# Breakdown film script using parsing algorithm

**Agung Wahana[1], Diena Rauda Ramdania[2], Dhanis Al Ghifari[3], Ichsan Taufik[4],**
**Faiz M. Kaffah[5], Yana Aditia Gerhana[6]**
[1-6]Department of Informatics, Faculty of Science and Technology UIN Sunan Gunung Djati Bandung, Indonesia
[5,6]Information and Communication Technology, Asia e University, Malaysia

## Article Info

## ABSTRACT

Breakdown script is a breakdown of the scenario into parts that describe each detail of the scene for shooting. The scenario is broken down into more detailed parts using the parsing algorithm. The film script used is a script in Bahasa Indonesia. The process starts from the film script file/scenario in FBX format uploaded to the website then is solved using a parsing algorithm into film elements such as cast members, extras, props, costumes, makeup, vehicles, stunts, special effects, music and sound. The results of this breakdown into sheets according to film elements. The purpose of this research is to produce breakdown sheets from film scripts according to film elements. The parsing algorithm test results showed the correct results of 12 scenes out of 19 scenes.

*Corresponding Author:*

Agung Wahana,
Department of Informatics, Faculty of Science and Technology,
UIN Sunan Gunung Djati Bandung,
105 A. H. Nasution St., Bandung, West Java, Indonesia.
Email: wahana.agung@uinsgd.ac.id

## 1. INTRODUCTION

The film is an enjoyable entertainment facility for the community and is a media that is liked by almost all members of society to get information and insights. Behind the creation of the film, there is a long process and involves many people from various units, working together and carrying out different tasks and functions. There are three stages of production before starting to make films, namely pre-production, production, and post-production. The pre-production stage usually takes a long time because many things must be prepared such as scripts, breakdown scripts, storyboards, director shots, time of shots, crew lists, cast lists, and budgeting. Next, in the production phase, almost the entire team began working on making films. Finally, in the post-production stage, all filming activities that have been carried out previously completed and reviewed again to minimize errors.

One of the work in the pre-production stage is to make a script breakdown. Breakdown script is a description of a screenplay, a part that describes every detail of the scene for shooting. All things needed for filming purposes are included in the script breakdown so that no problems occur while the shooting process is in progress. In the breakdown script, there is a breakdown sheet that contains all the scenes in the film. One breakdown sheet covers all purposes used in the film. The director does this script breakdown work. The director must do this because the director is considered to have a significant role in filmmaking both pre-productions to post-production and script breakdown is also an essential thing in filmmaking. The director must be creative, thorough, and detailed for the perfection of filmmaking.

The number of breakdown sheets for each scene and the long time for making the script breakdown become a problem in the script breakdown process. This has caused some young filmmakers to skip the script breakdown stage. This stage is critical in organizing every need in every scene in one film. The existence of film competition from the creative industry, especially in the field of a film involving young filmmakers, both students, and students, becomes an opportunity for them to assess how important the script breakdown process is in filmmaking.

The script breakdown is one of the stages that occur in the pre-production stage. Breakdown script is the process of breaking down each scene in a film scenario into a list that contains information about everything needed when shooting. This is done to know all the needs when shooting and make it easier to arrange production schedules and costs. To breakdown the film script, the film scenario must be in the final draft stage or in the final stage, and there are no revisions anymore. A breakdown script, a script breakdown sheet is needed, which contains information about each scene in the film. Every need to take pictures is broken down in each stage in one sheet breakdown sheet.

Parsing is the process of decomposition or breaks up part of an input in the form of vocabulary or files that produce a parsing tree (parsing tree) to carry out the next stage of semantic analysis [1-4]. The parsing tree itself is a connected graph where the root is the initial symbol of grammar, each node in the symbol is non-terminal, and the leaf or leaf is read from left to right is a row of input tokens. This parsing tree serves to describe how to get a string by lowering the variable symbols into a terminal symbol until no symbol has not been replaced. Semantic analysis is checking for errors from a series of instructions contained in the program, one of the roles of semantic analysis is checking the types of variables, for example, existing variables have been previously defined, and those variables are the same type if they want to be operated.

## 2. RESEARCH METHOD

The script breakdown is one of the stages that occur in the pre-production stage. Breakdown script is the process of breaking down each scene in a film scenario into a list that contains information about everything needed when shooting. This is done to know all the needs when shooting and make it easier to arrange production schedules and costs. To breakdown the film script, the film scenario must be in the final draft stage or in the final stage, and there are no revisions anymore. A breakdown script, a script breakdown sheet is needed, which contains information about each scene in the movies. Every need to take pictures is broken down in each scene in one sheet breakdown sheet. The breakdown sheet can be export to pdf format. Figure 1 explains the stages of the research, and the stages begin by uploading files in FBX format to the website, then parsing the results will be displayed on the website then tagging to parts of the script such as characters, dialogs, and scenes. The tagging results are displayed in the form of a breakdown sheet that can be exported in PDF format.
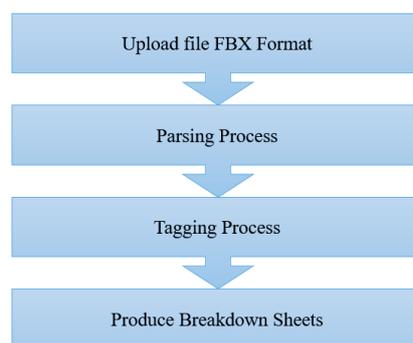
Figure 1. Stages of parsing algorithm breakdown script

### 2.1. Parsing algorithm

In computer science and linguistics, parsing, or, more formally, syntactic analysis, is the process of analyzing text, made of token sequences (for example, words), to determine the grammatical structure of what is given (more or less) formal grammar. Parsing is also an initial term for natural language sentence diagrams and is still used for inflected language diagrams, such as Roman or Latin languages. The term Pars parse comes from the Latin (ōrātiōnis), which means part (speech) [1, 5-9].

## 2.2. Parse tree

A parse tree or a parsing tree or a derivation tree or a concrete syntactic tree is a sorted rooted tree that represents the syntactic structure of a string based on context-free grammar. The term parse tree itself is used primarily in computational linguistics; in theoretical syntax, the term syntactic tree is more general   Parse tree is a rooted tree consisting of the syntactical structure of a word according to formal grammar rules. This tree serves to illustrate how to obtain a string by lowering the symbols of variable symbols into terminal symbols until nothing has been replaced.  Parse tree can be used to represent real-world constructs such as sentences. Tree hierarchy helps understand the order of evaluation for the whole expression. Building a parse tree is breaking the expression string into a token list [10-15].

## 2.2. JSON and XML

JSON (JavaScript object notation) is a lightweight data exchange format, easy to read and write by humans, and easy to translate and generate (generate) by a computer. JSON is made of two structures: a collection of name/value pairs and an ordered list of values [16-20]. XML is a markup language created by the world wide web consortium (W3C) to define syntax to encode documents that can be read by humans and machines. Companies are conducting their daily business by document exchange. Recent technologies allow the companies to package information into documents and route them through web services. XML has been used to describe documents and data in a standardized, textbased format that can be easily transported via web services. The standards of XML allow companies to agree on what their documents mean and facilitate document understanding. XML also allows the companies to combine and compare varieties of documents such as timesheets, balance sheet, income statement, and other types of documents [21-25].

## 3.    RESULTS AND ANALYSIS
### 3.1.   Parsing algorithm analysis
#### 3.1.1. Parsing algorithm process

The process of the parsing algorithm is by uploading the .fdx file (fileFDX_upload). Then the algorithm will read the uploaded file and detect the compatibility of the file with the XML file standard (SendToParse). The process (xmlDoc) determines whether the uploaded file is XML type. If it is of type XML, then the process continues to (ParseToJson. If it's not XML type, it means that the uploaded file is not suitable. In the process (ParseToJson), the XML file will be converted into JSON content.

The process (Json.FinalDraft) detects XML content that has been changed to JSON according to the final draft standard. If according to the standard, then proceed to the process (parseToHtml). If it does not conform to the standards, then the uploaded file does not match. The process (parseToHtml (script Array)) sorts out the contents of the final draft (JSON) according to the index elements. The elements in question are scene headings, actions, characters, parenthetical, dialogue, and general. Figure 2 explains the flowchart of the parsing algorithm, the FBX format file is uploaded and then parsed by the sendToParse class into XML format to validate whether the XML format is valid or not, if it is valid it will be changed in the JSON format; otherwise the process is complete. Valid XML will be changed in the final draft form, and the final draft form is checked whether valid or not if valid will be changed to HTML format if it is not finished.

#### 3.1.2. Parse tree

First, the parse tree divides the script file according to the placement of its elements. The type of elements in a scenario script file consists of scene headings, actions, characters, dialogs, and general. Each element has a derivative in accordance with its structure. Scene Heading is a description of the place, location name, and time. The place described in the scene heading consists of INT (interior) and EXT (exterior). As for the name of the location consists of several nouns or words that represent objects, people or places. In the case above, the name of the location of the uploaded script is "in front of the school gate." In the Big Indonesian Dictionary, those three words are nouns. And time consists of day, afternoon, and night.

Action is the structure of an effort that can be adjusted depending on the action to be parsed. But for structural standards that are following Enhanced Spelling (EYD) are S P O K Pel (Subyek/Subject, Predikat/Predicate, Obyek/Object, Keterangan/Description, Pelengkap/Complement). In the case above, the action of the uploaded script is, "Arya looks at the watch." The EYD structure that corresponds to when the word is "Arya" as a subject, "sees" as a predicate, "watches" as an object. Both of the object words are nouns [2-4]. Character is one of the most important elements in a scenario, and a character is the name of the role played by the actor or actress who plays in the film. The character only requires S (subject) for its derivatives. In this case, the character is Arya. Dialogue is a word or sentence that must be spoken by a character in a scene. The structure of dialogue can be adjusted depending on the dialogue to be parsed. But for structural standards

that EYD are SPOK Pel. EYD structure that corresponds to when the word is "Sir" as a subject, "open" as a predicate, "door gate "as an object. Both of the object words are nouns. Figure 3 explains the process of making a bitter melon tree based on a film script file into more detailed parts such as scene headings, actions, characters, dialogue, and general. Each part is then made in detail.

## 3.2. Testing the parsing algorithm
### 3.2.1. Parsing process
The parsing method Parsing or declining process is parsing or syntactic analysis is the process of analyzing a series of symbols, both in natural language or in computer language, according to formal grammar rules. The parsing method Parsing or declining process is parsing or syntactic analysis is the process of analyzing a series of symbols, both in natural language or in computer language, according to formal grammar rules. Algorithm testing is done by uploading scenarios into the website. Then the algorithm will read the uploaded file and detect whether the file is compatible with the algorithm. The following is the result of the detection process of the parsing algorithm. Figure 4 explains the display of Ike FBX uploads to the website, which shows the dialogue done by the actor/character.
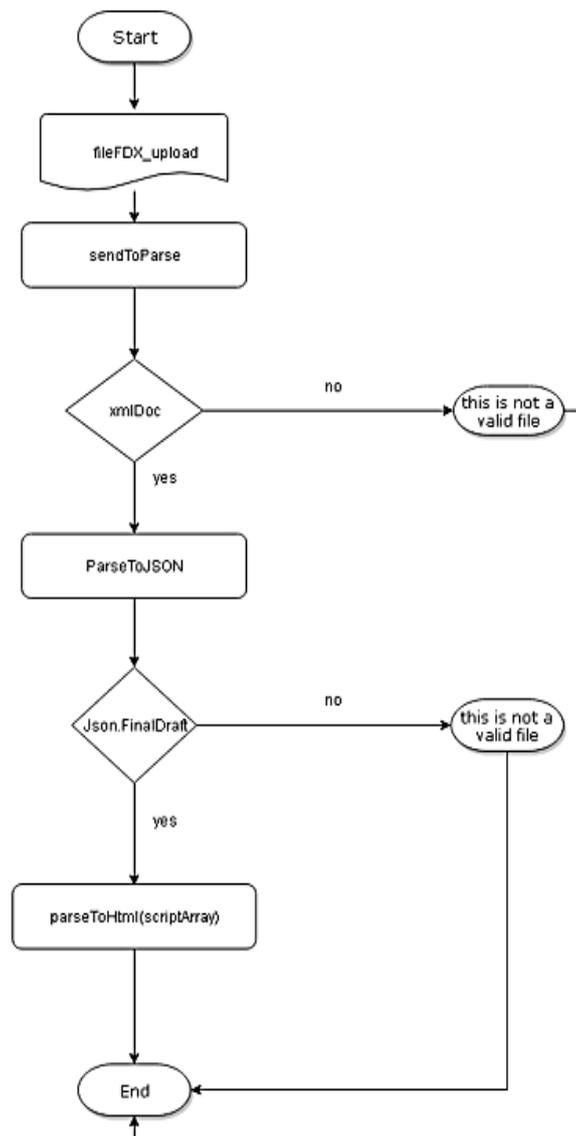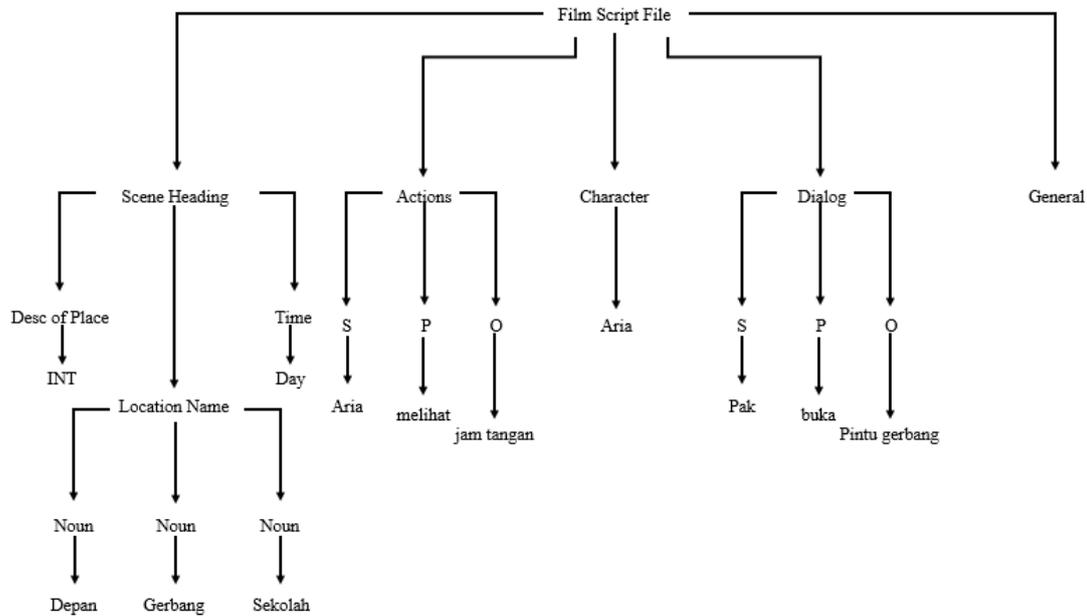
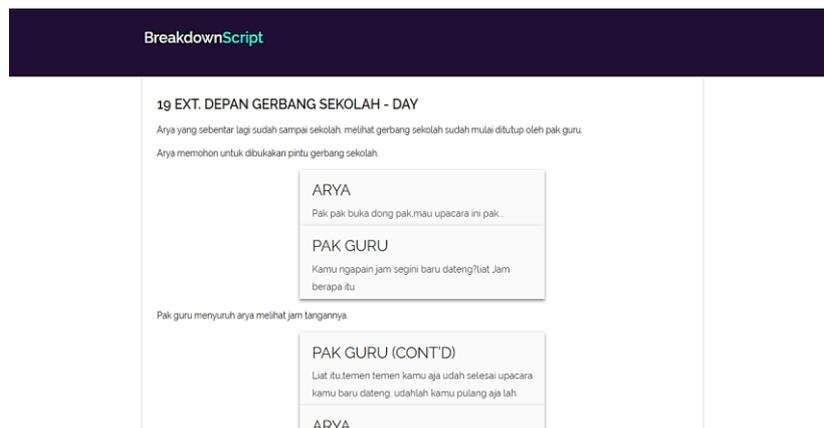Figure 2. Flowchart parsing algorithm

Figure 3. Parse tree



Figure 4. Display result on the website

### 3.2.2. Tagging/labeling process

The data to be observed in the tagging process is a collection of words or sentences in a film script which will then determine what class of words or tags are appropriate. The tagging process begins by considering all possible tag sequences for the sentence. The probability of labeling words using bigram is the probability that a tag appears only depending on the previous tag. The upload results are then tagged in each element based on the uploaded scenario. Figure 5 shows the results of the tagging process which are distinguished from the foreign colors of each element in the script file.

### 3.2.3. Breakdown Sheet

Breakdown Sheet is the process of breaking down each scene in a scene into a list that contains information about everything needed when shooting. This is done to find out the details of shooting requirements, simplify the shooting schedule and the required costs. Breakdown sheet is filled with various information, each department has the right to receive the breakdown sheet for reference in preparing everything during the production process. The division of labor and everything that is done during production also refers to the things listed on the breakdown sheet. The results of the tagging/labeling process can then be made breakdown sheets Figure 6 explains the results of a script file breakdown in the form of a breakdown sheet based on each film element that can be exported to PDF format. Testing of a film script file that managed to break into film elements correctly as many as 12 scenes from 19 scenes.
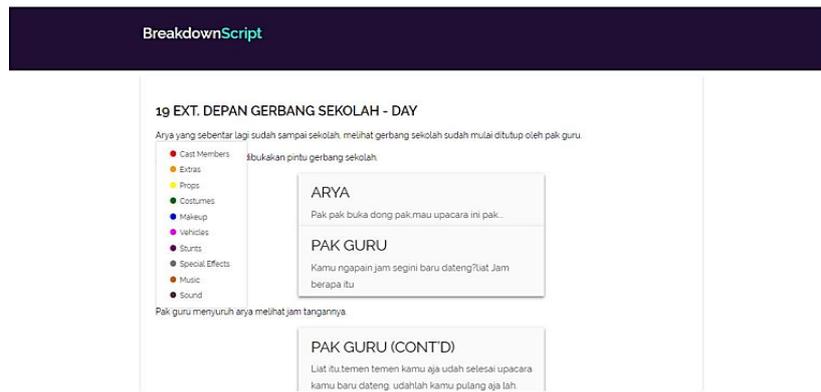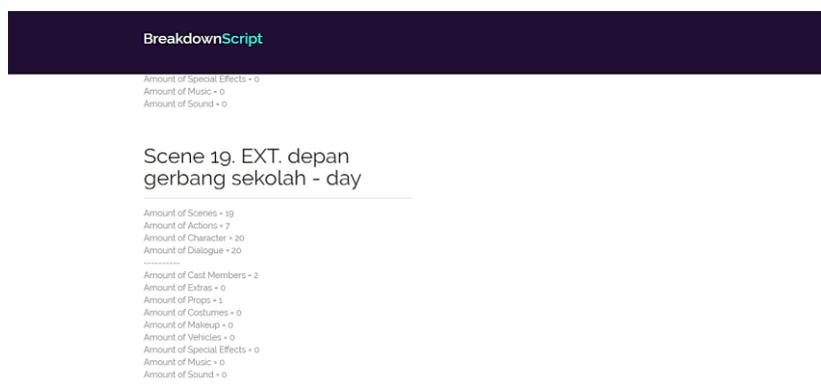
Figure 5. Display result after tagging



Figure 6. Display result breakdown sheet

## 4. CONCLUSION

The parsing algorithm implementation in the movie script breakdown is used when uploading .fdx files to the website to change the file so that it can be uploaded to the website. The parsing algorithm is able to process and divide parts of the film script file into several sections according to the scenario framework criteria consisting of scene headings, actions, characters, parenthical and dialogue which will be displayed after the user uploads the .fdx movie script file. The parsing algorithm is able to label every element in the movie script. The elements in question are cast members, extras, props, costumes, makeup, vehicles, stunts, special effects, music and sound. The parsing algorithm succeeded in doing a script breakdown from a film script file that succeeded in breaking down the film elements as many as 14 scenes from 19 scenes and made break sheets.

## ACKNOWLEDGMENTS

## REFERENCES

[1]  M. Song, R. Zhang, and D. Gang, "HTML tree parsing algorithm based on pre-extracted data," 2009 8th Int. Conf. Mob. Bus., pp. 249–254, 2009.
[2]  E. M. Sibarani, M. Nadial, E. Panggabean, and S. Meryana, "A study of parsing process on natural language processing in bahasa Indonesia," *2013 IEEE 16th International Conference on Computational Science and Engineering*, pp. 309–316, 2013.
[3]  X. Guo, Y. Wen, D. Ma, Y. Jin, and H. Yu, "Research on Scene Parsing Algorithm Cascading Object Detection Network," *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, pp. 459–464, 2017.

[4]    E. E. Luneva, P. I. Banokin, V. S. Zamyatina, and S. V. Ivantsov, "Natural language text parsing for social network user sentiment analysis based on fuzzy sets," *2015 International Conference on Mechanical Engineering, Automation and Control Systems (MEACS)*, 2016.

[5]    Joshua, "What is a PDF Parser? An introduction to PDF and Document Parsing," 2019. [Online]. Available: https://docparser.com/blog/pdf-parser/. Accessed: 30 July 2019.

[6]    G. Endignoux, O. Levillain, and J. Y. Migeon, "Caradoc: A Pragmatic Approach to PDF Parsing and Validation," *2016 IEEE Security and Privacy Workshops (SPW)*, pp. 126–139, 2016.

[7]    A. Bacchelli, A. Cleve, M. Lanza, and A. Mocci, "Extracting structured data from natural language documents with island parsing," *26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011)*, no. 2, pp. 476–479, 2011.

[8]    R. Zhang, L. Lin, G. Wang, M. Wang, and W. Zuo, "Hierarchical Scene Parsing by Weakly Supervised Learning with Image Descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 3, pp. 596–610, 2019.

[9]    S. Wang, Y. Wang, and S. C. Zhu, "Learning Hierarchical Space Tiling for Scene Modeling, Parsing and Attribute Tagging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2478–2491, 2015.

[10]   J. Kort and R. Lämmel, "Parse-tree annotations meet re-engineering concerns," *Proc. - 3rd IEEE Int. Work. Source Code Anal. Manip. SCAM 2003*, pp. 161–170, 2003.

[11]   J. Abrahams, "Code and parse tree for lossless source encoding," *Proceedings. Compression and Complexity of SEQUENCES 1997*, vol. 1, no. 2, pp. 113–146, 2014.

[12]   Y. Zhao and S. C. Zhu, "Scene parsing by integrating function, geometry and appearance models," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3119–3126, 2013.

[13]   S. Jaf, C. Calder, "Deep Learning for Natural Language Parsing," *IEEE Access*, vol. 7, pp. 131363–131373, 2019.

[14]   M. Aman, A. Bin Md Said, S. J. Abdul Kadir, and I. Ullah, "Key concept identification: A sentence parse tree-based technique for candidate feature extraction from unstructured texts," *IEEE Access*, vol. 6, pp. 60403–60413, 2018.

[15]   T. Ehsan and S. Hussain, "Analysis of Experiments on Statistical and Neural Parsing for a Morphologically Rich and Free Word Order Language Urdu," *IEEE Access*, vol. 7, pp. 161776–161793, 2019.

[16]   T. Santos and C. Serrao, "Secure Javascript Object Notation (SecJSON) Enabling granular confidentiality and integrity of JSON documents," *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 329–334, 2017.

[17]   Z. Niu, C. Yang, and Y. Zhang, "A design of cross-terminal web system based on JSON and REST," *2014 IEEE 5th International Conference on Software Engineering and Service Science*, pp. 904–907, 2014.

[18]   B. Lin, Y. Chen, X. Chen, and Y. Yu, "Comparison between JSON and XML in Applications Based on AJAX," *2012 International Conference on Computer Science and Service System*, no. February 1998, pp. 1174–1177, 2012.

[19]   A. A. Abd El-Aziz and A. Kannan, "JSON encryption," *2014 International Conference on Computer Communication and Informatics (ICCCI -2014)*, pp. 1–6, 2014.

[20]   M. Kleppmann and A. R. Beresford, "A Conflict-Free Replicated JSON Datatype," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 10, pp. 2733–2746, 2017.

[21]   M. Klempa et al., "JInfer: A framework for XML schema inference," *The Computer Journal*, vol. 58, no. 1, pp. 134–156, 2015.

[22]   R. D. Varma and G. V. R. Reddy, "Schema Based Parallel XML Parser: A Fast XML Parser Designed for Large XML Files," *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 3, pp. 379–389, 2014.

[23]   C. C. Chiang, "Engineering documents into XML file formats," *Fourth International Conference on Information Technology (ITNG'07)*, pp. 610–615, 2007.

[24]   A. A. Frozza, R. Dos Santos Mello, and F. De Souza da Costa, "An approach for schema extraction of json and extended JSON document collections," *2018 IEEE International Conference on Information Reuse and Integration for Data Science*, pp. 356–363, 2018.

[25]   A. Bauer, M. Braun, and K. R. Muller, "Accurate Maximum-Margin Training for Parsing with Context-Free Grammars," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 1, pp. 44–56, 2017.