened in professor X's office, etc. From a story understanding perspective, spatial relationships play a role in how we perceive the interaction between characters. The sentence *"X met Y near a train station."* in a story creates a mental image of the scene. Furthermore, in the case of specific authors, such as Herman Melville, space and places play a crucial role in understanding the narrative and overall plot.

Because stories (narrative or everyday life events) and perception are closely connected, in order to develop intelligent systems, we need to bridge story understanding with perception. We introduce the STORYSCAPE system that uses a set of story elements to construct spatial relations between actors, events and places and produce a graphical representation of the story. It enables discovering the spatial relations between story elements and makes inferences about the places where the events in the story take place. The representation of the system consists of three functional elements: PLACE, ACTION and RELATION. PLACE is the main element in which every actor is contained. This means even if the story does not explicitly describe any place or location, every action is represented inside a PLACE container. ACTION allows the system to discover transitions between places and time. The system uses an action dictionary to find whether an action causes a transition between places or not. For example, actions such as, *"to go", "to enter"* cause a transition. Finally RELATION represents places in relation to each other using containment, proximity and directionality descriptors in the story. For example, a room in a house is represented with two places in which one place is contained within the other (house contains room).

STORYSCAPE uses a set of spatial rules that allows us to make inferences. DIRECT INFERENCE rules determine where a particular actor is, and update the places that are affected by any change. For example, *"John goes to school."* triggers a direct inference rule:

> INFERENCE RULE: "if xx goes to yy, xx is in zz, xx exits zz, xx enters yy"
> RESULT: John is in Place 1. John exits Place 1. John enters school.

As demonstrated by the example, a generic place, Place 1, is generated to represent John's current location. Then John is removed from Place 1 and added to School. This is explicitly described in story.

PROBABILISTIC ACTION-PLACE rules allow us to determine what actions are more likely to be associated with a particular place. For example, the act of eating is likely to be associated with a restaurant place with high probability, while the act of reading is not. These rules are also used to determine a transition between places. For example, if John was planning to go to the museum and eat dinner afterwards, the system will use probabilistic action-place rules to determine that the act of eating has a low probability of happening at a museum. Therefore, John must have transitioned from the museum to a new place to eat dinner.

The system has two representations: CHRONOLOGICAL and VISUAL. CHRONOLOGICAL representations allow us to represent the same place in the different time states it takes during the story. For example, if John exits a room and enters that room again, there are three states generated for the room, accounting for when John is contained in the room and for when he is not. This allows the creation of causal bridges between different places through finding who is where at a given moment. VISUAL representations represent a moment in the story by taking into consideration the spatial relations between the places and trajectories of actors. The system is able to discover spatial transition patterns that are analogical to Concept Patterns in the Genesis system.

Our system contributes to the CBMM (Center for Brains, Minds and Machines) challenge. The central CBMM challenge problem is to understand pictures and videos computationally, developmentally, neurobiologically, and socially. The spatial understanding of stories plays a role, because pictures and videos tell stories. Our system also enriches the capabilities of the Genesis story understanding system.

2 Steps

2.1 Background Work

Before we introduce our system, we will first refer to some background work.

2.1.1 The Natural Language Question Answering System: START

To get spatial perception from stories, we need a way to analyze the structure and semantics of a sentence. We also need a way to look for and extract a particular word from a text or sentence. Parsing techniques enable us with such capabilities. The parsing software we use in our system is the START parser.

START (SynTactic Analysis using Reversible Transformations) was developed by Boris Katz at MIT's Artificial Intelligence Laboratory. Currently, the system is undergoing further development by the InfoLab Group, led by Boris Katz. The software was designed to answer questions that are posed to it in natural language. START parses incoming questions, matches the queries created from the parse trees against its knowledge base and presents the appropriate information segments to the user. In this way, START provides untrained users with speedy access to knowledge that in many cases would take an expert some time to find.

A key technique called "natural language annotation" helps START connect information seekers to information sources. This technique employs natural language sentences and phrases – annotations – as descriptions of content that are associated with information segments at various granularities. An information segment is retrieved when its annotation matches an input question. This method allows START to handle all variety of media, including text, diagrams, images, video and audio clips, data sets, Web pages, and others.

The natural language processing component of START consists of two modules that share the same grammar. The understanding module analyzes English text and produces a knowledge base that encodes information found in the text. Given an appropriate segment of the knowledge base, the generating module produces English sentences. Used in conjunction with the technique of natural language annotation, these modules put the power of sentence-level natural language processing to use in the service of multimedia information access.

Besides its capabilities, another reason our system uses START is because START is the parser incorporated in the Genesis system, which will be introduced in the next section. Further information about the START system can be found on its website [1].

2.1.2 The Genesis System

The Genesis Group at MIT, lead by Patrick H. Winston, supports the idea that human intelligence differs from that of other primates, past and present, because humans are able to understand, tell, and compose stories, and because they are able to direct their perceptual faculties to describe, and answer questions about, real and imagined situations. It follows that, in order to develop systems of human-level intelligence, there is a need for developing computational models of story understanding and directed perception. To understand how humans understand and learn from stories, we must first specify the behavior we want to model, then develop a suite of constraint exposing representations, then build a system to test the ideas. For this purpose, the Genesis story understanding system was created.

Genesis analyzes stories ranging from Shakespeare's plots to descriptions of conflicts in cyberspace. Genesis works with short story summaries, provided in English, together with low-level common-sense rules and higher-level concept patterns, likewise expressed in English. Using only a small collection of common-sense rules and concept patterns, Genesis demonstrates several story understanding capabilities. For example, Genesis determines that both *"Macbeth"* and the *"2007 Russia-Estonia Cyberwar"* involve *"revenge"*, even though neither the word *"revenge"* nor any of its synonyms are mentioned in the story summaries for both stories.

More information about the Genesis system and its capabilities can be found on its homepage [2].

Figure 1: Genesis determines that both *Macbeth* and the *2007 Russia-Estonia Cyberwar* involve revenge.

2.1.3 Causal Reconstruction

Another related work is Gary Borchardt's PATHFINDER [3]. In order to make a causal reconstruction an appropriate representation of the given description is required. Even though representing the statements related to causalities seems appropriate, there are cases that can not be captured by state representation, such as, temperature or elevation which can be differentiated with respect to time. Therefore, Borchardt proposes the Transition Space which represents causality as an association between changes of relevant attributes of objects through different time points. After that, events as short paths in the space can be associated to sequences of causalities or they can be combined for larger causal explanations. The transition space is composed of five main components: objects, attributes, time points, reference standards and predicates. The changes are compared using ten change characterization predicates consisting of boolean attributes (APPEAR, DISAPPEAR), qualitative attributes (CHANGE), and quantitative attributes (INCREASE, DECREASE), together with their negations (e.g. NOT-DISAPPEAR). In addition to these predicates, using EQUAL and GREATER with their negative pairs simplifies matching and inference processes. The comparison at or between two time points defines the event traces. Following the definition of event traces, partial matching is performed between them to construct an association structure. In addition to this, inference and explanatory transformations are applied to the structure, both for checking consistency and expanding it with further paths. In our work, we propose a similar approach to represent time states of places, triggered by actions. Using similar state descriptions such as enter/exit, change/not change, we are able to represent the states of places through the course of the story.

2.2 Elements of The system

The proposed system parses stories into places that are connected through transitional statements. We build the spatial representation using three main components: places, actions and relations.

2.2.1 Places

The main component of the system is the place representation. The idea behind it, is that all components of the story are contained in a place, whether it is explicitly stated or not. For example, the sentence *"John kissed Mary."*, includes a place that contains John and Mary, even if it is not stated in sentence. Therefore the system uses places to represent spatial containment of the story elements. This simple structure allows the system to represent the spatial relations and make inferences about story elements.

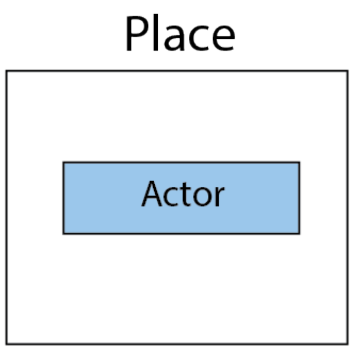

Figure 2: Place represents a container for story elements.

2.2.2 Actions

The second component of the system is the action representation. Actions plays two important roles in the system. First of all, every action causes a transition in the time state of the place, allowing us to determine the chronological order of the events that happened in the place. For example, the sentence *"John sat on the chair."*, causes the system to represent two states of the place: before and after John sits on the chair. Through the course of the story, the system uses the most recent state of the place for making connections to future events, while using past states to make inferences about parallel events in the same place or different places.

The second use of the action representation is to understand transitions between places. Two groups of actions are defined in the system's Action Dictionary as transitional and in-place actions. Transitional actions (go, enter, come, fly, walk to, etc.) cause the system to change the actor's place from the existing place to another one. If the story explicitly states the terminal place, the actor is placed in the mentioned place. Otherwise, the system probabilistically determines a category for the place based on the actions that occur in the same place. For example, the sentence *"John and Mary went to dinner."* causes the system to create a new place and look up in the Action-Place Dictionary to find out what is the most likely category of place where people have dinner (e.g. a restaurant). In-place actions do not cause any transition between places. However, in future iterations of the system, they will be used to determine spatial relations between objects and actors. For example, the sentence *"John put his phone on the table."* would cause to system to generate two states: one with John holding the phone and one with the phone on the table. The elements of the Action Dictionary are represented using actions as keys, and a set of tuples (type of action group, direction) as values: *action: (type, direction)*. Two possible examples would be *"go : [(transition, to) ...]"* and *"eat: [(in-place, none) ...]"*.

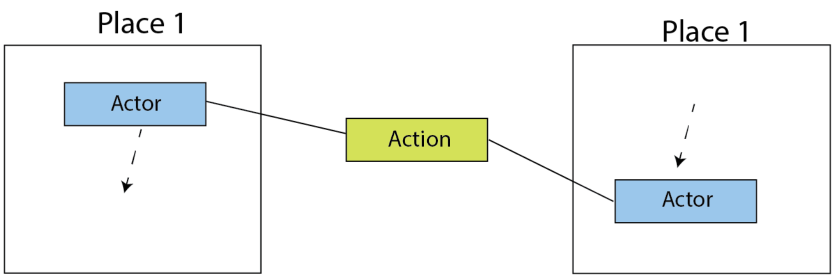

Figure 3: In-Place Actions generate Time States within the same Place.

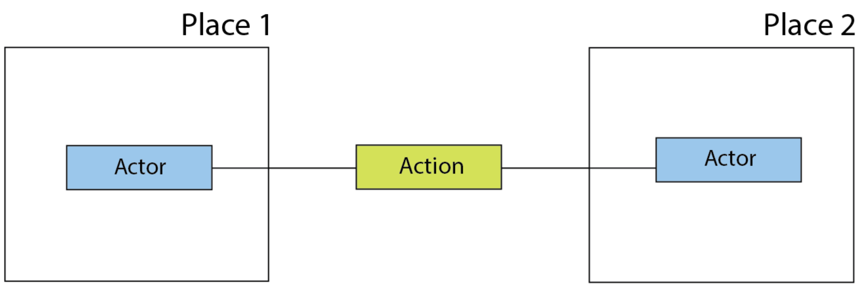

Figure 4: Transitional Actions trigger a transition between places.

2.2.3 Relations

RELATION represents places in relation to each other by using containment, proximity and directionality descriptors in the story. It forms the basis of the visual representation of the story, providing the information about how places spatially relate to each other. For example, if place A is next to place B, they are represented as two places next to each other in system. Containments describe whether a particular place is contained in another place, as in the example of house and a room. The container place directly inherits all spatial properties of the contained place, and contains its actors. For example, if John sleeps in the bedroom, he also sleeps in the house. Proximity describes how two places are related to each other in terms of distances. Places can be next to each other or far apart. Directionality provides information about where places are located in relation to each other, provided by descriptors such as north, west, top, in front, towards etc. For a detailed explanation, see the visual representation in Figure 5.

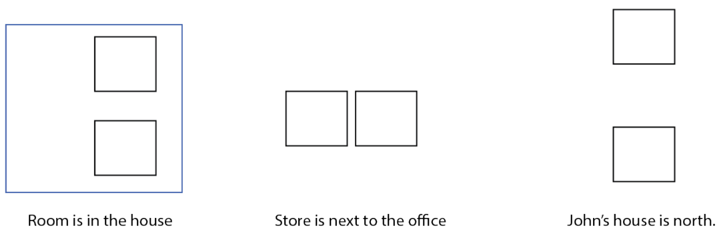

Figure 5: Relations dene how dierent places are related to each other in a visual context

2.2.4 Inference Rules in System

2.2.4.1 Direct Inference/ Replacement

STORYSCAPE makes inferences based on a set of inference rules. REPLACEMENT rule assures that every transition causes an actor to be removed from the start place - triggering a DISAPPEAR- and appear in the terminal place triggering APPEAR. Here, APPEAR/DISAPPEAR, similarly to Borchardt's work, allows the evaluation of a character's state in a given moment.

DIRECT INFERENCE generates spatial information based on transitional actions. For example, the sentence *"John enters the room."* includes the transitional action *"to enter"*, which requires a starting and a terminal place. The system infers that John was in another place, even if that was not stated in story, and generates a representation for the inferred place.

A slightly related rule is BACKWARD PROPAGATION, which generates the previous state of a place given a particular action. If the action is transitional and triggers an APPEAR rule, BACKWARD PROPAGATION creates the previous state of the place, applying a DISAPPEAR rule towards back in time. (Figure 7)

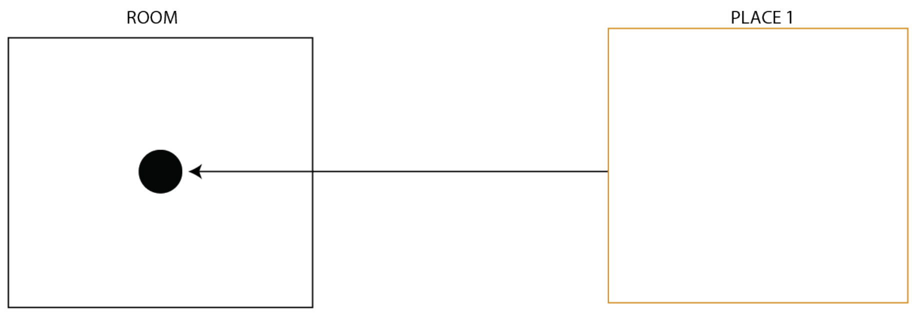

Figure 6: *"John enters the room"* triggers a DIRECT INFERENCE rule and causes the system to generate an inferred place

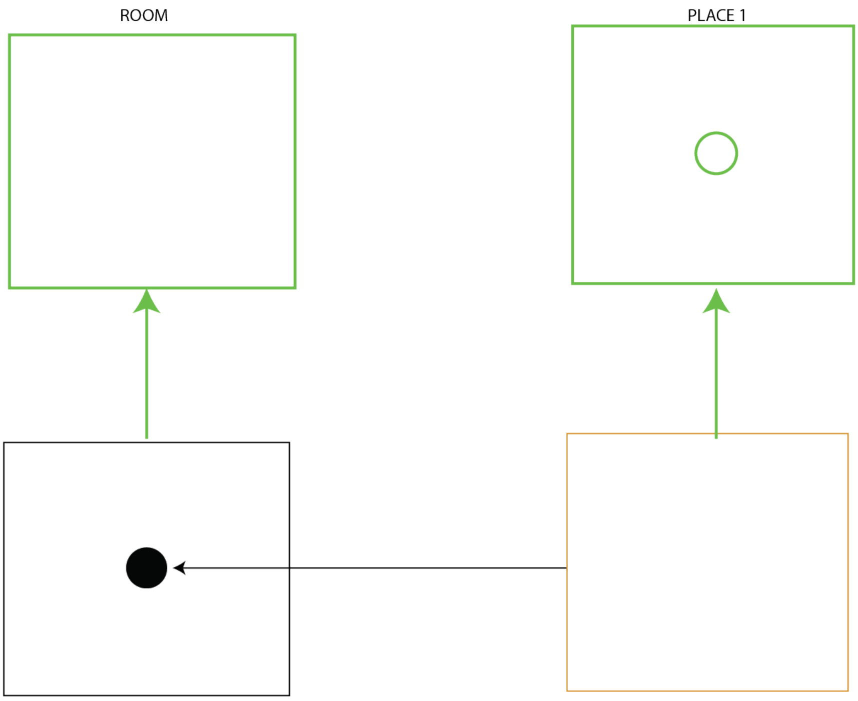

Figure 7: *"John enters the room"*, causes system to generate the previous state of the room, removing John through a BACKWARD PROPAGATION back in time (green), and adding John to the other place through an appear rule.

2.2.4.2 Chronological Action Space Representation

Using the inference and replacement rules described above, the system builds up a Chronological Action-Place representation, that represents the time states of different places, that are inferred from the actions. It also represents how a particular actor moves between places, creating trajectories that eventually allow us to discover patterns of transition. Time representation also allows us to make inferences about the concurrence of events that take place in the story, starting from the places where multiple actors meet, and back-propagating to the actions that each actor took until that point. (Figure 8)

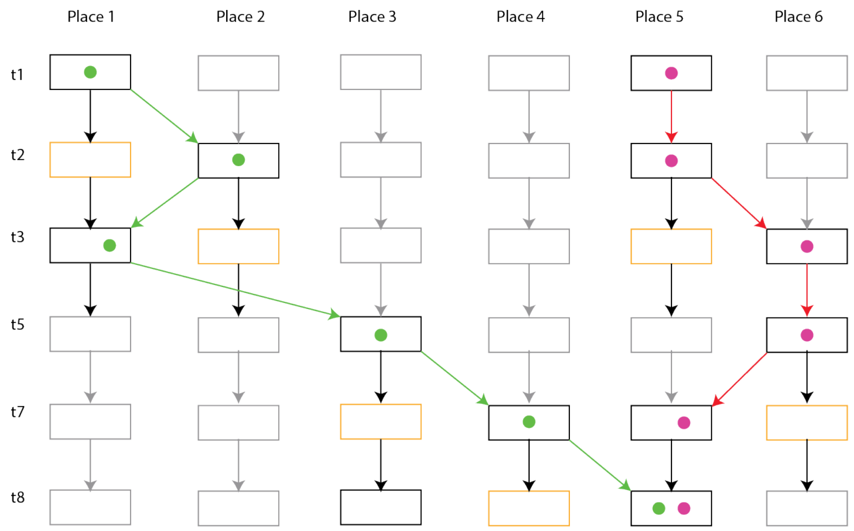

Figure 8: Chronological Action-Place Representation allows discovering concurrence of events in story in dierent places.

2.2.4.3 Visual Place Representation

Finally, the system is also able to generate visual representation of places in relation to each other. This representation enriches the chronological one, allowing the RELATION to take part. Places contained in each other and places that have certain proximity are included in this part.

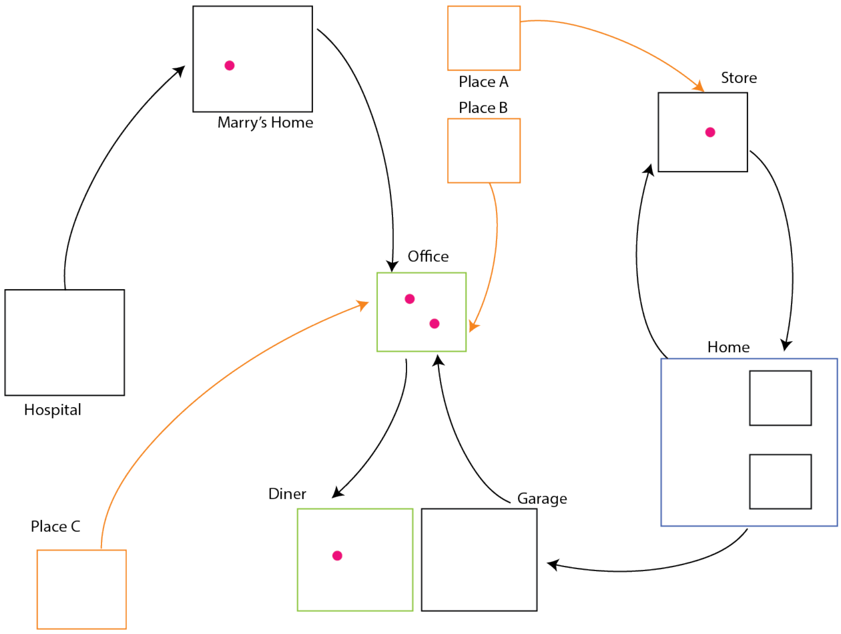

Figure 9: Visual Place Representation with Direct and Probabilistic Inferences.

2.2.4.4 Probabilistic Action-Place Dictionary

The Action-Place Dictionary is used to determine what actions are more likely to occur at a particular place. The dictionary is probabilistic because each action is associated with a certain probability. The elements of this dictionary are represented using places as keys, and a set of tuples (action, probability) as values: *place: (action, probability)*. A possible example would be *"restaurant: [(eat, 0.99), (read, 0.05) . . .]"*.

The Action-Place Dictionary is also used to determine a transition between places. For example, given the sentence *"John planned to go to the museum and eat dinner afterwards."*, the system will determine that the act of eating has a low probability of happening at a museum (probability of 0.05). The system will then conclude that John must have moved to a different place to have dinner.

3 Example

Here we provide a demonstration of how the system works, using the following story as an example:

"John goes to shop for Christmas presents. He buys snacks for his nephew. Mary enters the store. John exits the store. Mary discovers that she left her wallet at home. Mary exits the store. Mary sees John. John and Mary drive to Mary's home. Mary finds her wallet in the kitchen. John goes back to his home."

Direct Inference Rule: if xx goes yy, xx is in zz. xx disappears zz. xx appears yy.
Sentence: John goes to shop.
Result: John is in Place 1. John disappears in Place 1. John appears in Place 2.

Probabilistic Inference Rule: store: [(shop, 0.99) ...]
Result: Place 2 store

Direct Inference Rule: if xx enters yy, xx is in zz. xx disappears zz. xx appears yy.
Sentence: Mary enters store.
Result: Mary is in Place 3. Mary disappears in Place 3. Mary appears in store.

Direct Inference Rule: if xx exits yy. xx is in yy. xx disappears yy. xx appears zz.
Sentence: John exits store.
Result: John disappears in store. John appears in place 4.

Direct Inference Rule: if xx sees yy and if xx is in zz. yy is in zz.
Sentence: Mary sees John. John is in Place 4. Result: Mary is in place 4.

Direct Inference Rule: iff xx drives to yy. xx enters yy.
Sentence: Mary and John drive to Mary's home.
Result: Mary is in Place 4. John is in Place 4. John disappear Place 4. Marry disappear Place 4. John appear Mary's home. Mary appear Mary's home.

Proximity Rule: if xx drives yy. xx is in zz. zz is DISTANT yy.

Containment Rule: kitchen is contained in a house.
Sentence: Mary finds her wallet in the kitchen.
Result: Mary is in Mary's home. Kitchen is in Mary's home.

The images below represent the Visual and Chronological respresentations of the story. Blue elements are created using Probabilistic Inferences (store). Orange elements are created using Direct Inferences.

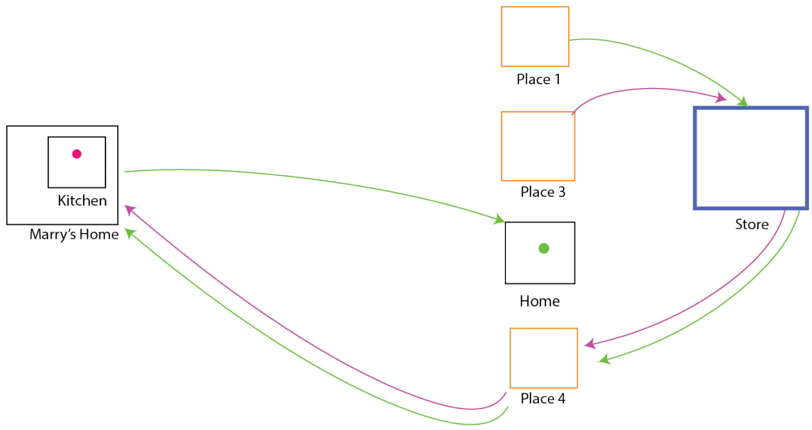

Figure 10: System generates the Visual representation of the story.

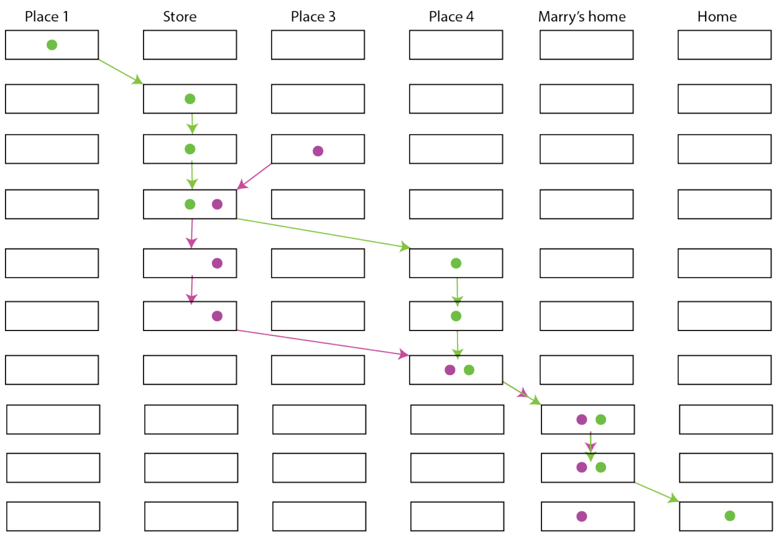

Figure 11: System generates the Chronological representation of the story.

4 Contributions

Below we list our contributions:

1.This study is a step towards bridging the gap between story understanding and the perceptual system.

2.We have proposed a spatial reconstruction system, STORYSCAPE, that can represent and evaluate spatial characteristics of stories, and make better inferences about stories by taking into account containment, proximity and directionality between story elements.

3.We developed two representations for spatial reconstruction: chronological and visual. Chronological representations provide temporal information about which actor is contained in what place in a given time in story, and make inferences based on co-occurrences of actors. Visual representations provide a graphical map-like representation of places and trajectories of actors through the course of the story and allow for detection of transition patterns for the actors. For example, the system can tell whether an actor commutes between two places (e.g. is actor going and coming back to the same place).

4.We provided a probabilistic dictionary to infer what categories of place are more likely to be associated with a particular action in the story. This eventually allows the system to generate new places when necessary.

5.We have provided examples of how the system works.

6.This project contributes to the CBMM challenge by proposing a method for reconstructing the story in a visual context, and using perceptual information about containment, proximity and directionality to make inferences about the story elements. In future iterations the same methodology can be applied to reconstruct the events in the same place and make inferences about the interactions between actors and objects.

7.The system also enriches the capabilities of the Genesis story understanding system.

References

[1] START System Website: http://start.csail.mit.edu/index.php

[2] Genesis Home Page: http://groups.csail.mit.edu/genesis/

[3] Gary C. Borchardt, *Causual Reconstruction:* http://start.csail.mit.edu/publications/Borchardt-AIM1403.pdf