Extracting Conceptual Relations from Children's Stories

Briane Paul Samson and Ethel Ong

Center for Language Technologies, De La Salle University, Manila, Philippines
briane.samson@dlsu.edu.ph, ethel.ong@delasalle.ph

Abstract. Automatic story generation systems require a collection of commonsense knowledge to generate stories that contain logical and coherent sequences of events appropriate for their intended audience. But manually building and populating a semantic ontology that contains relevant assertions is a tedious task. Crowdsourcing can be used as an approach to quickly amass a large collection of commonsense concepts but requires validation of the quality of the knowledge that has been contributed by the public. Another approach is through relation extraction. This paper discusses the use of GATE and custom extraction rules to automatically extract binary conceptual relations from children's stories. Evaluation results show that the extractor achieved a very low overall accuracy of only 36% based on precision, recall and F-measure. The use of incomplete and generalized extraction patterns, insufficient text indicators, accuracy of existing tools, and inability to infer and detect implied relations were the major causes of the low accuracy scores.

Keywords: Natural Language Processing, Text Analysis, Relation Extraction, Commonsense Knowledge.

1 Introduction

People use storytelling as a natural and familiar means of conveying information and experience to each other. During this interchange, people understand each other because we rely on a large body of shared commonsense knowledge. But computers do not share this knowledge, causing a barrier in human-computer interaction in applications requiring computers to understand and generate coherent text. To support this task, computers must be provided with a usable knowledge about the basic relationships between concepts that we find everyday in our world.

Creative text generation systems such as T-PEG [1] have utilized a semantic network representation of commonsense concepts to identify the relationships of words found in human puns. The extracted word relationships are then used as templates to enable computer systems to generate punning riddles, given the same repository of commonsense knowledge. This repository is ConceptNet [2].

Story generation systems, specifically Picture Books [3, 4], on the other hand, generate children's stories of the fable form for children age 4-8 years old by using a semantic ontology of commonsense concepts whose design was patterned after ConceptNet. The manually built knowledge repositories of the story generators contain

B.H. Kang et al. (Eds.): PKAW 2014, LNCS 8863, pp. 195–208, 2014.

binary conceptual relations about objects, activities and their relationships in a child's daily life. Later on, Yu and Ong [5] explored using a two-layer ontology with ConceptNet [6] as one of the major resource utilized to provide the upper story world knowledge about commonsense concepts.

Building knowledge repositories for story generation systems required a lot of manual effort. One way to address this was through the use of crowdsourcing techniques to acquire knowledge from the community, specifically from children [7] in order to learn concepts that are relevant for story generation. Another approach is through relation extraction.

Research works in relation extraction have achieved significant progress in extracting facts and concepts in the domains of newspapers [8], biographies [9], and legal documents [10]. For domain-independent data, systems like KnowItAll [11] extracted entities using generic noun phrase patterns. TextRunner [12], on the other hand, do not rely on predefined relation types but discovers relation triples automatically. These triples represent binary relations (arg1, relation, arg2) that were identified and extracted from a sentence.

However, limited work has been done on stories. Stories contain not only facts and conceptual entities, but also sequences of actions that characters perform or experience at various points in the story world. These descriptions on story events may span multiple sentences. Knowledge about how these events are ordered and the constraints under which they can occur must also be extracted.

This paper presents an approach to extracting binary conceptual relations from children's stories by defining a set of extraction rules that were then fed to GATE[1]. We refer to such relations as assertions representing storytelling knowledge and model them as an ontology of commonsense concepts. Section 2 identifies the types of storytelling knowledge needed by story generation systems with particular emphasis on domain-specific commonsense concepts. This is followed by a description of the process in defining and extracting conceptual relations on commonsense concepts from children's stories in Section 3. An analysis of the quality of the extracted assertions is then presented in Section 4. The paper ends with a discussion of issues and recommendations for future work.

2 Storytelling Knowledge

The knowledge needed by story generators can be classified into two broad categories, namely the operational knowledge about narrative structures and story plots to drive the flow of the story; and the domain knowledge that describes the story characters, the world, and the causal chain of actions and events. In this paper, the commonsense storytelling knowledge we referred to, specifically concepts and events, are classified under the domain knowledge. Concepts include concrete objects and their descriptions. Events include actions that story characters explicitly perform in the story world, the events that occur as a result of these actions, and events that are from

[1] GATE: General Architecture for Text Engineering, University of Sheffield. `https://gate.ac.uk`

naturally occurring phenomenon. Assertions in the form of binary semantic relations relate two concepts or events.

This section gives a brief overview of the different categories of semantic relations that are used to represent assertions in the commonsense knowledge repository of Picture Books. This serves as the basis for the types of assertions that are targeted by our relation extractor. A more detailed discussion of the knowledge representation of Picture Books and how this was utilized for story planning are beyond the scope of this paper and can be found elsewhere [3, 4].

2.1 Assertions Describing Concepts

A story world is comprised of various objects that interact with one another to achieve some form of a narrative plot. These objects, which include characters and the things that they manipulate, are described as part of the sequence of events that comprise the story flow. Character and object descriptions are two of the major factors that can motivate characters to exhibit certain behaviors, thus prompting them to perform actions in the story.

Character descriptions include roles, physical attributes, physical and mental states, capabilities to execute some actions, and emotions that a character may experience before or after the occurrence of event. Sample assertions for each of these are shown in Table 1.

Table 1. Semantic relations to describe story characters used in Picture Books [3, 4]

Category	Concept1
Concepts describing Character States	IsA(fighting, problem) IsA(itchy, discomfort) IsA(grounded, punishment)
Concepts describing Character Emotions	IsA (happy, emotion) IsA (scared, emotion) Feels(character, scared) Feels(character, sleepy)
Character Reaction to Events	EffectOf (break object, scared) EffectOf (meet new friends, smile)
Character States after an Event	EventsForGoalState(play games, happy)
Character Capabilities	CapableOf(character, hide)
Roles and Responsibilities	HasRole(character, king) ResponsibleFor(king, rule country)

The *IsA*, *EffectOf*, *EventsForGoalState* and *CapableOf* relations are adapted from ConceptNet. The *HasRole* and *RoleResponsibleFor* relations are used to model assertions that describe roles that characters may play in stories as well as the tasks associated with that role. These are currently not included in Picture Books. The rest of the relations in Table 1 were defined by Picture Books based on the requirements of the story planning task.

Objects are described based on their classification, properties and compositions, possible locations and co-located objects, and the actions that they can be used as instruments. Sample assertions for each of these are shown in Table 2. The *IsA*, *Property*, *PartOf*, *UsedFor* and *CapableOf* relations are adapted from ConceptNet. Concepts that model locations can also be associated with other concepts that describe them or their usage, as shown in Table 2.

Table 2. Semantic relations to describe objects used in Picture Books

Category	Concept1
Classification	IsA(doll, toy) IsA(ball, toy) IsA(marshmallow, food)
Properties and Compositions	HasProperty(lamp, fragile) HasProperty(marshmallow, fluffy) PartOf(wheel, truck) MadeOf(cake, sugar)
Location	LocationOf(toys, toy store) LocationOf(swing, park) OftenNear(swing, slide) HasProperty(camp, far) UsedFor(camp, camping) UsedFor(park, picnic)
Usage	UsedFor(toy, play) UsedFor(food, eat) UsedFor(water jug, drink)
Events on Objects	CapableOf(lamp, break) CanBe(toys, scattered)

2.2 Assertions Describing Events

Stories are comprised of sequences of events, which include explicit or voluntary events in the form of intentional character actions, and implicit or involuntary events that arise due to the execution of these actions or due to natural causes. Relations between events can be signified in two ways: temporal succession and causality. Temporal succession uses time to show the sequential relationship between two events. For instance, *Event A* happens before *Event B*. On the other hand, causality means *Event B* happened as a result of the occurrence or execution of *Event A*.

ConceptNet provides a number of relations to describe events, namely *EffectOf*, *FirstSubEventOf*, *EventForGoalEvent* and *EventForGoalState*. Table 3 shows sample assertions in Picture Books using these relations as well as additional relations that were defined specifically for story planning. The same set of relations is used to define both events and concepts. The story planner uses some other relations, e.g., *CapableOf* and *CanBe*, to signify that an event concept is an action that a character can perform. In Picture Books 2 [4], the *IsTransition* relation has been defined to model event assertions that describe the appearance, disappearance or movement of an object or character between two scenes in a story. The *negate* relation is used to model complementary events, usually a positive and a negative one.

Table 3. Semantic relations to describe events and actions

Category	Concept1
Causality of Events	EffectOf(hearing sound, scared) EffectOf(eating, sleepy) EffectOfIsState(become dirty, feel itchy)
Events to Achieve Goals	EventForGoalEvent(go to store, buy food) EventForGoalState(clean up, be neat)
Event Components	HasSubevent(sleep, brush teeth) HasSubevent(sleep, pray) HasSubevent(sleep, read story book)
Usage	UsedFor(toy, play) UsedFor(food, eat) UsedFor(water jug, drink)
Events on Objects	EventRequiresObject(play, toy) CapableOf(lamp, break) CanBe(toys, scattered)
Transition	IsTransition(bring, appearance) IsTransition(eat, disappearance) IsTransition(walk, movement)
Complements	Negate(sleep early, sleep late) Negate(eat healthy food, eat junk food)

Since stories are sequences of events, their analysis may necessitate the creation of new relations to represent sequences of events, temporal relations between events, as well as the constraints under which certain events may take place. For example, during testing, evaluators noticed that one of the generated stories of Picture Books occurred at an inappropriate time; specifically, the first segment of the story that introduces the day, the place, and the main character, contained the following text:

> *The evening was warm. Ellen the elephant was at the school. She went with Mommy Edna to the school.*

Although the temporal properties of events can be easily modeled in the Suggested Upper Merged Ontology as shown in the works of Cua et al for SUMO Stories [13], Picture Books' knowledge base currently does not provide relations about when certain events can take place. Furthermore, some granularities may be needed to model various aspects of time, namely season (planting can only occur during spring, snow can only fall during winter), month (Christmas in December, Valentine's in February), or even weeks, days, hours, and minutes. Assertions such as *Happens(Christmas, December)* and *Happens(going to school, morning)* can be defined by adopting the predicates used by Mueller [14] to model event occurrences.

3 Design and Implementation

The extraction process started with the gathering and preprocessing of the input corpus; followed by the creation of extraction templates; and lastly, the extraction of target relations using the open-source tool, GATE.

3.1 Data Gathering and Preprocessing

The input corpus for this research is comprised of 30 children's stories and include titles from the following: five (5) Topsy and Tim stories published by Ladybird Books and written by Jean Adamson and Gareth Adamson about the adventures of twins, and sixteen (16) stories from the Little Life Lessons: A Treasury of Values collection published by Publications International for children age 4-7 year olds; and seven (7) stories from the Jump Start series published by Scholastic, and two (2) Winnie the Pooh stories published by Disney Press for children age 8-10 year olds.

Each story in the corpus was modified to clean the data of any inconsistencies. Dialogues, for instance, show inconsistencies because of different writing styles, its conversational nature, use of informal language and colloquial words, and have incomplete thoughts. Thus, most dialogues were transformed into declarative sentences. The objective of these modifications was to convert the dialogues into complete and coherent sentences in order to yield proper extractions. It is important to note that even though the story has changed in terms of writing, the theme was retained. The intention was to make the actions and facts more apparent to the extraction tool.

Aside from dialogues, other modifications include: changing interjections into the emotions conveyed; expanding of contractions; removing of the punctuation marks (period) that do not denote the end of a sentence; and removing of words made-up by story characters such as *splendiferous* from the Winnie the Pooh titles.

3.2 Target Relations and Extraction Templates

For the purpose of this research, sixteen (16) relations were identified to be extracted; they were deemed relevant and helpful to the development of commonsense knowledge for the children's story domain. Table 4 shows the sixteen conceptual relations targeted in this research.

Extraction templates were defined for each of the target relations; they are the different ways a certain relation is manifested in a sentence or a span of text. Some of the templates were adopted from the ones used by ConceptNet to crowd-source data; the others were manually derived, especially for the *EventForGoalEvent*, *EventForGoalState*, *EffectOf*, *EffectOfIsState*, and *Happens* relations.

Table 4. Target relations

Relation	Description
IsA	Specifies what kind an entity is.
PropertyOf	Specifies an adjective to describe an entity.
PartOf	Specifies the *parthood* of an entity in another entity.
MadeOf	Specifies a component of an entity.
CapableOf	Specifies what an entity can do.
OftenNear	Specifies an entity near another entity in most instances.
LocationOf	Specifies the location of an entity.
UsedFor	Specifies the use of an object in an activity.
EventForGoalEvent	Specifies an event that causes the fulfillment of a goal event.
EventForGoalState	Specifies an event that causes the fulfillment of a goal state.
EffectOf	Represents a cause-effect between two events
EffectOfIsState	Represents a cause-effect between an event and an end state.
Happens	Specifies the time an event/state happens.
HasRole	Specifies the role of a person in the story.
RoleResponsibleFor	Specifies an action done by a role.
Owns	Specifies the ownership of an object.

Table 5 shows the different elements present in an extraction template; all these are tagged by the open-source tool, GATE. The first nine elements (9) in Table 5 can be tagged by default; the others were custom tags created for this research. The <*Indica-*

tor> element tag denotes the presence of a relational structure in a sentence, which is usually identified by the use of a transition word. Transition words are used mainly to aid in identifying explicit relations. The indicators were collated from ConceptNet, Picture Books 2 [4], and sentences from the corpora.

Table 5. Template elements

Default Tags	Custom Tags
<NP>, <NP:JobTitle>, <Noun:Possessive>, <AP>, <Pronoun:Possessive>, <Verb>, <VP>, <VP:Gerund>, <PP:Temporal>	<Event>, <GoalEvent>, <GoalState>, <Cause>, <Effect>, <EffectState>, <Indicator>

Shown in Table 6 are the templates used for the *PartOf* relation which can be extracted within a single sentence.

Table 6. Extraction templates of *PartOf* relation

Templates
<NP> <PartOfIndicator> <NP>
<Noun:Possessive> ... <NP>
<Pronoun:Possessive> ... <NP>

Here is a sample sentence to show the existence of the first template:

A window is a part of a house.

In the example, “A window” is the noun phrase or <NP>, “part of” is the <PartOfIndicator>, and “a house” is the second <NP>. When the extraction tool recognizes these three elements in this order, the *PartOf* relation will be extracted.

Table 5 shows the templates used for the *EventForGoalEvent* relation which can be manifested within a span of 2 sentences. Here are example sentences for this relation:

Kisha wants to buy a car. She saved all her lunch money.

The verb phrase “buy a car” is the <GoalEvent> in the first sentence, while “saved all her lunch money” is the <Event> in the second sentence. These occurrences indicate the existence of an *EventForGoalEvent* across the 2 sentences.

Table 7. Extraction templates of *EventForGoalEvent* relation

Templates
<GoalEvent> ... <Event>
<Event> <MotivationIndicator> <GoalEvent>

3.3 Extraction Tool

The architecture of the system was implemented using the open-source tool GATE. Figure 1 shows the GATE Application Pipeline designed for this research.

Each children's story in the corpora is processed by creating a GATE Document. These are then added to a GATE Corpus. First, the GATE application resets the input of any previous annotations. This applies only if the input has already been annotated before passing through GATE. After cleaning, the input is parsed into tokens. The input was then split into sentences and each token was annotated with their respective part-of-speech tags.

After that, named-entities, like common story character names and locations, were annotated. In resolving unique and story-specific named-entities like *Pooh* and *Tree Fort Island*, new gazetteer resources were used and new characters, locations and roles are added. New gazetteers were also created for indicators, world states and emotions, among others.

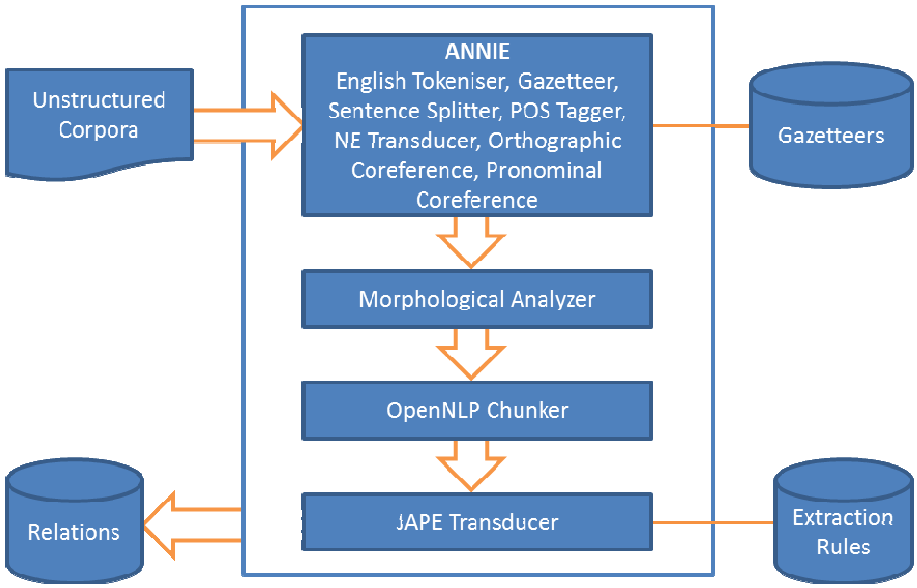

Fig. 1. GATE Application Pipeline

Then, pronouns are matched with the named-entities they are referring to in the text. Then, each token are processed to identify their lemmas, affixes and chunks. Lastly, the input text is run through a transducer to identify all the target relations and create the appropriate annotations for each. The transducer uses the defined templates and allows multiple relation types to be extracted from a single sentence.

4 Extraction Results and Analysis

The extracted relations were evaluated on their accuracy and completeness, and usability in the Picture Books system to generate stories. In evaluating the accuracy and completeness of the relations that were extracted, a gold standard was created where each story was manually tagged with the appropriate relations. Subsequently, precision, recall and F-measure values were computed and analyzed. Table 8 shows the results after comparing the automatically extracted relations from both the raw and modified versions of the stories to the gold standard.

Table 8. Evaluation results based on gold standard

Stories	Gold Standard	Extraction	False Positive	False Negative	Correctly Extracted	P	R	F
Overall (RAW)	663	662	421	422	241	0.36	0.36	0.36
Overall (MODIFIED)	663	615	403	451	212	0.34	0.32	0.33
Everybody Cries (RAW)	456	392	255	319	137	0.35	0.30	0.32
Everybody Cries (MODIFIED)	456	410	254	300	156	0.38	0.34	0.36
Start School (RAW)	139	150	103	92	47	0.31	0.34	0.33
Start School (MODIFIED)	139	173	119	85	54	0.31	0.39	0.35
Hopsalot's Garden (RAW)	68	73	45	40	28	0.38	0.41	0.40
Hopsalot's Garden (MODIFIED)	68	79	48	37	31	0.39	0.46	0.42

Overall, the extraction tool had a precision of 0.34; recall of 0.32; and an F-measure of 0.33. Furthermore, it was able to get relatively high accuracy scores (F-measure) for the *PropertyOf*, *PartOf*, *CapableOf* and *Owns* relations; each having scores of 0.40, 0.42, 0.47, and 0.62, respectively. The high accuracy of the aforementioned extracted relations can be attributed to the simplicity of their most common pattern in a sentence. For example, majority of the stories in the corpus contain the following patterns used to extract the *Owns* assertion:

<Noun:Possessive> ... *<Noun>*
<Pronoun:Possessive> ... *<Noun>*

The *CapableOf* relation looks for a noun and an immediate verb in a sentence, thus, easily extracting assertions from sentences such as the following:

Pierre threw a ball to Puppy. ==> *CapableOf(Pierre, throw)*

On the other hand, approximately 64% of the extracted relations were false positives, especially for the *IsA*, *EffectOf*, *UsedFor*, *Happens*, and *HasRole* relations. These are extracted relations not found in the gold standard. Such results suggest that the defined templates were too generalized and all-encompassing. Consider the following pattern for extracting *IsA* relations:

<Noun or Gerund> <,>[0,1] <Determiner> + ... <Noun>

This yielded the assertion *IsA(having, day)* from the following story text:

The best thing about having the worst day ever is that tomorrow will be a lot better.

Some inaccuracies of the part-of-speech tagger and issues with the gazetteer contributed to the extraction of false positives as well. Consider the sentence below:

Bear's first game is tomorrow.

The relation *IsA(game, tomorrow)* was extracted because both the words *game* and *tomorrow* were tagged as noun, which satisfies the template:

<Noun> is <Noun>

In the case of the *HasRole* relation, the assertion *HasRole(he, driver)* was extracted from the following sentence without resolving the pronoun *he*.

He cannot tell if the driver sees him, though.

The *EffectOf* relation also received a number of false positives with the following generic template pattern used more often:

<Event:VP> < . > ... <Event:VP>

The pattern accepts any tagged *Events* from two adjacent sentences, leading to the extraction of the assertion *EffectOf(made the team, gets closer)* from the following pairs of story text:

Bear made the team.
As he gets closer, Pig sees that Puppy has been crying.

Lastly, there were target relations that did not have any extractions at all because of their high dependence on indicators, incorrect part-of-speech tags and limitation on the number of sentences it can extract from. For example, the assertion *Happens(walk to school, today)* was extracted because the system used the word *today* as the time indicator for the given activity.

On the quality of the extraction, it is important to note that for event relations like *EffectOf* and *EventForGoalEvent*, the extracted relations seem to be longer and more specific because the extractor uses whole phrases as concepts. This may be different from the concepts of Picture Books that are more generalized. Here are some example extractions:

EffectOf(looked at the map,checked the wind)
EffectOf(pours something into the volcano,stopped him)
EventForGoalEvent(called everyone,go to the ship)

Out of all the extracted relations, only 6 relations were acceptable to be used by Picture Books, and all of them are *PartOf* relations. The other extracted relations are not aligned with the existing themes, thus generating incoherent story text; or are too specific to the story they were extracted from.

5 Conclusion and Further Work

Researches in the field of natural language processing (NLP) seek to finds ways to make human-computer interaction more fluent. But human-computer communication is hampered by the lack of a shared collection of common sense knowledge that people rely on when they communicate in order to understand each other. In order to make computers achieve the same level of expressiveness as humans, we must provide them "a common knowledge with richness that more closely approaches that of the human language." [15]

Although dedicated IE systems have been developed to extract information from various domains, this research is a first step towards extracting relations from children's stories. And based on the results obtained through the evaluation of the extraction tool, it was proven possible to extract new semantic relations from children's stories and feed them into Picture Books' ontology. However, the extraction tool was found to be inaccurate in doing so. Overall, it only got a precision of 0.34; recall of 0.32; and an F-measure of 0.33. Therefore, the automatically extracted relations were mostly incorrect and the extraction tool was not able to extract all expected relations in a given text. As for their quality, it was greatly affected by the common sentence structures in a story, the quality and accuracy of the part-of-speech tagging, the limitations of the defined extraction templates, and the completeness of the indicators.

After evaluation, it is conclusive that as the sentence structures become more complex and the length of the story increases, the extractions get less accurate. It exposes a limitation on the templates used as they can only successfully handle simpler sentences and simpler manifestations of a relation in a text.

For the extraction rules, an attempt to handle all sentence patterns with the least number of rules has caused exceptional cases to not be covered. Moreover, these rules cannot handle implied and inferred relations, and the different senses of a word. Lastly, the templates were limited to extract event relations from one or two adjacent sentences only.

The prevalent use of indicators in most of the extraction templates posed a limitation on the number and quality of extractions done. First, in most cases, indicators are not always used because of their formality. This also assumes that the concepts constituting a relation are within a sentence. If not, it is assumed that the second concept is in the next sentence, the subject pronoun referring to the first concept, and the whole thing signaled by an indicator.

There are also redundant extracted relations that were not generalized into a single binary relation which made the ontology cluttered. Lastly, the binary nature of the relations caused some to become unusable. Most event relations were too specific, because of the existence of direct and indirect objects, and character names.

To address these deficiencies, it is recommended to incorporate as many patterns as possible to improve the extraction rules. Such patterns should also include those that span more than two sentences. There must be an increased focus on extracting event relations as they are not usually explicitly indicated in a span of text. Such relations also constitute the bulk of a story. Building an accurate cause-effect chain of events would be very beneficial for most creative text generation systems.

And in improving the quality of the ontology, future works should consider storing metadata, like frequency and direct object, with the binary relations to reduce specificity and improve usability. Lastly, relations can be further refined by using a language resource that can supply accurate semantic information.

References

1. Hong, B.A., Ong, E.: Generating Punning Riddles from Examples. In: Proceedings of the Second International Symposium on Universal Communication, Osaka, Japan, pp. 347–352 (2008)
2. Liu, H., Singh, P.: ConceptNet - A Practical Commonsense Reasoning Tool-kit. BT Technology Journal 22(4), 211–226 (2004)
3. Hong, A.J., Solis, C., Siy, J.T., Tabirao, E., Ong, E.: Planning Author and Character Goals for Story Generation. In: Proceedings of the NAACL Human Language Technology 2009 Workshop on Computational Approaches to Linguistic Creativity, pp. 63–70. ACL, Colorado (2009)
4. Ang, K., Yu, S., Ong, E.: Theme-Based Cause-Effect Planning for Multiple-Scene Story Generation. In: Proceedings of the 2nd International Conference on Computational Creativity, Mexico City, Mexico, pp. 48–53 (2011)
5. Yu, S., Ong, E.: Using Common-Sense Knowledge in Generating Stories. In: Anthony, P., Ishizuka, M., Lukose, D. (eds.) PRICAI 2012. LNCS (LNAI), vol. 7458, pp. 838–843. Springer, Heidelberg (2012)
6. Liu, H., Singh, P.: Commonsense Reasoning in and over Natural Language. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3215, pp. 293–306. Springer, Heidelberg (2004)
7. Chua Jr., R.C., Ong, E.: Commonsense Knowledge Acquisition through Children's Stories. In: Richards, D., Kang, B.H. (eds.) PKAW 2012. LNCS (LNAI), vol. 7457, pp. 244–250. Springer, Heidelberg (2012)
8. Muslea, I.: Extraction Patterns for Information Extraction Tasks: A Survey. In: Proceedings of the AAAI 1999 Workshop on Machine Learning for Information Extraction (1999)
9. Alani, H., Kim, S., Millard, D., Weal, M., Lewis, P., Hall, W., Shadbolt, N.: Automatic Extraction of Knowledge from Web Documents. In: Proceedings of ISWC 2003 Workshop on Human Language Technology for the Semantic Web and Web Services, Florida (2003)
10. Cheng, T.T., Cua, J., Tan, M., Yao, K.: Information Extraction for Legal Documents. Unpublished undergraduate thesis, De La Salle University, Manila (2010)

11. Etzioni, O., Cafarella, M., Downey, D., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S., Yates, A.: Unsupervised Named-Entity Extraction from the Web: An Experimental Study. Artificial Intelligence 165, 91–134 (2005)
12. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open Information Extraction from the Web. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, pp. 2670–2676 (2007)
13. Cua, J., Ong, E., Manurung, R., Pease, A.: Representing Story Plans in SUMO. In: NAACL Human Language Technology 2010 Second Workshop on Computational Approaches to Linguistic Creativity, pp. 40–48. ACL, Stroudsburg (2010)
14. Mueller, E.: Modelling Space and Time in Narratives about Restaurants. Literary and Linguistic Computing 22(1), 67–84 (2007)
15. Niles, I., Pease, A.: Towards a Standard Upper Ontology. In: Proceedings of the International Conference on Formal Ontology in Information Systems, pp. 2–9. ACM, New York (2001)