

스토리 정보의 검색을 위한 웹 기반의 협업적 스크립트 분석 플랫폼 개발

박승보*, 김현식**, 백영태***, 유은순*

Development of Collaborative Script Analysis Platform Based on Web for Information Retrieval Related to Story

Seung-Bo Park*, Hyun-Sik Kim**, Yeong-Tae Baek***, Eun-Soon You*

요 약

영화의 스토리에 대한 분석은 영화의 설계도와 같은 스크립트를 분석하여 용이하게 이루어질 수 있다. 영화의 스크립트는 파이널 드래프트(Final Draft)의 형식화된 구조로 기술되어 있지만, 웹상에 공개된 스크립트들은 대부분 형식이 파괴되어 문장 내용에 대한 분석이 없으면 형식을 복원하기 힘든 상황이다. 이를 위해 자동으로 스크립트를 파싱한 후에 사용자가 협업적으로 결과물의 오류를 자유롭게 점검하고 수정할 수 있도록 웹기반의 스크립트 분석 소프트웨어를 개발하고 제공할 필요가 있다. 따라서 본 논문에서는 사용자에 의해 스크립트의 형식 오류가 수정되고 걸러지게 하여 완성도 높은 영화 데이터 축적을 가능케 하는 웹 기반의 협업적 스크립트 분석 플랫폼의 구조를 제시하고, 구현 결과에 대한 성능평가를 진행하였다. 실험을 통해 자동 파싱 성공률이 64.95%로 나타났고, 협업적 오류 수정은 5단계를 거친 후에 대부분의 오류가 수정되어 99.58%의 파싱 정확도를 보여주었다.

▶ Keywords : 스토리, 씨네메트릭스, 스크립트, 협업, 파서

Abstract

Movie stories can be retrieved efficiently by analyzing a script, which is a blueprint of the movie. Although the movie script is described in the formatted structure of Final Draft, it is hard to restore the type without analyzing the story of the sentences since the scripts open on the website are mostly broken. For this purpose, it is necessary to develop and provide the web-based script analysis software so that users collaboratively and freely check and correct the errors in the

•제1저자 : 박승보 •교신저자 : 유은순

•투고일 : 2014. 8. 19, 심사일 : 2014. 8. 25, 게재확정일 : 2014. 9. 3.

* 단국대학교 미디어콘텐츠연구원(Institute of Media Content, Dankook University)

** 단국대학교 (Dankook University)

*** 김포대학교 멀티미디어과(Dept. of Multimedia, Kimpo University)

※ 이 논문은 2014년 한국컴퓨터정보학회 제50차 하계학술대회에서 발표한 논문("스토리 차원의 Cinemetrics를 위한 협업적 스크립트 분석 플랫폼(14)")을 확장한 것임

results after automatically parsing the script. Hence, in this paper we suggest the structure of the web-based collaborative script analysis platform that enables users to modify and filter the type error of the script for high level of film data accumulation and performance evaluation for the implementation results is conducted. Through the experiment, accuracy of automatically parsing appears to be 64.95% and performance of modification by collaboration showed 99.58% of accuracy of parsing with errors mostly corrected after passing through 5 steps of modification.

▶ Keywords : Story, Cinemetrics, Script, Collaboration, Parser

I. 서론

영화에 존재하는 다양한 정보들을 측정하고 시각화 또는 검색하는 분야를 씨네메트릭스(cinemetrics)라 한다[12]. 씨네메트릭스의 지금까지의 연구는 영화의 영상신호나 썸트 길이와 같은 기초적인 정보를 이용한 형태의 연구들이 주를 이루었다[1,2]. 하지만 영화는 스토리를 가진 미디어로 스토리 차원의 정보 검색을 위해서는 영화의 다양한 형태의 데이터들을 분석할 필요가 있다.

스크립트는 영화의 설계도와 같은 것으로 그 속에 장면, 등장인물, 대사, 액션 등과 같은 영화를 제작하기에 필요한 다양한 정보가 존재한다. 따라서 영화 분석 또는 정보검색을 위해서는 영화의 설계도와 같은 스크립트에 대한 데이터 수집과 분석이 필요하다.

전세계의 영화시장을 대부분 차지하고 있는 할리우드 영화는 제작을 위한 공동 작업을 위해 공인된 파이널 드래프트(Final Draft) 형식으로 스크립트를 작성하고 공유한다[3,4]. 파이널 드래프트 형식으로 표현된 스크립트를 활용하여 영화의 다양한 정보를 추출하여 활용한 다양한 연구들이 존재한다[5,6]. 스크립트를 활용하여 영화 정보를 검색하는 기존의 연구들은 영화의 스크립트가 오류가 없다고 가정하고 스크립트를 분석하여 영화 정보를 추출하고 활용한다[5,6].

하지만 웹에 공개된 스크립트들은 기본적으로 파이널 드래프트 형식을 지키고 있지만, 스크립트 작성 소프트웨어의 종류나 작가의 표현방식에 따라서 저장된 형식에 차이가 발생한다. 또한 파이널 드래프트 형식의 PDF 파일을 텍스트로 변환하면서 특정 요소들이 형식에서 이탈되는 경우가 발생한다. 따라서 스크립트를 이용한 영화 분석 또는 정보 검색을 위해

서는 웹상에 존재하는 스크립트의 손상된 오류들을 개선하는 전처리 작업들이 선행되어야 한다. 즉 오류가 발생한 부분들을 모두 보정할 수 있는 스크립트를 파싱하고 분석하는 소프트웨어가 필요하다. 하지만 이 모든 경우를 보장할 수 있는 스크립트 파서를 설계하고 제작하는 것은 매우 어려운 일이다. 이보다는 다양한 사용자들이 스크립트를 협업적으로 분석하는 플랫폼을 개발하여 사용자에 의해 스크립트 데이터를 분석하고 활용하도록 하는 시스템을 개발할 필요가 있다.

따라서 사용자에 의해 오류가 손쉽게 수정되어 완성도 높은 영화 데이터 축적을 가능케 하는 협업적 스크립트 분석 플랫폼을 설계하고 개발하고자 하는 것이 본 연구의 목적이다.

이를 위해 2장에서 씨네메트릭스와 협업적 영화 분석과 관련된 연구들을 소개하고, 3장에서 본 연구에서 제안하는 협업적 스크립트 분석 플랫폼에 대해 기술한다. 또한 4장에서 협업적 스크립트 분석 플랫폼에 대한 구현 결과를 설명한 후에 5장을 통해 본 연구의 결론을 맺는다.

II. 관련 연구

씨네메트릭스는 영화를 측정하기 위한 데이터를 가공하거나 시각화 해주는 연구 분야이다. 씨네메트릭스를 위해서는 영화의 데이터 분석을 통해 영화의 변화 그래프 또는 정형화된 데이터를 표현하여 영화의 측정 데이터로 활용하는 연구가 필요하다[1,2,9,14].

Brodbeck은 영화의 썸트에서 썸트 길이, 칼라의 평균값, 음성, 프레임들의 색정보 등을 추출하여 썸트의 특징 값을 활용한 연구를 진행하였다. 원형으로 썸트를 누적인 후 프레임 내의 움직임량을 추출하여 애니메이션 되도록 하여 그림 1과 같이 영화의 지문(Fingerprint)로 생성하였다[1,9]. 지문은

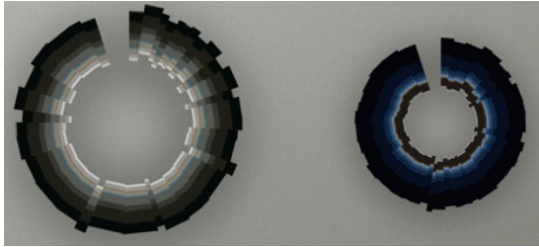


그림 1. 씨네메트릭스의 지문(영화 'Solaris'의 (좌)1972년판 (우)2002년 판의 지문)

Fig. 1. Fingerprints for Cinemetrics (Movie 'Solaris' Edition of 1972 (Left) and 2002 (Right))

영화마다 고유의 형태를 갖는다. Solaris 영화인 그림 1의 경우 샷들의 시간이 짧고 프레임 간의 움직임이 많기 때문에 2002년의 Solaris 리메이크 영화에서 펄스의 두께가 얇고 빠른 펄스의 움직임이 나타난다[9]. 이 기술은 한눈에 영화 간의 비교가 가능토록 시각화했다는 데는 의미가 있으나 영화 특징들 간의 관계가 고려되지 않아서 스토리 차원의 시각화나 정보 검색에 활용되기에는 한계가 있다[9].

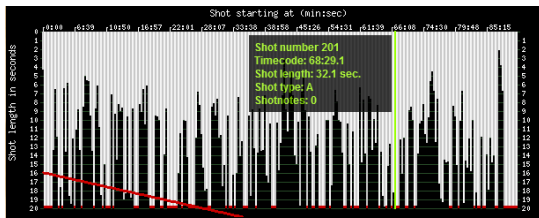


그림 2. 영화 'Camp 14: Total Control Zone'의 샷 길이 그래프
Fig. 2. Shot Length Graph of Movie 'Camp 14: Total Control Zone'

Nothelfer 등은 영화에서 시네메트릭스를 위해 하나의 영화에 대해 샷들의 길이를 측정한 후에 시간 순서대로 나열하는 연구를 진행하였다. 그들은 그림 2와 같은 그래프를 완성하여 영화의 특징으로 활용하였다[2]. 이 연구는 현재에도 지속적으로 연구가 진행되고 있으며, 사용자들이 샷들의 길이를 파악하는 소프트웨어를 다운로드 받아서 직접 분석한 결과를 웹사이트에 공유할 수 있도록 하였다. 그리고 샷들의 개수와 샷의 길이 등을 영화의 특징 요소로 추출한 후에 감독별, 시대별 또는 장르별로 영화의 변화 추이나 차이를 분석하는 연구를 진행하였다. 하지만 샷의 길이와 형태는 영화의 편집 유형 변화나 스토리 전개 속도를 파악하는 데는 중요한 요소이지만, 관객이 인식하는 영화 스토리의 내용을 파악하기에는 부족하다[9].

2. 파이널 드래프트

영화 스크립트는 영화의 스토리 진행과 촬영, 편집에 필요한 배경, 시간, 등장인물의 행동, 대사, 카메라 효과에 대한 요소들을 설명하는 문서이다. 영화 스크립트를 작성하는 세계적인 표준 형식은 파이널 드래프트 형식이다. 이 형식은 파이널 드래프트라는 스크립트 작성 저작도구의 저장양식으로 개발되었다가, 미국의 영화 시나리오 작성의 표준으로 활용되고 있다. 현재는 국제적으로 이 형식을 활용하거나 유사한 형태의 스크립트 작성 형식을 사용하고 있다[9].

```

INT. HOUSE - STAIRWAY - NIGHT

Edward descends the long, curved stairway. He has Vance in tow.

                EDWARD
            (all business)
            Prepare the filing papers. Have them ready by the middle of the week. I'll tell you when to submit them.

The PEOPLE talking and drinking on the stairway gravitate toward Edward.

                A MAN
            Edward!

                EDWARD
            (turning on the charm)
            Yes!

                THE MAN
            Will we see you at the polo matches Wednesday?

                EDWARD
            Wouldn't miss it.
            (abruptly; to Vance)
            I want this done fast and easy. No slip-ups. I want to be back in New York by Saturday. Got it?

He deftly keeps making his way through the crowd.
    
```

그림 3. 파이널 드래프트 형태의 영화 시나리오의 예
Fig. 3. A Example of Script of Type of 파이널 드래프트

영화 스크립트는 일반적으로 'title', 'scene heading' (장면 머리글), 'action', 'character name' (등장인물명), 'dialogue' (대사), 'parenthetical', 'extension', 'transition', 'shot' 등으로 구성된다[5,6,9,13]. 그림 3은 'scene heading', 'action', 'character name', 'dialogue', 'parenthetical'으로 구성된 영화 'Pretty Woman' 시나리오의 일부분이다. 'scene heading'은 'INT' 또는 'EXT'로 시작하며, 배경 장소와 시간을 지시한다. 'action'은 장면 상황을 기술 하거나, 등장인물의 상태

를 설명한다. 'character name'은 등장인물의 이름으로 대문자와 숫자와 일부 특수문자만을 사용하여 표시한다. 'character name'에 이어서 등장인물이 말하는 대사인 'dialogue'가 바로 나타난다. 'dialogue' 중간에는 'parenthetical'이 나타나기도 하는데, 이것은 등장인물의 심리상태나 상황을 짧게 지시하는 부분으로 괄호로 둘러싸여 있다[9]. 스크립트의 각 요소들은 각각의 들여쓰기를 적용 받는다.

3. 협업적 영화 분석 플랫폼

영화 정보의 수집부터 분석까지 다양한 정보를 협업적으로 처리하는 분석 플랫폼에 대한 연구들은 이전부터 꾸준히 나타나고 있다[2,7,8,11].

Nothelfer 등은 사용자가 직접 영화를 정하여 영화의 쏫의 길이를 측정하여 업로드 하는 정보 수집 방법을 제안하고 활용하였다[2]. 이를 위해 데스크탑 버전의 쏫 길이를 추출하는 소프트웨어를 개발하여 배포하였다.

IMDb는 사용자가 영화에 대한 리뷰와 사진들을 올리고 태깅을 하는 협업적 정보 수집 및 편집의 활동을 진행하여 영화 정보를 구축하였다[7]. 하지만 영화에 대한 개략적인 정보를 수집하기 때문에 대사나 배경과 같은 상세한 내용에 대한 정보 검색의 토대가 되는 데이터로 활용하기는 어려운 문제점이 있다.

Wikipedia는 사용자가 영화와 관련된 데이터들을 직접 작성하고 수정하여 협업적으로 영화 정보를 구축하였다[8]. 하지만 영화에 대한 개략적인 정보를 수집/분석하기 때문에 영화의 상세한 내용에 대한 정보 검색의 토대가 되는 데이터로 활용하기는 어려운 단점이 존재한다.

따라서 영화의 실제 동영상 정보로부터 정보를 추출하여 활용하거나, 스크립트를 협업적으로 분석하여 영화의 상세한 데이터로 구축하여 정보 검색의 토대가 되도록 활용할 필요가 있다. 특히 본 논문은 기술적 난이도로 인해 정확도가 떨어지는 동영상 분석보다는 스크립트를 기반의 텍스트 분석을 통한 협업적 스크립트 분석 플랫폼에 대한 연구를 진행한다.

III. 웹 기반의 협업적 스크립트 분석 플랫폼

웹 기반의 협업적 스크립트 분석 플랫폼은 영화의 정보를 활용하기 위해 스크립트를 자동으로 파싱하는 기능과 사용자가 협업적으로 파싱된 결과를 수정할 수 있는 기능을 웹 상에서 지원한다. 이 두 기능을 통해 영화 스크립트로부터 정제된 영화 정보들을 추출하여 분석하게 된다. 또한 웹 상에서

파싱과 수정 작업을 지원하므로 인해 누구나 쉽게 오류를 수정하고 정제된 스크립트 데이터를 활용할 수 있도록 하였다.

스크립트의 자동 파싱을 통해 장면(scene heading), 등장인물(character name), 대사(dialogue) 등을 계층적으로 구조화하여 저장하는 기능을 제공하며, 오류가 확인될 경우 계층적으로 구조화된 스크립트를 사용자가 Script 원본과 비교하면서 오류를 수정할 수 있는 기능을 제공한다.

웹 기반의 협업적 스크립트 분석 플랫폼은 그림 4와 같이 클라이언트(Client)의 UI(User Interface)와 서버단에 존재하는 Automatic Script Parsing Part(스크립트 자동 파싱부)와 Collaborative Script Analyzing Part(협업적 스크립트 분석부)로 구성된다. UI는 클라이언트 단에서 동작하는 모듈로써 영화 스크립트를 업로드하고, 파서를 동작시키는 명령과 스크립트 오류를 수정하는 인터페이스를 제공한다.

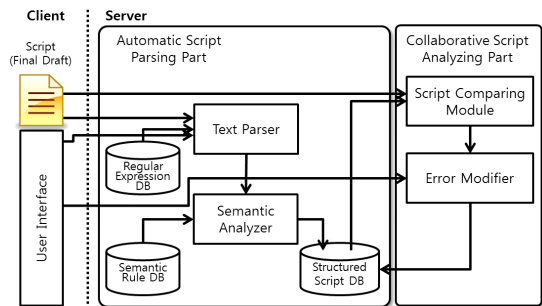


그림 4. 웹 기반의 협업적 스크립트 분석을 위한 시스템 구조
Fig. 4. System Architecture of Collaborative Script Analyzer Platform Based on Web

1. 스크립트 자동 파싱부

이 부분은 스크립트 파싱을 자동화하여 사용자에게 제공하는 부분이다. 입력은 클라이언트로부터 제공되는 파일형식의 영화의 텍스트 스크립트 파일이 활용되며, 출력된 결과는 Structured Script DB에 저장된다. Automatic Script Parsing Part는 Text Parser, Semantic Analyzer, Structured Script DB의 3개의 컴포넌트로 이루어진다.

Text Parser는 Text Parser와 Regular Expression DB로 이루어진다. Text Parser는 파일형식의 스크립트 화일로부터 장면(scene heading), 등장인물(character name), 대사(dialogue) 등을 추출하며, Regular Expression DB는 영화 스크립트를 파싱하기 위해 장면과 등장인물, 대사 등을 추출할 수 있는 표 1과 같은 정규 표현식을 테이블 형태로 저장하고 있다.

표 1. 스크립트 파싱을 위한 정규 표현식
Table 1. Regular Expressions for Script Parsing

Elements	Regular Expression
SCENEHEADING	<code>^t((EXT INT)\.)*\$</code>
CHARACTER	<code>^t\t\t\t(\w+)(.*)\$</code>
ACTION	<code>^t(?!(:?EXT INT))((^\t)+.*)\$</code>
PARENTHETICAL	<code>^t\t\t\t((^\t)+.*)\$</code>
DIALOGUE	<code>^t\t((^\t)+.*)\$~</code>

Semantic Analyzer는 Semantic Analyzer와 Semantic Rule DB로 이루어진다. Semantic Analyzer는 Text Parser에 의해 추출된 장면, 등장인물, 대사 등의 포함 관계를 판정하여 계층화된 Script의 형태로 변환하는 역할을 하며, Semantic Rule DB는 장면, 등장인물, 대사의 포함관계를 판정하기 위한 Rule들을 저장하고 있다.

Structured Script DB는 파일별 드래프트 형식의 영화 스크립트에 대한 파싱된 결과물을 저장하는 데이터베이스로 계층적으로 장면, 등장인물, 대사가 구조화되어 저장되어 있는 Script DB이다.

2. 협업적 스크립트 분석부

Collaborative Script Analyzing Part는 입력으로 파일별 드래프트 형식의 영화 스크립트 파일과 Structured Script DB가 활용되며, 수정된 결과물이 Structured Script DB로 다시 저장된다.

Collaborative Script Analyzing Part는 Script Comparing Module과 Error Modifier로 이루어져 있다. Script Comparing Module은 파일별 드래프트 형식의 영화 스크립트와 Structured Script DB를 비교하여 판단할 수 있는 뷰어(Viewer)를 사용자에게 제공한다. 또한 Error Modifier는 사용자가 영화 스크립트와 Structured Script DB 사이에 차이나 오류를 발견하였을 경우 수정을 할 수 있는 기능을 제공한다. 수정은 들여쓰기의 정도를 조정하는 기능이다. 그리고 수정된 결과는 Structured Script DB에 다시 저장된다.

IV. 평가 및 고찰

1. 구현 및 운용 환경

제안된 웹 기반의 협업적 스크립트 분석 플랫폼은 그림 5에 보여지는 것처럼 Finaldraft Script Analyzer라는 이름으로 개발되었다. 서버단의 Automatic Script Parsing Part와 Collaborative Script Analyzing Part를 구현하고 클라이언트 단의 UI가 서버단의 기능들을 제어할 수 있도록

구현하였다.

개발은 HTML 5 기반으로 진행하였다. 개발환경은 자바스크립트와 표 2와 같은 라이브러리들을 이용하여 진행하였다.

표 2. 사용된 라이브러리 목록
Table 2. Used Libraries

라이브러리	사용 목적
Twitter Bootstrap	전체적인 화면 UI 드로잉의 구현
AngularJS	웹 앱 내부의 데이터 플로우 처리
CodeMirror	에디터 윈도우와 스크립트 파서
jQuery	DOM 트리 조작
d3js	데이터 시각화 라이브러리
CoffeeScript	커피스크립트(Language) 컴파일러

라이브러리들 중에 CoffeeScript는 개발 단계에서만 사용하는 라이브러리로 실제 파싱을 위한 작업에서는 제거되는 코드이다. 또한 모듈 의존성 관리를 위해 Bower 유틸리티를 사용하였다. HTML5의 최신 기능을 다수 사용하고 있어서 구글(Google)의 크롬(Crome) 브라우저와 모질라(Mozilla)의 파이어폭스(FireFox) 브라우저에서 정상 작동되는 것을 확

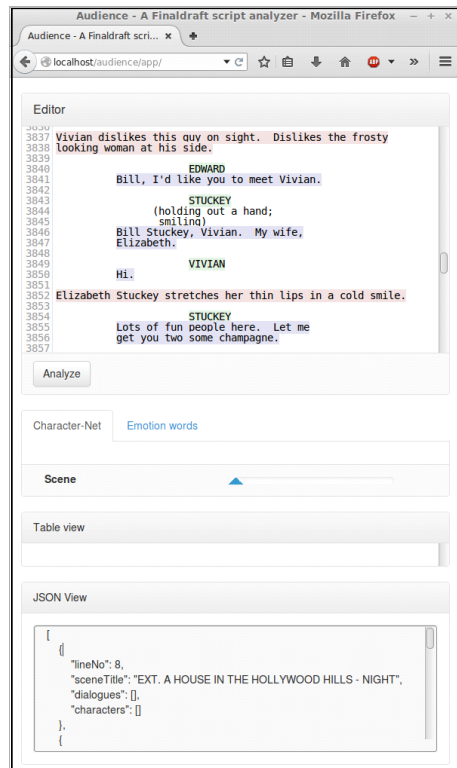


그림 5. 협업적 스크립트 분석 플랫폼을 이용한 영화 "귀여운 여인"의 스크립트 파싱의 예
Fig. 5. A Screen Shot of Collaborative Script Analyzer Platform for Movie "Pretty Woman"

인하였다. 협업적 스크립트 분석 플랫폼의 홈페이지에 연결하면, 서버로부터 HTML 5의 코드를 다운받아서 기능이 동작하도록 하였다. 파싱을 비롯한 대부분의 기능을 클라이언트 단에서 동작하도록 구현하여 사용자가 빠르게 실행하고 실험 결과를 확인하도록 하였다.

전체적인 화면 UI 드로잉의 구현을 위해서는 Twitter Bootstrap을 활용하였다. Automatic Script Parsing Part를 위해 스크립트를 파싱하는 소프트웨어를 개발하고 구조화된 스크립트 결과물을 업로드 하는 절차를 지원하는 웹 모듈을 개발하였다. 파싱된 결과는 그림 5의 밑에 보여지는 것처럼 Json 자료형태로 데이터베이스에 저장된다.

Json 형식은 스크립트의 라인별로 파싱된 결과를 가지고 있게 된다. Json 형식으로 변환되면서 Semantic Analyzer에 의해 장면과 등장인물과 대사 등의 포함관계가 그림 6과 같이 지정되면서 계층적인 관계로 저장되게 된다. 길이로 인해 중간에 일부분의 등장인물과 대사는 생략하였다.

Collaborative Script Analyzing Part는 사용자들이 Automatic Script Parsing Part에 의해 변환된 Structured Script 상의 오류를 수정하여 재 업로드 할 수 있도록 한 부분이다. 이것 역시 그림 5의 Finaldraft Script Analyzer의 일부분으로 그림 5의 Editor 화면에서 사용자가 오류를 수정할 수 있도록 하였다. 저장된 스크립트를 불러오면, Editor에 스크립트가 보여진 후 해당하는 Json 파일이 Json View에 보여지게 된다. 이후에 사용자는 Editor 상에

```

{
  "lineNo": 17,
  "sceneTitle": "INT. HOUSE STUDY - NIGHT",
  "dialogues": [
    {
      "character": "EDWARD",
      "lineNo": 25,
      "dialogue": "Left pocket."
    },
    {
      "character": "INVESTMENT BANKER 1",
      "lineNo": 42,
      "dialogue": "Our banks certainly doesn't want to discour"
    },
    {
      "character": "VANCE",
      "lineNo": 54,
      "dialogue": "Look, we're talking about hard assets and a"
    },
    {
      "character": "STUCKEY",
      "lineNo": 60,
      "dialogue": "This is a #no brainer#. There's no risk"
    },
    <----- Omitted ----->
    {
      "character": "EDWARD",
      "lineNo": 160,
      "dialogue": "Where are you, Cynthia? I see. No, I don't"
    }
  ],
  "characters": [
    "EDWARD",
    "INVESTMENT BANKER 1",
    "VANCE",
    "STUCKEY"
  ]
}
    
```

그림 6. 영화 '귀여운 여인'이 Json 형태로 변환된 예
Fig. 6. A Sample converted to Json Form for Movie "Pretty Woman"

스크립트와 Json View를 비교하면서 발견된 오류에 대해 수정한 후에 즉시 파싱된 결과가 올라온지 확인할 수 있다.

이렇게 Finaldraft Script Analyzer를 통해 사용자들은 원하는 영화의 스크립트를 파싱하고 오류를 수정할 수 있도록 하였으며, 다른 사람이 수정한 스크립트를 검색하여 활용할 수 있도록 하였다.

2. 평가 및 고찰

구현된 웹 기반의 협업적 스크립트 분석 플랫폼의 성능을 평가하기 위하여 표 3과 같은 로맨틱 코미디 장르의 영화 10편을 선정하여 실험을 진행하였다. 영화의 스크립트는 IMSDb (The Internet Movie Script Database)[10]에서 수집하였다.

표 3. 사용된 영화 스크립트 목록
Table 3. Movie Script List

ID	영화명	장르
M01	Pretty Woman	Comedy/Romance
M02	As Good As It Gets	Comedy/Drama/Romance
M03	The Back-Up Plan	Comedy/Romance
M04	Crazy, Stupid, Love	Comedy/Drama/Romance
M05	It's a Wonderful Life	Drama/Romance/Fantasy
M06	The Proposal	Comedy/Drama/Romance
M07	The Ugly truth	Comedy/Romance
M08	500 Days of Summer	Comedy/Drama/Romance
M09	The Bounty Hunter	Action/Comedy/Romance
M10	Larry Crowne	Comedy/Drama/Romance

성능평가는 Automatic Script Parsing Part에 의한 파싱 정확도과, 사용자들의 참여에 의한 협업적 오류 수정 후에 나타나는 파싱 정확도의 변화로 실시하였다. 이 결과는 그림 7, 8과 같이 나타났다. 그림 7은 Automatic Script Parsing Part에 의한 스크립트 자동 파싱 정확도로서 평균적으로 64.95%로 나타났다. M03, M06, M10의 경우 대사

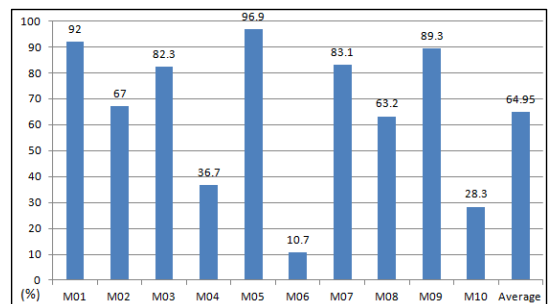


그림 7. 자동 파싱 정확도
Fig. 7. Accuracy by Automatic Script Parsing

의 들여쓰기가 대부분 손상된 경우로 파싱 성공률이 매우 낮게 나타났다.

웹 상에 존재하는 스크립트 파일들은 원본인 PDF 화일을 텍스트 형태의 스크립트 파일로 변환하면서 들여쓰기 형식이 손상되면서 오류가 많이 나타나는 것으로 보인다. 오류가 많이 나타나는 부분은 대사의 들여쓰기가 'action'이나 'scene heading'과 동일하거나, 'scene heading' 앞에 장면번호가 있거나 페이지 번호가 나타난 경우이다. 들여쓰기가 손상된 경우 대사와 action 간의 차이를 컴퓨터가 자동으로 구분하기 위해서는 문장의 의미를 분석하여 대화인지 상황설명인지를 판단하여야 한다. 하지만 현재의 컴퓨터 기술로는 해결할 수 없는 작업으로써 사람이 직접적으로 판단하는 방법밖에 없다.

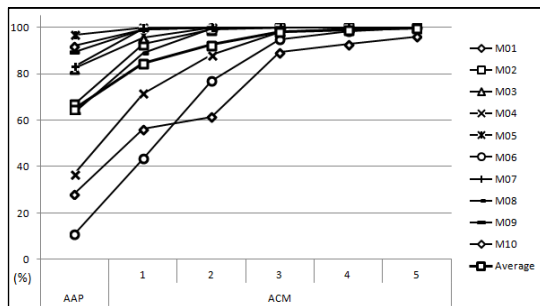


그림 8. 협업적 스크립트 파싱의 단계별 정확도
Fig. 8. Accuracies According to Steps of Collaborative Script Parsing

자동 파싱의 결과물을 사용자들의 참여에 의한 오류 수정을 협업적으로 수행할 경우 대부분 5단계를 거치면 그림 8과 같이 99.58%의 스크립트 파싱 정확도에 도달하였다. 단계는 영화의 스크립트에 대한 사용자들의 오류수정 참여 회수를 의미한다. 3단계가 넘어갈 경우 대부분의 스크립트는 98% 이상의 정확도를 나타낼 정도로 대부분의 오류가 수정되었다. 하지만 M06과 M10의 경우 초기의 오류량이 커서 사용자들이 부분적인 오류만 수정하는 형태로 스크립트를 개선하여 정확도의 개선이 느리게 나타났다. 이 현상의 개선을 위해서는 초기 오류가 높은 경우 사용자들이 편리하게 전체적인 오류 수정을 실행할 수 있도록 오류 패턴을 일괄 수정할 수 있는 매크로 기능이 도입될 필요가 있을 것이다. 스크립트에서 나타나는 오류는 유사한 오류 패턴의 연속으로 매크로 기능이 있을 경우 오류 수정이 용이해 질 것으로 예상된다.

V. 결론

본 논문은 영화 데이터 축적을 가능케 하는 웹 기반의 협업적 스크립트 분석 플랫폼을 제안하고, 설계한 내용과 개발된 결과물에 대한 성능평가를 진행하였다.

웹 기반의 협업적 스크립트 분석 플랫폼은 클라이언트-서버 형태를 갖는다. 세부적으로는 UI, Automatic Script Parsing Part와 Collaborative Script Analyzing Part의 3부분으로 구성된다. 플랫폼을 통해 사용자들은 원하는 스크립트를 검색하고 분석할 수 있도록 하였고, 사용자에게 의해 오류가 수정되고 하여 완성도 높은 영화 데이터가 축적될 수 있도록 하였다. 실험을 통해 Automatic Parsing Part를 통해 이루어지는 자동 파싱 성공률이 64.95%로 나타났다. 또한 Collaborative Script Analyzing Part를 통한 협업적 오류 수정은 5단계를 거친 후에 대부분의 오류가 수정되어 99.58%의 파싱 정확도를 보여주었다. 3단계를 거칠 경우 98% 이상의 정확도를 보여주었지만 초기에 오류가 높을 경우 사용자들이 부분적인 오류 수정 후 업데이트를 실행하여 5단계에 도달하여야 대부분의 오류가 수정되는 결과가 나타났다. 초기 오류가 높은 경우 사용자들이 일괄적으로 오류를 수정할 수 있는 매크로 기능이 도입될 필요가 있을 것이다. 향후에는 웹 기반의 스크립트 분석 플랫폼의 사용 편리성의 증대를 위해 다수의 영화 스크립트에 대해 실험을 진행하여 사용자가 어려움을 갖는 부분들에 대한 파악을 통해 기능개선을 진행할 예정이다.

사사

이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2013R1A1A2061737)

참고문헌

[1] Y. Tsivian, "Cinematics, Part of the Humanities' Cyberinfrastructure," Digital Tools in Media Studies: Analysis and Research: an Overview, Bielefeld: Transcript Verlag, pp. 93-100, 2009.
[2] C. E. Nothelfer, J. E. DeLong, and J. E. Cutting,

"Shot Structure in Hollywood Film," Indiana Undergraduate Journal of Cognitive Science, Vol. 4, pp. 103-113, 2009.

[3] S.-B. Park, H.-N. Kim, H. Kim, and G.-S. Jo, "Exploiting Script-Subtitles Alignment to Scene Boundary Detection in Movie," Proceeding of IEEE International Symposium on Multimedia (ISM2010), pp. 49-56, Dec. 2010.

[4] T. Cour, C. Jordan, E. Mitsakaki, and B. Taskar, "Movie/script: Alignment and parsing of video and text transcription," Computer Vision - ECCV 2008: 10th European Conference on Computer Vision, Part IV, pp. 158-171, Oct. 2008.

[5] [http://en.wikipedia.org/wiki/Final_Draft_\(software\)](http://en.wikipedia.org/wiki/Final_Draft_(software))

[6] <http://www.screenwriting.info>

[7] <http://www.imdb.com/>

[8] http://en.wikipedia.org/wiki/Main_Page

[9] S.-B. Park, Y. T. Baek, "A Study of Story Visualization Based on Variation of Characters Relationship by Time," Journal of The Korea Society of Computer and Information, Vol. 18, No. 3, pp. 119-126, Mar. 2013.

[10] <http://www.imsdb.com/>

[11] S.-B. Park, E.-S. You, "Story-based Information Retrieval," Journal of Intelligent Information System, Vol. 19, No. 4, pp. 81-96, Dec. 2013.

[12] S.-B. Park, E. You, and J.J. Jung, "A cinemetric approach to sentimental processing on story-oriented contents," Quality & Quantity, Vol. 48, no 1, pp. 49-62, Jan. 2014.

[13] S.-B. Park, G.-S. Jo, "Role Grades Classification and Community Clustering at Character-net," Journal of The Korea Society of Computer and Information, Vol. 14, No 11, pp. 169-178, 2009.

[14] S.-B. Park, M.-S. Hong, H. Kim, Y. T. Baek, E.-S. You, "Collaborative Script Analysis Platform for Cinemetrics Based on Story," Proceedings of The Korea Society of Computer and Information Summer 2014, pp. 29-32, July 2014.

저 자 소 개



박 승 보

1995 : 인하대학교
전기공학과 공학사
1997 : 인하대학교
전기공학과 공학석사
2011 : 인하대학교
정보공학과 공학박사
현 재 : 중앙대학교
컴퓨터공학부 전임연구원
관심분야 : 멀티미디어 정보검색,
스토리공학, 감성 검색
Email : molaal@naver.com



김 현 식

2009: 인하대학교
컴퓨터정보공학 공학사.
2011: 인하대학교
정보공학과 공학석사.
현 재: 중앙대학교 연구원
관심분야: 스토리공학
Email : wbstory@storymate.net



백 영 태

1989 : 인하대학교
전자계산학과 이학사
1993 : 인하대학교
전자계산공학과 공학석사
2002 : 인하대학교
전자계산공학과 공학박사
1993-1998 : 대상정보기술(주)
정보통신연구소 선임연구원
1998-현 재 : 김포대학교
멀티미디어과 부교수
관심분야 : 멀티미디어 정보검색,
웹교육시스템, 모바일시스템
Email : hannaek@kimpo.ac.kr



유 은 순

1995년: 인하대학교 불어불문학과 학사

2000년: 프랑스 브장송대학교 언어학 석사

2007년: 프랑스 브장송대학교 언어학 박사

2011년 - 2012년: 단국대학교

미디어콘텐츠연구원

전임연구원

2012년~현재: 단국대학교

미디어콘텐츠연구원

리서치 펠로우

관심분야 : Digital storytelling, Ontology,

Machine Translation,

Social Media, Emotion

Email : tesniere@naver.com