

A Verb-centered Ontology Modelling for Text Understanding

Pyung Kim

Dept. of Computer Education
Jeonju National University of Education
Jeonju, Korea
pyung@jnue.kr

Abstract—The role of illustration is to illustrate specific words used in sentence or to express a specific situation. To automatically generate an illustration, the sentences used in contents must be understood and expressed as pictures. This requires an ontology consisting of parsing results, characters, words, affiliations, backgrounds, props, location and time. Manual correction is also required once the scene is configured. In order to improve the productivity of contents based on auto-generation of illustrations, studies for the auto-generation of illustrations are continuously performed. This study models a verb-centered ontology to be used for character-background-prop arrangement, arrangement and emotion of characters, time, weather and seasonal expression necessary for constructing illustrations of Korean traditional fairy tales. As part of the ontology used to construct the illustrations, the verb ontology not only plays a fundamental role for each verb type but also can be used for more naturally representing the behavior, placement, and emotion of characters appearing in the illustration, using different verb concepts and various relationships between concepts.

Keywords-ontology modelling; verb-centered ontology; text understanding

I. INTRODUCTION

An illustration is a picture inserted to a book (copy, print, etc.), magazine, newspaper, advertisement, etc. for the purpose of supplementing or emphasizing the contents of sentences [1]. The illustration is to explain specific words used in a sentence or to express a specific situation using a picture, and the technique for auto-generating the illustration helps save time and cost of making illustrations in the process of publishing a book.

Automatic generation of illustrations require understanding of sentences used in the content and expressing them with pictures. Several steps are needed, such as parsing sentences, recognizing entity names, resolving cross-references, identifying paragraphs that comprise a scene, composing a scene, and manually adjusting the final illustration through manual correction. In this process, it should be possible to identify and arrange the background, characters and props necessary for generating the illustration and to express the characters' motion and emotion as well. Especially, in a fairy tale, illustrations can be considered more important because they can have a great influence on understanding the contents of the story.

Ontology is a knowledge expression model and is widely used in intelligent services or semantic-based information services because it is suitable for expressing the relationship among concepts [2]. In this study, we model a verb-centered ontology as a relational model for verbs used for generating illustrations. For this purpose, we select verbs that affect constructing scenes for illustrations for Korean traditional fairy

tales, and classify them into action, emotion, arrangement, direction, time, season, and weather verbs.

Chapter 2 of this paper describes the related studies on auto-generation of illustrations and on verb-centered ontology; Chapter 3 the steps of auto-generation of illustrations and entire elements of ontology necessary for illustration; Chapter 4 the roles of each verb type appearing in sentences used for fairy tale illustration and the verb-centered ontology; and Chapter 5 conclusions and future works.

II. RELATED STUDIES

The study of text as scenes has been performed for a long time in order to help understand sentences or incidents [3,4,5]. NALIG [3] accepts a description as a set of simple spatial relations between objects and draws an imaginary scene on a screen by means of a sequence of simple phrase. To generate 3D scenes and images from a user's description, WordsEye [4] focuses on translating the semantic intent of the user. WordsEye works by converting a parse tree to a dependency representation, which is then converted into a semantic representation. Carsim[5] can enable people to imagine a traffic situation and understand the course of events properly. It is composed of natural language processing module, spatio-temporal planning and inference module and graphical module. Carsim analyzes texts describing car accidents and visualizes them in a 3D environment.

Studies of the image auto-generation have continued[6,7]. To generate 3D scene, Chang et al. [6] use a dataset of 3D scenes annotated with natural language descriptions and learn from this data how to ground textual descriptions to physical objects. To perform detailed text-to-scene conversion from

books, Copper [7] suggests an approach to inferring objects using Wikipedia and WordNet. In addition to these studies, studies to generate sentences to describe 2D images [8,9] and to express specific objects of the images have been performed[10].

The ontology has been used in various ways for the services to define and utilize the relationship between meanings of words and meanings, and the studies on auto-generation of ontology through sentence parsing have also been conducted[11,12]. ASUIM [11] learns subcategorization frames of verbs and ontologies from syntactic parsing of technical texts in natural language. TextOntoEx [12] constructs ontology from natural domain text using semantic pattern-based approach. TextOntoEx is a chain between linguistic analysis and ontology engineering. TextOntoEx analyses natural domain text to extract candidate relations and then maps them into meaning representation to facilitate constructing ontology.

Ma et al. [13] suggest the notion of visual valency and use it as a primary criterion to recategorize eventive verbs for language visualization (animation) in our intelligent multimodal storytelling system, CONFUCIUS. IMAGACT [14] is a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech corpora, which makes use of the universal language of images to identify the different action types extended by verbs referring to action in English, Italian, Chinese and Spanish. IMAGACT makes explicit the variation of meaning of action verbs within one language and allows comparisons of verb variations within and across languages.

III. AUTOMATIC GENERATION OF ILLUSTRATION

To automatically generate illustrations for a fairy tale, we should be able to understand the text centering around the objects appearing in the text of the fairy tale. Various elements such as background, persons, animals, plants, props, arrangement of characters or props, and expression of motion or emotion are used to generate the illustration. This chapter describes the steps to create the illustration, the ontology that defines the elements required for generating the illustration and the relationship among the elements, and the detailed steps to construct the scenes.

A. Overall Process for Automatic Generation of Illustration

Generation of the illustration consists of 6 steps as shown in Fig.1. Step 1 is to parse sentences and analyze their structures; Step 2 to tag entity names on a dictionary basis, to identify types of the entity names, and to determine their meanings based on a predetermined dictionary; Step 3 to determine the continuity of characters, places, and time through cross-reference resolution; Step 4 to segment paragraphs for constructing one scene in consideration of continuity of sentences, in which it is determined whether or not to insert an illustration; Step 5 to compose a scene using elements shown in

the sentences; and Step 6 to manually correct the illustration that is automatically generated through the above steps.

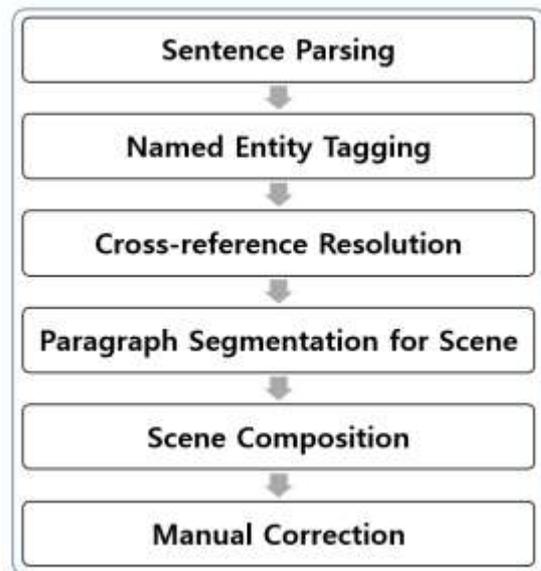


Figure 1. Overall Process for Generation of Illustration

B. Ontology for Automatic generation of Illustration

To automatically generate illustrations to express the meaning of sentences, it is necessary to define the elements that make up the illustration and to understand the relationships among those elements. In the process of automatic generation of the illustration, the ontology constructed for auto-generation of illustration is utilized. In this study, the elements that make up the illustration are defined with the ontology, as shown in Fig. 2. The ontology defines 6 classes: character, prop, affiliation, verb, time, and background. The character is divided into sub-classes of person, animal and plant, and the background is divided into such subclasses as building and nature.

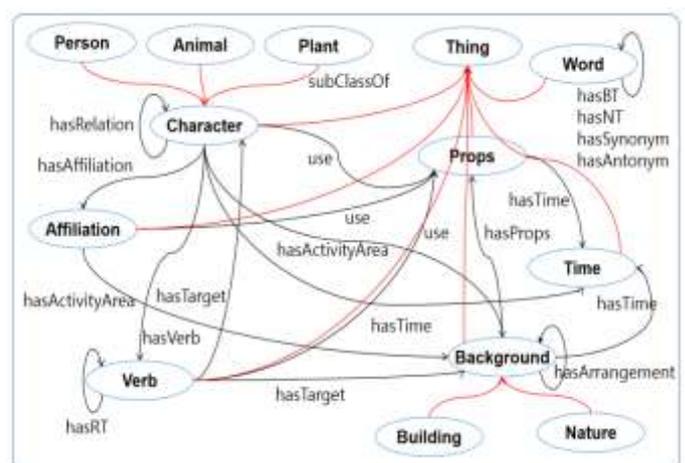


Figure 2. Ontology of All Components for Illustration

The roles of classes constructing the ontology are as follows:

- Characters: Those who make up the scene, such as persons, animals, and plants. An animal(s) and/or a plant(s) can be a subject(s) that may perform an action(s), in fairy tales.
- Props: Those required to compose a scene, which can be divided into those used for setting up the background and those required for composing the characters' occupation or actions.
- Affiliation: Expresses the character's occupation, and affects background or props, and basic shape of characters.
- Time: Represents time, season, and age.
- Background: Constructs scenes, with nature, buildings, and props, and includes basic arrangement of them.
- Word: Defines Broad Terms (BTs), Narrow Term (NTs), synonyms and antonyms of the verbs in a fairy tale.
- Verb: Describes the information required for composing the placement and direction among characters, the placement and direction between background and characters, and the expression and motion of characters, and the relationship among verbs (Related Term: RT)

IV. VERB CATEGORIZATION FOR SCENE COMPOSITION

This Chapter defines the detailed steps required for composing scenes, and the process of extracting and categorizing the verbs used for generating illustrations.

A. Detailed Process for Scene Composition

In the scene composition step, various information including parsing results, ontology, arrangement, age information and temporal information is required. Accuracy of the scene composition in this step significantly affects the manual correction in the next step.

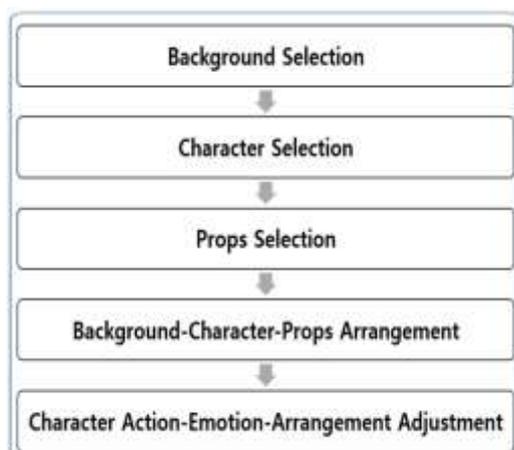


Figure 3. Detailed Process for Scene Composition

As shown in Fig. 3, the steps that constitute the scene, which is the 5th step of the auto-generation of illustrations, can

be subdivided into another 5 steps. Step 1 is to select a background needed for creating illustrations from sentences separated by semantic units, which is a process of selecting and arranging buildings and nature considering the time and age. Step 2 is to select characters, such as persons, animals, and plants in consideration of time, age, occupation, and places of activities. Step 3 is to select props to be used in the scene, which can be used as components of a background but also as a tool used by a character. In other words, props can be used to express a character's occupation or behavior. Step 4 is to determine the layout of background, characters, props, etc., and there is a basic layout of persons, animals, plants, buildings and nature, and a semantic arrangement to adjust characters and backgrounds depending on the meaning of verbs. Step 5 is to express the characters' behaviors or emotion, by determining the characters' line of sight, placement and movement based on verbs.

B. Verb Extraction and Categorization

We extracted verbs from 30 Korean traditional fairy tales, and categorized them. The Korean traditional fairy tales unfold stories with persons, animals, and plants as characters, in various backgrounds such as mountain, sea, river, and fields. Due to the nature of the fairy tale, animals and plants may have been personified, and there are also unusual backgrounds such as in the sea or inside a whale's stomach.

Extraction and categorization of the verbs used for generating the illustration are done manually and consists of the following steps. Step 1 is to extract verbs by parsing the fairy tales. Step 2 is to select main sentences to be used for the illustration by comparing the illustration and the sentences. As one or more sentences can be reflected in the illustration in this step, all sentences that affect the illustrations should be selected. Step 3 is to classify the extracted verbs into words related to movement, placement, emotion, direction, place, time, season, and weather. Step 4 is to select synonyms and categorize them into key concepts. The processing of synonyms and antonyms is handled by the entire ontology, not by the verb-centered ontology. Step 5 is to set up the relationship among words, to enable the processing of words belonging to a specific category when they affect other categories. In addition, in order to process the case where verbs are used as different meanings when they are used in conjunction with a particular word, verbs are distinguished with phrases and the phrase patterns are compared with the parsing result.

The verbs used for composing a scene are divided into 9 groups, and then assigned to the respective groups. They are used to arrange characters, props, background, and basic layout based on the parsing result. Props are used to express tools and actions to represent the characters' occupation, or the background. If a number of characters appear in the illustration, the props should be arranged in a way that characters are not overlapped. Depending on the verb group, action verbs express

the character’s action; emotion verbs their emotion with action and facial expression; direction verbs indicate the characters’ direction; arrangement verbs determine the placement of characters depending on action and emotion. They are also used to determine time, place, season, or weather or used as idioms and distinguish verbs that are used differently from their ordinary meanings. Meaning of each verbgroup and its effect on scene composition are as follows:

- Action Verbs: Express the character’s action, which is the most basic verb category in the verb-centered ontology. Actions can be divided into several groups: Stand-Sit-LayDown-Fall Down, Walk-Run-Fly-Swim, Give-Catch-Drop, and Speak-Fight-Dance-Shake. Verbs in the action group can determine the placement of a character, his or her emotion, and the place where he or she is. For example, a verb belonging to “Fight” group may have characters face one another, and render their emotion as angry.
- Emotion Verbs: Express the character’s emotion. Emotions can be grouped intoPleasure-Anger-Sadness-Joy. A verb in the emotion group may determine the placement and actions of characters. For example, a verb belonging to “Pleasure” group may have characters face or surround one another, and have them dance.
- Arrangement Verbs:Express the arrangement of characters, which can be grouped into BeOpposite-Follow-Surround-TurnBack-Accompany.A verb in the arrangement group may determine the action or emotion of a character. For example, a verb belonging to “TurnBack” groupmay illustrate the character as standing and render his or her emotion as angry.
- Direction Verbs: Express the direction a character is facing, as Left-Right-Up-Down.
- Time Verbs: Express time, as Morning-Afternoon-Dinner-Night.
- Season Verbs: Express a season, as Spring-Summer-Fall-Winter.
- Location Verbs: Express a location, Land-Sky-Sea_River-Mountain-Road-Underground. A verb in this group can determine the action of a character that may be taking place in the corresponding location. For example, a verb in “Sky” group may illustrate a character as waiving a hand.
- Weather Verbs: Express weather, as Sun-Rain-Cloud-Wind-Snow.
- Phrase: Expresses a special meaning, not general meaning of a verb or verbs, in conjunction with the result of parsing. For example, “hit the ground” may mean an act of physically hitting the ground but can also be used to express the emotion of being angry or sad.

One verb may be assigned to two or more groups. Also, one verb can be related two or more groups such as action, emotion, arrangement, time, and location. Verb groups can be correlated and individual verbs in each group can also be correlated to one another within the group. Table 2 shows the number of verbs in each group, extracted from 30 fairy tales, together with respective verb samples.

TABLE I. VERB LIST BY CATEGORY

Category	Verb List (Example)
Action Verb	run, eat, slip, call, stand, fight, sit, kneel down, catch, like, collapse, reconcile, work, play, reject, greet
Emotion Verb	delight, surprise, get angry, be sad, ardent, boast, embarrass, regret, envy
Direction Verb	look <u>upward</u> , look <u>down</u> , look <u>left</u> , look <u>right</u>
Arrangement Verb	oppose, accompany, follow,be <u>opposite</u> , turn <u>back</u>
Weather Verb	rain, snow, freeze, be <u>cold</u> , be <u>hot</u> , blow
Time Verb	darken, be bright, dawn
Season Verb	sprout, freeze, harvest,
Location Verb	swim, fly, walk, climb, run
Phrase	shed <u>a tear</u> -> sadness, hit <u>the ground</u> -> sadness, speak <u>loudly</u> -> anger or happy

V. VERB-CENTERED ONTOLOGY FOR SCENE COMPOSITION

In this study, we modelled a verb-centered ontology in the process of auto-generating illustrations for fairy tales, based on verbs required for expressing the arrangement of characters and their emotion and facial expressions. As shown in Figure 4, the verb-centered ontology classifies verb types into following 9 sub-classes: action, emotion, arrangement, direction, time, season, location, weather, and phrase.

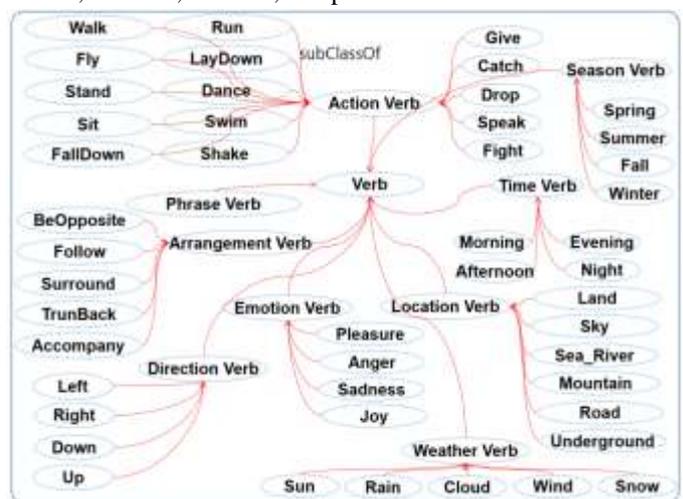


Figure 4. Class Hierarchy of Verb-centered Ontology.

Each sub-class of verbs has various “hasRT” relationships with other sub-classes, such as action-emotion-arrangement, emotion-action-arrangement, arrangement-action, location-action, weather-action, season-action, and time-action-emotion. These relationships are used to more naturally express the action, arrangement and emotion of characters shown in illustrations.

VI. CONCLUSION

In order to automatically generate illustrations, it is necessary to understand sentences used in the contents and be able to express them with pictures. For this, sentences should be parsed and appropriated sentences should be selected for scene composition. For scene composition, characters, props and background should be selected, and their arrangement and actions also be designated.

In this study, we categorized the verbs used for composing scenes in the Korean traditional fairy tales into groups and then modelled a verb-centered ontology. The verbs were grouped to determine action, emotion, arrangement, direction, time, season, and location, and they are co-related to each other. The verb-oriented ontology is useful for expressing the arrangement of character-background-props as well as the placement, action, and emotion of characters.

We modelled a verb-centered ontology for the auto-generation of illustrations. For the auto-generation of illustrations, however, it is necessary to expand the ontology to include characters, background, and props in addition to verbs. It is also necessary to complement verbs from various documents and learn various patterns. This requires expanding the verb-centered ontology to cover the sentence patterns used to express scenes in fairy tales and developing ontology for various components.

ACKNOWLEDGMENT

This work was supported by the cost of education, research and student guidance at Jeonju National University of Education in 2016.

REFERENCES

- [1] <https://en.wikipedia.org/wiki/Illustration>
- [2] D. Preuveneers, et al., "Towards an extensible context ontology for ambient intelligence," European Symposium on Ambient Intelligence. Springer Berlin Heidelberg, 2004.
- [3] G. Adorni, M. Di Manzo and F. Giunchiglia, "Natural language driven image generation," In Proc. of the 10th international conference on Computational Linguistics, pp. 495–500, 1984.
- [4] B. Coyne and R. Sproat, "WordsEye: an automatic text-to-scene conversion system," In Proc. of SIGGRAPH, pp. 487-496, 2001.
- [5] R. Johansson, A. Berglund, M. Danielsson and P. Nugues, "Automatic Text-to-Scene Conversion in the Traffic Accident Domain," IJCAI, vol. 5, 2005.
- [6] A. Chang, W. Monroe, M. Savva, C. Potts and C. D. Manning, "Text to 3D Scene Generation with Rich Lexical Grounding," ACL 2015.
- [7] A. Cropper, "Identifying and inferring objects from textual descriptions of scenes from books," OASICS-Open Access Series in Informatics, vol. 43, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.
- [8] G. Kulkarni, et al., "Babytalk: Understanding and generating simple image descriptions," IEEE Transactions on Pattern Analysis and Machine Intelligence 35.12, pp. 2891-2903, 2011.
- [9] A. Karpathy, A. Joulin and F. F. F. Li, "Deep fragment embeddings for bidirectional image sentence mapping," Advances in neural information processing systems, pp. 1889-1897, 2014.
- [10] N. FitzGerald, Y. Artzi and L. Zettlemoyer, "Learning Distributions over Logical Forms for Referring Expression Generation," EMNLP, 2013.
- [11] D. Faure and C. Nédellec, "A corpus-based conceptual clustering method for verb frames and ontology acquisition," LREC workshop on adapting lexical and corpus resources to sublanguages and applications, vol. 707, no. 728, 1998.
- [12] M. Y. Dahab, H. A. Hassan and A. Rafea, "TextOntoEx: Automatic ontology construction from natural English text," Expert Systems with Applications 34.2, pp. 1474-1480, 2008.
- [13] M. Ma and P. Mc Kevitt, "Visual semantics and ontology of eventive verbs," International Conference on Natural Language Processing, Springer Berlin Heidelberg, 2004.
- [14] M. Moneglia, S. Brown, F. Frontini, G. Gagliardi, F. Khan, M. Monachini and A. Panunzi, "The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the representation of lexical encoding of Action," In LREC pp. 3425-3432, 2014.
- [15] F. Frontini, et al., "Verb interpretation for basic action types: annotation, ontology induction and creation of prototypical scenes," 24th International Conference on Computational Linguistics, 2012.