

Interactive Media Creation by Reusing Video Contents

Minho Han, Wonjoo Park, Jeong-Woo Son, Sun-Joong Kim

ETRI (Electronics and Telecommunications Research Institute), Korea

mhhan@etri.re.kr, wjpark@etri.re.kr, jwson@etri.re.kr, kimsj@etri.re.kr

Abstract— Currently, to create an interactive media provided multimodal interface, an expert shoots scenes based on scenario, and then edits scenes using special program. In the market, the technology that user can easily create an interactive media by reusing video contents is required. In this paper, we propose the method for creating an interactive media by only user input scenario.

Keywords— Interactive Media, Video Contents, Ontology

I. INTRODUCTION

Interactive media normally refers to products on digital computer-based systems which respond to the user's actions [1]. Here, the interactive media is limited to the video that is played based on the user's response. To create the interactive media, natural language-based video searching and user interaction insertion technology is required.

Natural language processing (NLP) is a technology for enabling the computer to understand human language. NLP is composed of morpheme analysis, part-of-speech (POS) tagging and phrase analysis. Morpheme analysis is the first step of NLP for extracting meaning from natural language string [2]. That is to extract the words of the smallest mean from the sentence. POS tagging is to attach the correct part-of-speech for a word to resolve ambiguousness [3]. Phrase analysis is the final step of determining the structure of sentence using morpheme analysis and part-of-speech information [4]. When the phrase analysis is completed, a parse tree is generated.

In computer science and information science, an ontology is a concept belonging to a specific domain as a data model for describing a specific field, and consists of a set of fixed form vocabulary for describing the relationship between the concepts [5]. Scene ontology was built, and in order to get the desired scene through the ontology, you need to create a RDF (Resource Description Framework) triple using a parse tree. If RDF triples can be converted to SPARQL query, it is possible to get the desired scene from the ontology using this query [6].

Multimodal interface is the base technology for analysis and encoding information on human gesture and hand movement, including keyboard, pen and voice for communication between man and machine. It usually allows a driver to safely and intuitively manipulate an infotainment system while driving [7]. If the user can insert the multimodal

interface between the scenes, it may be created the interactive media that is played based on the user's response.

In this paper, we present to a method that anyone can easily create an interactive media.

II. PROPOSED SCHEME

Figure 1 depicts the system architecture to create the interactive media. It takes a natural language scenario as input, and finally returns an interactive media as output. Once the user inputs a natural language scenario, NLP (Natural Language Processing) module first produces a parse tree. Upon receiving the parse tree, the searching module processes it in two steps:

- The parse tree is translated into SPARQL queries.
- And then, it finds the scene index from ontology.

Finally, the controller gets the scene from storage by using this index.

The scenario is composed of several sentences, and the controller gets the scene in sentence unit. Therefore, several scenes are searched based on scenario, and the user can create an interactive media by inserting the user interaction between scenes.

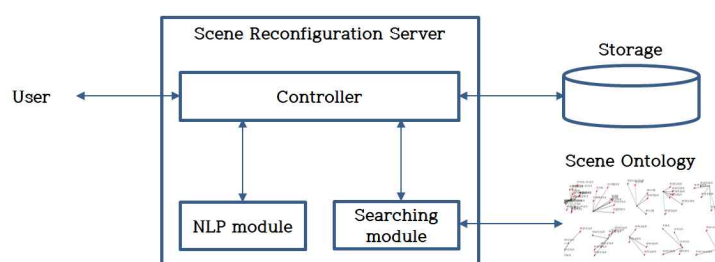


Figure 1. System Architecture

Prior to the operation, the video contents should be divided into scenes and stored in storage. Also, the administrator has to build ontology to find the stored scene.

A. Controller

The controller has interfaces with user, NLP module, searching module and storage. Upon receiving the scenario from the user, the controller processes it in three steps:

- The controller performs the natural language processing by sentence unit through NLP module.
- And then, the controller finds the scene index from ontology through searching module.
- Finally, the controller gets the scene from storage by using this index.

Users can connect to a controller (Scene Reconfiguration Server) via the Web or dedicated apps. Also, it may be searched several scenes about each sentence, and the user can select a preferred scene through controller.

B. NLP module

Figure 2 depicts the architecture and interface of NLP module. The NLP module has an interface with a controller, and performs a natural language process on the scenario input by the user. As the first step of natural language analysis, the morpheme analyzer provides the ability to change the input string into morphemes. The PoS tagger provides part-of-speech (PoS) tagging based on morphological analysis result. And finally, the phrase analyzer performs the phrase analysis based on part-of-speech tagging results, and then generates several parse trees.

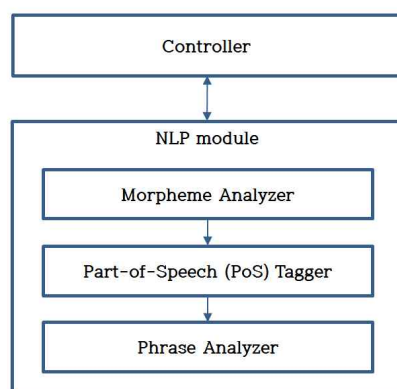


Figure 2. Architecture and Interface of NLP module

C. Searching module

Figure 3 depicts the architecture and interface of searching module. The searching module has an interface with a controller, and provides the ability to search the scene index by using natural language processing result. NLP module (phrase analyzer) generates several parse trees. Therefore, the parse tree analyzer provides the ability to find the optimal parse tree among them. The RDF Triple generator extracts objects, relationships and events by using the optimized parse

tree, and then generates the RDF triples by using the scene ontology. The SPARQL query generator generates the SPARQL query by using the RDF triples, and then finds the scene index from ontology by using this query.

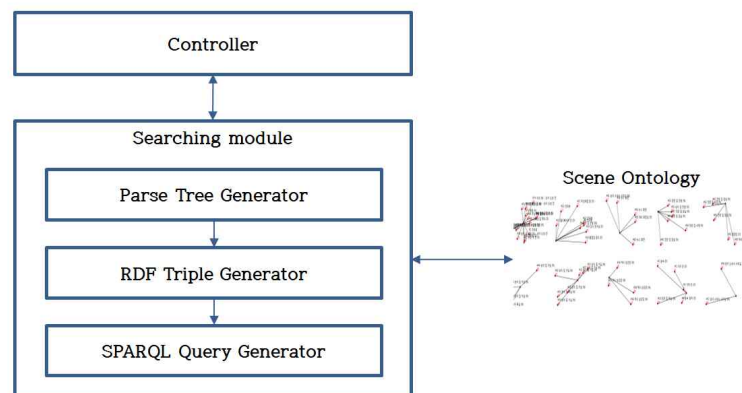


Figure 3. Architecture and Interface of Searching module

III. OPERATING SCENARIO

Figure 4 depicts the operating scenario for interactive media creation. The operating scenario of system is as follow:

- (1) To create interactive media, the user inserts a natural language-base scenario in each scene unit. For example, to create interactive media that is composed of the N scenes, the user must insert each natural language-based sentence about the scene 1, scene 2, ..., scene N.
- (2) The scene reconfiguration server performs the natural language processing by sentence unit, and searches the scene about each sentence.
- (3) It may be searched several scenes about each sentence, and the user selects a preferred scene among them. In addition, the user reconfigures the searched scenes based on scenario.
- (4) To be played according to user interaction, the user inserts a user's interaction between the reconfigured scenes.
- (5) The user requests an upload for the reconfigured scenes to the server.
- (6) The server stores the reconfigured scenes in the storage.

When the reconfigured scenes are stored in the storage, the server does not store a duplicate scene in the storage. Instead, the indexes of the reconfigured scenes and user's interaction between the scenes are only stored. For example, if the interactive media generated according to a user's scenario is composed of scene S1, scene S3, and user interaction I4 (action to raise right hand), the server stores the index 'S1-I4-S3'. This structure increases the efficacy of the storage space.

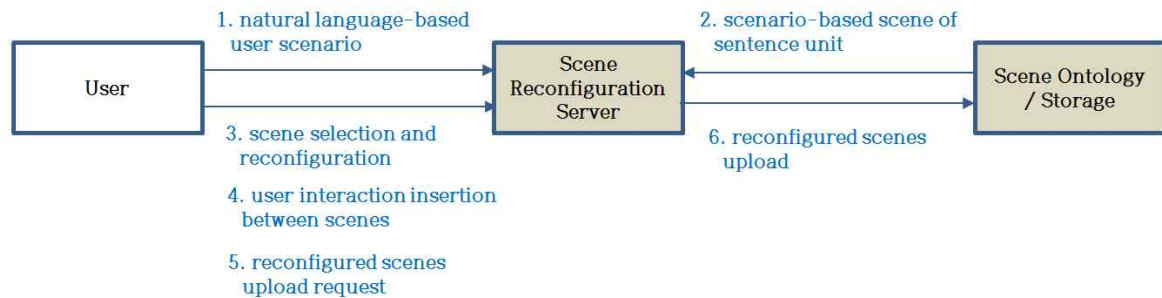


Figure 4. Operating Scenario for Interactive Media Creation

If the user requests a playback of interactive media consisting of 'S1-I4-S3' to the scene reconfiguration server through a variety of cross-platform, it is processed by the following steps:

- The server brings the scene S1 from the storage, and then transmits this scene to the cross-platform
- After streaming of the scene S1 is finished, the server waits for a user interaction I4 input via a cross-platform.
- If user interaction is input, the server brings the scene S3 from the storage, and then transmits this scene to a cross-platform

IV. CONCLUSIONS

NLP, ontology-base searching and multimodal interface technology have been studied for many years. In this paper, we propose a system and method to create interactive media by combining these technologies. The interactive media market is expected to become active in near future. Therefore, a method that anyone can easily create an interactive media is required. In this paper, we propose the overall system architecture and operation method for interactive media creation, and future studies on video segmentation and storage method is needed.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [16ZC2200, Development of programmable interactive media creation service platform based on open scenario]

REFERENCES

- [1] https://en.wikipedia.org/wiki/Interactive_media
- [2] S. Kang, "Korean morphological analysis using syllable information and multiple-word units" Ph.D. thesis, Department of Computer Engineering, Seoul National University. 1993. (written in Korean)

- [3] Brill, "A simple rule-based part-of-speech tagger", In Proceedings of the conference on applied natural language processing, 1992.
- [4] J. Son, H. Yoon, S. Park, K. Cho, and W. Ryu, "Consolidation of Subtasks for Target Task in Pipelined NLP Model", ETRI Journal, vol. 36, no. 5, pp. 714-720, 2014.
- [5] <https://en.wikipedia.org/wiki/Ontology>
- [6] C. Wang, M. Xiong, Q. Zhou, Y. Yu, "Panto: A portable natural language interface to ontologies", In: The SemanticWeb: Research and Applications, pp. 473-487, Springer, 2007.
- [7] Jinwoo Kim, Jae Hong Ryu, and Tae Man Han, "Multimodal Interface Based on Novel HMI UI/UX for In-Vehicle Infotainment System," ETRI Journal, vol. 37, no. 4, Aug. 2015, pp. 793-803.



Minho Han received his BS and MS degree in computer engineering from Chungnam National University, Daejeon, Rep. of Korea in 2001. He joined ETRI, Rep. of Korea in 2000, where he is currently senior researcher. His research interests include natural language processing and information security.



Wonjoo Park received her MS degrees in information and communication engineering from Chungnam National University, Daejeon, Rep. of Korea in 2000. She joined ETRI, Rep. of Korea in 2000, where she is currently senior researcher. Her research interests include data mining, topic model, and ontology



Jeong-Woo Son received his MS and Ph.D. degrees in computer engineering from Kyungpook National University, Daegu, Rep. of Korea in 2007 and 2012 respectively. Since 2013, he has been with ETRI, Daejeon, Rep. of Korea. He focuses on machine learning, NLP, and information retrieval.



Sun-Joong Kim received her BS degree in computational statistics and her MS degree in computer science from Chungnam National University, Daejeon, Rep. of Korea, in 1989 and 2000 respectively. In February 1989, she joined ETRI, Daejeon, Rep. of Korea, where she is currently principal researcher and director. Her research interests include convergence service control, smart TV, content knowledge mining.