Explorations in Cyber International Relations

Massachusetts Institute of Technology Harvard University

Story Retrieval and Comparison Using Concept Patterns

Caryn E. Krakauer

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

Patrick Henry Winston

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

May 26, 2012

This material is based on work supported by the U.S. Office of Naval Research, Grant No. N00014-09-1-0597. Any opinions, findings, conclusions or recommendations therein are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.

Story Retrieval and Comparison using Concept Patterns

Caryn E. Krakauer, Patrick H. Winston

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, Massachusetts
krakauer@mit.edu, phw@mit.edu

Abstract

Traditional story comparison uses key words to determine similarity. However, the use of key words misses much of what makes two stories alike. The method we have developed use high level *concept patterns*, which are comprised of multiple events, and compares them across stories. Comparison based on concept patterns can note that two stories are similar because both contain, for example, *revenge* and *betrayal* concept patterns, even though the words *revenge* and *betrayal* do not appear in either story, and one may be about kings and kingdoms while the other is about presidents and countries. Using a small corpus of 15 conflict stories, we have shown that similarity measurement using concept patterns does, in fact, differ substantially from similarity measurement using key words. The Goldilocks principle states that features should be of intermediate size; they should be not too big, and they should not too small. Our work can be viewed as adhering to the Goldilocks principle because concept patterns are features of intermediate size, hence not so large as an entire story, because no story will be exactly like another story, and not so small as individual words, because individual words tend to be common in all stories taken from the same domain. While our goal is to develop a human competence model, we note application potential in retrieval, prediction, explanation, and grouping.

Keywords: Goldilocks principle, story retrieval, intermediate features, concept patterns

1. Story comparison and precedent retrieval

Any full account of precedent-based reasoning must provide an account of how potentially relevant precedents are retrieved from memory. Considerable psychophysical research, reviewed in Finlayson and Winston (Finlayson and Winston, 2005), indicates that novices in a domain retrieve using superficial features, whereas experts in a domain retrieve retrieve using structure.

Finlayson and Winston showed how a range of behavior, from novice to expert, corresponds to an increase in the maximum chunk size considered by a matcher, starting with individual objects and ending with collections of objects and the relations among them. Thus, expert behavior corresponds to matching not on objects, nor on entire precedents, but on chunks of intermediate size, which led Finlayson and Winston to frame what they call the Goldilocks principle.

The Finlayson and Winston work was based on structure mapping theory (Falkenhainer et al., 1989; Gentner and Forbus, 1991), and thus requires computationally expensive graph matching. Our work, in contrast, is based on what we call *concept patterns*, which are reminiscent of plot units (Lehnert, 1981), and capture aspects of what we mean when we talk of, for example, *revenge* or *selling out*. As in Dehghani's work on analogy and moral decision making, retrieval is sensitive to known narratives (Dehghani et al., 2009).

In our implemented system, potential precedents are stored along with the concept patterns determined to lie within them. Then, at retrieval time, story-to-precedent matching is done by a fast dot-product computation on a concept-pattern vector derived from a story with concept-pattern vectors derived from potential precedents.

Thus, relative to the Finlayson-Winston work, our approach is fast and our concept patterns may span long chains of connected relations, while still retaining the flavor of retrieval based on intermediate features. [1]

Of course, once a potential precedent is retrieved, analysis begins, and judgments of similarity involve not just events and how they are arranged but also concept patterns and how they are arranged. Accordingly, we have begun to study the role of concept patterns in similarity judgments in general, not just in retrieval.

2. The Genesis substrate

We build on the Genesis System (Winston, 2011), a story understanding system that reads simple English, elaborates on what it reads by applying commonsense rules, and performs searches to detect concept patterns. At the commonsense level, Genesis notes, for example, that if you are killed, you become dead, and if I harm you, I harm your friends.

Concept patterns are higher level structures in which events are said to lead to other events, with possibly many intermediate events. We generally supply concept patterns in English, instructing Genesis directly, as in the following examples:

[1]Finalyson and Winston's thinking about intermediate features was originally inspired by Shimon Ullman's work on face finding in images (Ullman et al., 2002). He matched, for example, using features such as a nose and mouth combination, rather than, say, an eye or an entire face.

```
Start description of "Revenge".
xx is an entity.
yy is a entity.
xx is my friend.
xx's harming yy leads to yy's harming xx.
The end.

Start description of "Pyrrhic victory".
xx is an entity.
ll is an action.
zz is an entity.
xx's wanting ll leads to xx's becoming happy.
xx's wanting ll leads to zz's harming xx.
The end.
```

Concept patterns are chosen by the user, and are generally include two to three events. Users choose concept patterns based on the users' understanding of the stories. We are developing a method of automatically generating concept patterns, which is discussed later in the paper.

As stories are read by Genesis, commonsense rules are deployed, which have a tendency to connect the story's explicit events via causal relations. The search machinery that looks for satisfied *leads to* relations then exploits those causal connections to locate concept patterns. In our simple, abbreviated rendering of the plot in Shakespeare's Hamlet, that search machinery finds *Pyrrhic Victory*, *Leadership Achieved*, *Suicide*, and three instances of *Revenge*

3. Comparing stories using concept patterns

We compared stories on a concept level using three different methods, each of which serves a different purpose in story comparison. The methods are: comparing the number of concept patterns in common, noting the longest common substring of concept patterns, and weighted comparing usin concept pattern rarity.

Vector-angle Mode

Our first method compares the number of concept patterns in common for fast retrieval. We save story concept pattern counts in vectors. Then, using these vectors, our method calculates the angle between story vectors to determine similarity, with the metric varying between 0.0 and 1.0.

For example, the highest match for the Bay of Pigs Invasion by vector-angle is the Cambodia-Vietnam Invasion. Both conflicts have a *allied offense*, an *invasion*, and a *victory*. In both conflicts, a larger political entity supported a smaller group's invasion.

Order-sensitive Mode

Our second method takes into account the ordering of the concept patterns. For example, a *revenge* that is the result of a *betrayal* is different than a *betrayal* that is the result of a *revenge*. The importance of ordering can also be seen in the comparison of the American Revolution and Afghan Civil War, as shown in figure 1.

In the stories as provided, both the American Revolution and the Afghan Civil War contain *defense* and *allied support*, but they appear in different orders. In the American Revolution, the American people received allied support from France after Britain's attack. In the Afghan Civil War however, Russia gave allied support to Najibullah before the attack happened. In one, an ally came to the support of an already embattled nation. In the other, an ally helped stockpile weapons for an impending conflict. The ordering of two stories makes a difference in their overall similarity.

Rarity-sensitive Mode

The rarity of each concept pattern is also important in comparing stories. We recognize three variations:

- **Rare among a group of stories:** If a concept pattern is rare among a group of stories, it can be seen as more important when comparing similar stories. For example, when looking at a group of Disney-style fairy tales, two stories that have a *princess marries a prince* concept pattern, they do not seem as similar as two stories in which a concept pattern indicating *princess ditches the prince and marries a poor commoner*, because the ditch-the-prince concept pattern is rare.
- **Very common among a group of stories:** If a concept pattern is very common among some but not all stories, it may be useful for grouping stories. If a group of stories have concept patterns in common, but those concept pattern are much rarer among all stories, than that group of stories may make up a genre. For example, the "Disney-style fairy tale" genre may have concept patterns such as *princess and prince fall in love*, *villain causes prince and princess to be kept apart*, and *prince and princess live happily ever after*. If a new story is read with similar concept patterns, it may also be a Disney-style fairy tale.
- **Very common among most stories:** If a concept pattern is very common among most stories, then it is not particularly useful in deciding whether two stories are similar.

As an example of the influence of rarity, consider the rarities of the concept patterns in the American Revolution, with the concept rarities shown in table 1.

Concept Pattern	Rarity
Legal Disagreement	0.027
Invasion	0.167
Rebellion	0.069
Unwanted succession	0.042
Conflict	0.083
Allied Defense	0.014
Victory	0.208
Victory Defensive	0.069

Table 1: The rarity of concept patterns found in the American Revolution story. An example of the *victory*' pattern is the most common concept pattern, while an *allied defense*' is the most rare. Rarity is calculated by dividing the number of times a concept pattern appears by the total number of concept patterns seen.

The most common concept pattern is *victory*. A victory occurs in almost every story in the set, and so is very common. Because of this, a *victory* is a poor measure of similarity between these stories, but a very good indicator that the story

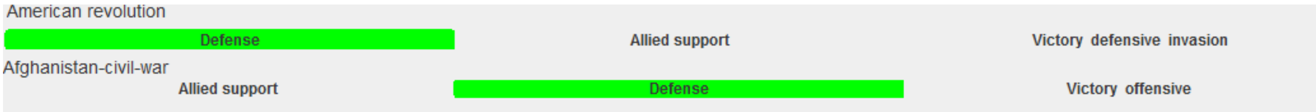

Figure 1: The in-order comparison of the American Revolution and the Afghan Civil War. While these stories both have defense and allied support, the two concept patterns appear in differing order. If order was not taken into account, then the two stories would share two in-common concept patterns. However, with order taken into account, the maximum sub-plot has only a length of one. The maximum sub-plot is *defense* which is highlighted in the figure.

is about a conflict. On the other hand, an *allied defense* is much more rare and therefore more important when measuring story similarity in the conflict domain.

We currently calculate rarity by dividing the number of times a concept pattern appears by the total number concept patterns seen.

4. Experimental results

We have run our system, in Vector-angle Mode, on 15 conflict summaries previously used in the work of Finlayson and Winston (Finlayson and Winston, 2005). These includes rebellions, wars, and political conflicts. Here, for example, is the American Civil War as provided to our system:

> *Start story titled "American Civil War". The United States is a country. The Confederacy is an entity. The Union is an entity. The Confederacy was a region of the United States. The Union was a region of the United States. The Union disliked the Confederacy because the Confederacy possessed slaves. The Confederacy left the United States because the Confederacy disliked the Union. The Confederacy left the United States because the Union possessed the Confederacy. The Union wanted the Confederacy to stay at the United States. The Union attacked the Confederacy because the Confederacy left the United States and the Union wanted the Confederacy to stay at the United States. The Confederacy attacked the Union because the Union attacked the Confederacy. The Union was stronger than the Confederacy. The Union defeated the Confederacy because the Union attacked the Confederacy and the Union was stronger than the Confederacy. The Union controlled the Confederacy because the Union defeated the Confederacy. The Union forced the Confederacy to return to the United States because the Union controlled the Confederacy and the Union wanted the Confederacy to stay at the United States. The end.*

Stories have been simplified mainly to get them through the front-end natural-language parser. Accordingly, the need for simplification will diminish as natural-language parsers improve.

Story simplification introduces the possibility of simplifier bias. From one simplifier's perspective, an attack might be recast as an *invasion* while from another, it might be described as a *counter-attack*.

If two interpretations are different enough, there may be a change in the analysis, but we view this as a feature, not a bug. If a simplifier thinks of two wars very differently, say one war was a *justified first strike* and another as an *unjustified invasion*, they would not be considered similar by the simplifier and likewise would not be considered similar by our system .

Figure 2 illustrates the differing results using concept-pattern vectors (top) and word vectors (bottom). Black represents a similarity score of zero and white represents a similarity score of 1.0, the maximum possible value.

When comparing the conflict stories on a word level, the difference between the similarity scores of most story pairs is small. Because all of the stories are on the same topic, they all share many keywords. Stories compared with themselves are white because the keywords are exactly the same, but when compared to other stories, the comparison scores are relatively low and do not change much from story to story. The mean and standard deviation for each method are as follows:

Method	Mean	Standard Deviation
Keyword	0.267	0.119
Concept Pattern	0.364	0.200

Table 2: The mean and standard deviation of similarity scores generated by each method. The standard deviation of story comparison by concept pattern is almost twice that of keyword comparison. Similarity scores are on a scale from 0.0 (not similar) to 1.0 (identical).

The Cambodian-Vietnam Invasion compared with the China War with Vietnam is an outlier. These two stories are two parts of an overall conflict, so the actors in both conflicts are the same.

We found that comparing stories using concept patterns performs in more congruence with our own interpretations. For example, the deviation of similarity score values is much higher than in keyword comparison on the fifteen conflict stories on which we ran experiments, just as we view story pairs as varying considerably in similarity. Following are three examples where concept pattern comparison finds similar stories but keyword comparison falls short.

- **American Revolution and the American Civil War:** Concept pattern comparison picks out the American Revolution and the American Civil War as being similar giving them a similarity score of 0.67, as they have several concept patterns in common (*unwanted succession*, *victory*, *conflict*, *legal disagreement*). This makes

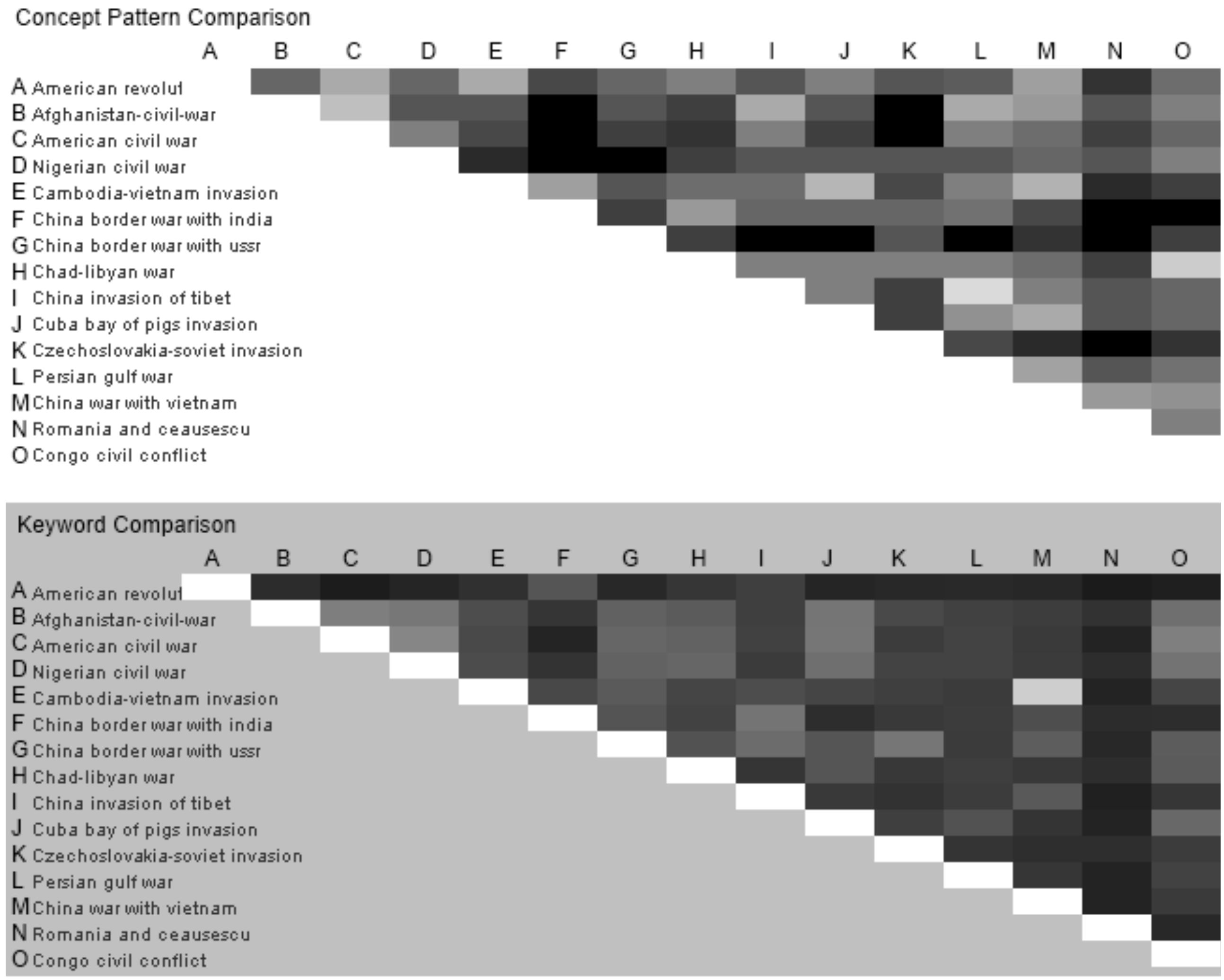

Figure 2: Top: similarity scores produced by concept patterns. Bottom: similarity scores from keywords. Comparison based on concept patterns has a greater diversity of scores and sensitivity to high-level structural matches even in the absence of low-level keyword correspondence. Similarity scores are created using the vector angle calculation.

sense, as both stories are about a part of a country rebelling from the main country over legal disputes (taxes in one case, slaves in the other). In the word comparison, these stories have a very low similarity score of 0.1 (as shown by the red). By using concept patterns to compare stories, more meaningful story comparison is performed.

- **China border War with India and the Cambodia-Vietnam Invasion:** Another example of the concept pattern comparison succeeding while the keyword comparison fails is the comparison between the China border War with India and the Cambodia-Vietnam Invasion. In both cases, two countries fought over an area of land (the Mekong Delta in the Cambodia-Vietnam conflict, and the Assam in the China-India conflict). The relevant concept patterns found are a land dispute along with two invasions (one by each country into the disputed region), which gives the comparison score of 0.71. The keyword comparison however, rates them as relatively unsimilar with a score of 0.26.

- **Afghanistan Civil War and the Czechoslovakia Soviet Invasion:** An example where keyword comparison has decided that two stories are similar, where in fact they are not, are the Afghanistan Civil War and the Czechoslovakia Soviet Invasion. Keyword comparison gives a score of 0.48, which is very high for keyword comparisons. However, the concept pattern comparisons give them a score of 0.0. The stories, while both involve the Soviet Union, are very different conflicts. In the Czechoslovakia Soviet Invasion, the Soviet Union invaded Czechoslovakia due to political reform. In the Afghanistan Civil War, the Soviet Union funded one side of a civil war, but did not actually attack. Thus, the two conflicts are quite different, which is shown by the concept pattern comparison.

5. Concept Pattern Generation

Our current work includes concept-pattern discovery directly from stories, circumnavigating the need to supply all concept patterns in English. It is in the same spirit as the concept-pattern discovery work of Mark Finlayson (Finlayson, 2012).

Our discovery process works by searching for concept patterns, consisting of two or three events in leads-to relations, common to two or more stories. In principle, there could be $O(n^3)$ such concept patterns in a story, where n is the number of events; in practice, there are far fewer, because only event pairs connected by causal chains qualify as potential concept patterns.[2] In addition, we filter out concept patterns that are too rare. We ignore concept patterns that only appear in a single story. This is reminiscent the approach taken by Chambers and Jurafsky in their work on unsupervised learning. (Chambers and Jurafsky, 2008) Due to their large amount of data, their system was preforming poorly. Accordingly, they eliminated rare occurrences of verb pairs, improving performance.

Once concept patterns in one story are computed, they can be compared to concept patterns appearing in one or more previously read stories In order for two concept patterns to be the same, they must align. In order to align, their structure must be the same, their events must be similar, and the concept pattern's actors must align. In order for two events to be similar, their actions must be similar. For example, *A invades B* is similar to *C attacks D*. Every word has a thread which is defined by WordNet. The thread for *invade* is {action, contend, attack, invade} while the thread for *attack* is {action, contend, attack}. Two words a similar if they are the same word, share a common parent, or if one of the words is a parent of the other. A "parent" is defined as the immediate parent to the word (so attack for invade, and contend for attack).

- Same Structure: Two concept patterns have the same structure if their leads to relations are the same. So *a→b→c* has the same structure as *d→e→f*, but not *d→e d→f*.
- Similar Events: Concept patterns can have different event types, as long as they are sufficiently similar. For example *punch*, *kill*, *insult*, and *murder* are considered sufficiently similar because they are all kinds of *harm*. Our system uses WordNet to determine if two words are similar in meaning.
- Aligned Actors: Two concept patterns can be aligned if the actors from each event correspond. So if the two concept patterns are: *a harms b leads to b harms a* and *c harms d leads to d harms c*, then the events align but *a harms b leads to b harms a* and *c harms d leads to e harms c* do not.

Below are examples of concept patterns generated by the system. The names were provided by us after the fact and are not known the the system.

- **Giving Aid (two events):**
  American revolution:
    France helps America
    France gives money to America
  Cambodia-vietnam invasion:
    China helps Cambodia
    China gives weapons to Cambodia

[2]This idea emerge in a discussion that included Finlayson, fortuitously initiated by a fire drill in our building.

- **Revenge Attack (two events):**
  Afghanistan-civil-war:
    Najibulla attacks Mujahideen
    Mujahideen attacks Najibulla
  American civil war:
    Confederacy attacks Union
    Union attacks Confederacy
- **Wanting an entity to stay, and dislike between entities, leads to a defeat (three events):**
  Nigerian civil war:
    Nigeria wants NigerianEast to stay
    Nigeria defeats NigerianEast
    NigerianEast dislikes Nigeria
  American civil war:
    Union wants Confederacy to stay
    Union defeats Confederacy
    Confederacy dislikes Union
- **Wanting an invasion leads to an invasion, which is defeated (three events):**
  Cuba bay of pigs invasion:
    UnitedStates wants exiles to invade Cuba
    Exiles invade Cuba
    Soldiers defeat Exiles
  China war with Vietnam:
    Vietnam does not want China to invade Vietnam
    China invades VietNam
    Vietnam defeats China

Because there are many potential concept patterns in stories, care must be taken to only select the concept patterns that are meaningful when measuring story similarity. A concept pattern that only appears in just two stories is not likely to be important, as it can serve no role in demonstrating story similarity more generally. Likewise, a concept pattern that appears in all stories is not useful because it has no discriminatory power.

In our next step, we will attempt to use a mutual information metric to establish which of the candidate concept patterns are useful in story comparison.

6. Potential application

Our main goal in this work is to model human story retrieval, and in that connection, we are planning a series of psychological experiments. In passing, we note that our approach to similarity matching offers a promising approach to prediction, understanding, and grouping.

- *Retrieval and prediction:* By finding patterns in similar stories, the ending of a new story can be predicted by way of precedent. This is especially useful for understanding how a person from a culture different from our own will respond to a proposed course of action. If our system is loaded with stories that characterize a culture of interest, and is then presented with the beginning half of a course of action, its predictions may well be different from those predicted in the absence of those culture-characterizing stories. Suppose, for example, a person is presented with a story: "Charlie and Bob were friends.

Charlie hit Bob in the face." If the person is then asked to predict the ending, his answer will depend on his culture and upbringing. By understanding people's reactions to situations, we can better predict the outcome of events.

- *Retrieval and explanation:* By finding a similar story, one better understood that a current story, explanations for events can be discovered. Consider, for example, the scenario: "Bob bought Jill flowers." Without any explanation for the action, a program would not understand the reasoning behind the action. By retrieving a similar story, the program may find an explanation for the action. If the similar story contained: "Mary bought Larry chocolates because Mary liked Larry." The program could extrapolate from the similar events that Bob may like Jill, causing him to give her a gift. By finding similar stories, unexplained events can be better understood.

- *Grouping:* By using concept patterns, stories can be grouped into categories. A group of concept patterns that are rare overall, but common among a group of stories may constitute a genre. For example, conflict stories may generally involve an attack and a victory, while fairy tales may involve falling in love and living happily every after. Grouping stories together helps to organize information, and can make story retrieval faster, because if stories are pre-grouped, a retrieval system only has to search in one or a few genres to find the most similar story.

7. Contributions

- We have implemented several mechanisms for story comparison based on concept patterns.

- We have shown, with a small corpus of 15 conflict stories, that retrieval based on concept-pattern vectors produces precedents more like those found by domain experts (structure) than those found by novices (superficial features).

- We have demonstrated, at an illustration-of-concept level, a mechanism that discovers concept patterns in story ensembles by searching for parallel event patterns.

The next step in the development of the program is to conduct studies in which human subjects are given stories and asked to compare them on a concept level. This will establish a ground truth of story similarity, and will allow better testing of our system's modeling fidelity

8. Acknowledgments

Research on Genesis has been supported, in part, by the National Science Foundation (IIS-0413206), the Office of Naval Research (N00014-09-1-0597), the Air Force Office of Scientific Research (A9550-05-1-0321), and the Defense Advanced Research Projects Agency (FA8750-10-1-0076).

9. References

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*.

Morteza Dehghani, Dedre Gentner, Kenneth D Forbus, Hamed Ekhtiari, and Sonya Sachdeva. 2009. Analogy and moral decision making. *New Frontiers in Analogy Research Proceedings of the Second International Conference on Analogy*.

Brian Falkenhainer, Kenneth D. Forbus, and Dedre Gentner. 1989. The structure mapping engine: Algorithm and examples. *Artificial Intelligence*, 41(1):1–63.

Mark A. Finlayson and Patrick Henry Winston. 2005. Intermediate features and informational-level constraint on analogical retrieval. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pages 666–671.

Mark A. Finlayson. 2012. *Learning Narrative Structure from Annotated Folktales*. Ph.D. thesis, MIT.

D Gentner and K Forbus, 1991. *MAC/FAC: A model of similarity-based access and mapping*.

Wendy Lehnert. 1981. Plot units and narrative summarization. *Cognitive Science*, 5(4):293–331.

Shimon Ullman, Michel Vidal-Naquet, and Erez Sali. 2002. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7).

Patrick Henry Winston. 2011. The strong story hypothesis and the directed perception hypothesis. In Pat Langley, editor, *Technical Report FS-11-01, Papers from the AAAI Fall Symposium*, pages 345–352. AAAI Press.