

Instantly Check Your Writing

Trusted by millions of students, faculty, and professionals worldwide. Try now

Grammarly

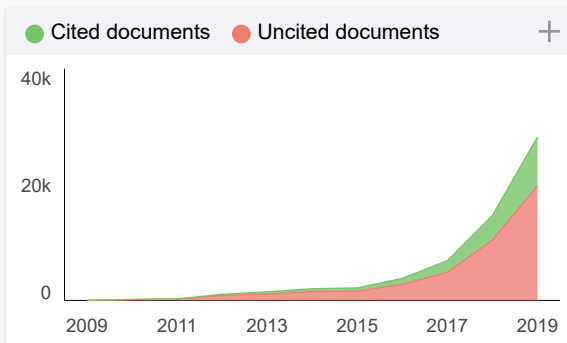
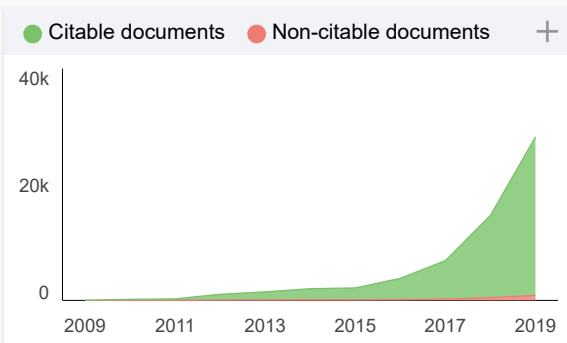
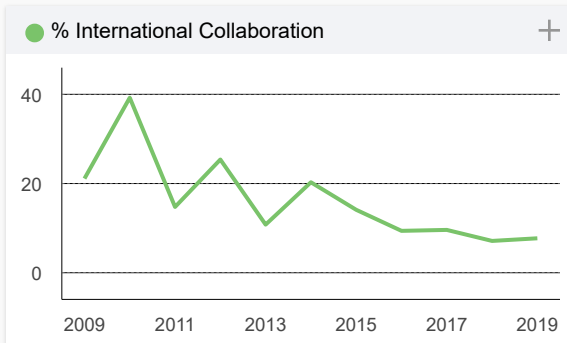
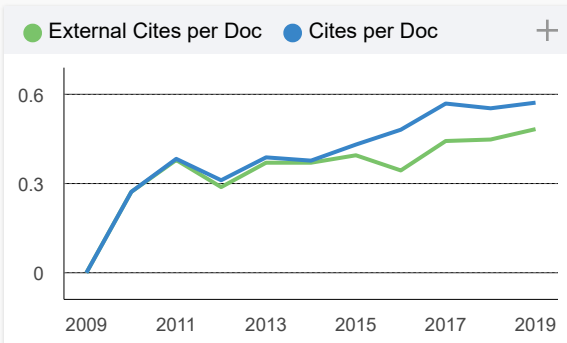
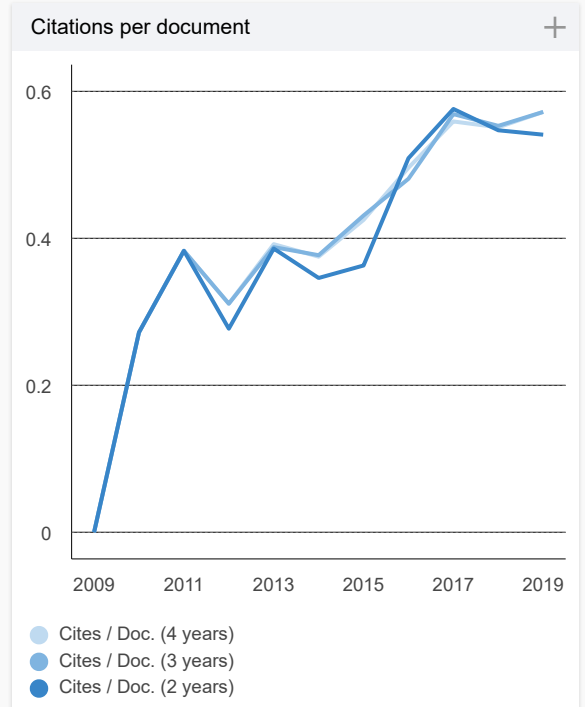
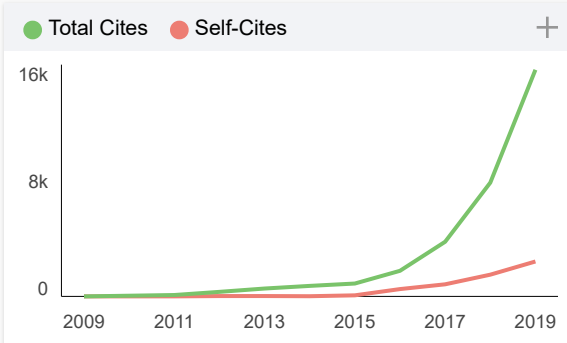
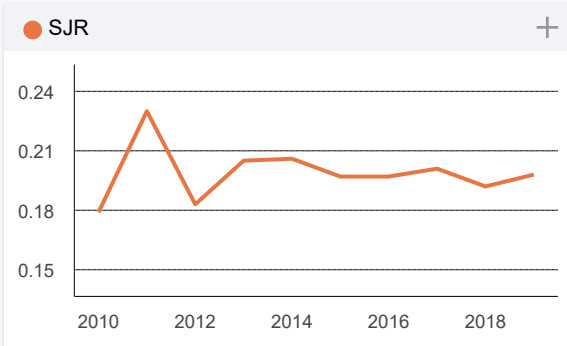
DOWNLOAD

IOP Conference Series: Materials Science and Engineering

31

H Index

Country	United Kingdom - SJR Ranking of United Kingdom
Subject Area and Category	Engineering Engineering (miscellaneous) Materials Science Materials Science (miscellaneous)
Publisher	IOP Publishing Ltd.
Publication type	Conferences and Proceedings
ISSN	17578981, 1757899X
Coverage	2009-2020
Scope	The open access IOP Conference Series provides a fast, versatile and cost-effective proceedings publication service for your conference. Key publishing subject areas include: physics, materials science, environmental science, bioscience, engineering, computational science and mathematics.
	Homepage How to publish in this journal Contact
	Join the conversation about this journal





Jakarta, 16 November 2019

No. : 160-TIM/7440/UNTAR/IX/2019

FULL PAPER ACCEPTANCE NOTIFICATION

Reference Number : TICATE-178

Title : COMMENTS SCRAPING APPLICATION FOR REVIEW YOUTUBE
CONTENT

Authors : Viny Christanti M

Dear Mr./Madam

Thank you for your paper submission to Tarumanagara International Conference on the Application of Technology and Engineering 2019.

We are pleased to inform you that your submission is accepted for presentation in TICATE 2019. Due to some outstanding reviews for other submissions, however, the review result is not available for viewing yet. We expect to publish the review results will begin on 18 November 2019, you need to register to conference before 18th November 2019.

In the meantime, we recommend that you check your manuscript to minimize obvious errors, such as formatting and grammatical errors.

If you have further questions, please don't hesitate to contact us. We are looking forward to seeing you at UNTAR. Thank you very much for your cooperation.

Sincerely,

Dr. Hugeng, S.T.,M.T.(SMIEEE)

Chairman

PAPER • OPEN ACCESS

Preface

To cite this article: 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **852** 011001

View the [article online](#) for updates and enhancements.

2nd Tarumanagara International Conference on the Applications of Technology and Engineering 2019

Preface

On behalf of the organizing committee of 2nd Tarumanagara International Conference on the Applications of Technology and Engineering (TICATE) 2019, I would like to welcome all delegates to Jakarta, Indonesia with great pleasure. Being held from November 21 to 22, 2019 at Campus I- Jl. Letjen. S. Parman No. 1, Jakarta, the international conference is organized by Universitas Tarumanagara (UNTAR) and technically sponsored by IOP Publisher.

TICATE 2019 has attracted many academicians, scientists, engineers, postgraduates and other professionals from many countries. This conference accepted 215 papers from 7 different countries, those are Australia, Taiwan, India, Malaysia, Japan, Peru and Indonesia. The aim of the conference is to promote exchange of ideas among engineers, researchers, and scientists active in the related areas of technology and engineering.

Our special thank goes to our Rector, Prof. Dr. Agustinus Purna Irawan, who has initiate this international conference, to our Plenary Speakers, Dr.-Ing. Joewono Prasetijo from Universiti Tun Hussein Onn, Malaysia, Prof. Dr. Tjokorda Gde Tirta Nindhia from Udayana University, Indonesia, Prof. Dr. Srikantappa A.S. from Cauvery Institute of Technology, India, and Prof. Dr. Mohd. Zulkifli Abdullah from Universiti Sains Malaysia, Malaysia, and Prof. Yasuyuku Nemoto, Ph.D. from Ashikaga University, Japan.

Our special thank also goes to Tarzan Photo and PT. Astaguna Wisesa as our patrons. Also to all individuals and organizations such as the members of international editorial board, the conference organizers, the reviewers and the authors, for their contribution in making TICATE 2019 as a successful international conference and a memorable gathering event. I am also grateful for the support of publication service of IOP Publisher. We hope that the conference could present you wonderful memories to bring home in addition to new insights and friendship congregated during the event.

We truly value your participation and support for the conference. We hope that you will enjoy TICATE 2019 and culture and tradition in Jakarta.

Dr. Hugeng, S.T., M.T. (SMIEEE)



2nd TICATE 2019
Conference Organisation

INITIATOR & ORGANIZING INSTITUTION

Universitas Tarumanagara, Jakarta

Supporting Institution

IOP Publisher, Tarzan Photo, PT. Astaguna Wisesa

Honorary Chair

Prof. Dr. Agustinus Purna Irawan
R. M. Gatot Soemartono, Ph.D.

Chairman

Dr. Hugeng, S.T., M.T. (SMIEEE)

Co-chairperson

Dr. Fransisca Iriani Roesmala Dewi

Secretary

Bagus Mulyawan, M.M.

Program & Sponsorship

Mei Ie, M.M.

Herlina Budiono, M.M.

Parallel & Scientific Session

Dr. Hetty Karunia Tunjungsari

Sinta Paramita, M.A.

Treasurer

Wulan Purnama Sari, M.Si.

Keynote Speakers

Prof. Dr. Tjokorda Gde Tirta Nindhia, Universitas Udayana, Indonesia

Dr.-Ing. Joewono Prasetijo, Universiti Tun Hussein Onn, Malaysia

Prof. Dr. Srikantappa A.S., Cauvery Institute of Technology, India

Prof. Dr. Mohd. Zulkifli Abdullah, Universiti Sains Malaysia, Malaysia

Prof. Yasuyuku Nemoto, Ph.D., Ashikaga University, Japan

Editorial Board / Reviewers:

Prof. Dr. rer. nat. Alexander Ferrein, University of Applied Sciences Aachen, Germany

Dr.-Ing. A. Ruggeri Toni Liang, Karlsruhe Institute of Technology, Germany

Dr. -Ing Stephan Herzog, TU Kaiserslautern, Germany

Dr. Thomas Marconi, Inside Secure, France

Prof. Yifan Chen, Ph.D., University of Waikato, New Zealand

Dr. Soh Sie Teng, Curtin University, Australia

Dr. Channing Chuang, Kun Shan Univeristy, Taiwan

Prof. Mohd Zulkifli bin Abdullah, Universiti Sains Malaysia, Malaysia

Prof. Zaidi Mohd. Ripin, Universiti Sains Malaysia, Malaysia

Dr. -Ing. Joewono Prasetijo, Universiti Tun Hussein Onn, Malaysia

Dr. Filbert H. Juwono, Curtin University, Sarawak Malaysia

Prof. Dr. Tresna P. Soemardi, Universitas Indonesia, Indonesia

Dr. -Ing. Eko Adhi Setiawan, Universitas Indonesia, Indonesia

Prof. Dr. Jamasri, Universitas Gadjah Mada, Indonesia

Dr. Bambang Kismono Hadi, Bandung Institute of Technology, Indonesia

Prof. Eko Sedyono, Universitas Kristen Satya Wacana, Indonesia

Prof. Tjokorda Gde Tirta Nindhia, Universitas Udayana, Indonesia

Dr. Rianti Ariobimo, Universitas Trisakti, Indonesia

Dr. Richard Napitupulu, Universitas HKBP Nommensen, Indonesia

Prof. Dr. Dyah Erny Herwindiati, Universitas Tarumanagara, Indonesia

Prof. Dr. Leksmono Suryo Putranto, Universitas Tarumanagara, Indonesia

Harto Tanujaya, Ph.D., Universitas Tarumanagara, Indonesia

Jap Tji Beng, Ph.D., Universitas Tarumanagara, Indonesia

Lina, Ph.D., Universitas Tarumanagara, Indonesia

Dr. Steven Darmawan, Universitas Tarumanagara, Indonesia





PAPER • OPEN ACCESS

Peer review statement

To cite this article: 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **852** 011002

View the [article online](#) for updates and enhancements.

Peer review statement

All papers published in this volume of *IOP Conference Series: Materials Science and Engineering* have been peer reviewed through processes administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a proceedings journal published by IOP Publishing.



PAPER • OPEN ACCESS

Comments Scraping Application For Review Youtube Content

To cite this article: M. Viny Christanti *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **852** 012167

View the [article online](#) for updates and enhancements.

COMMENTS SCRAPING APPLICATION FOR REVIEW YOUTUBE CONTENT

Viny Christanti M.^{1*}, Walda¹, Tri Sutrisno¹

¹Computer Science Department, Faculty of Information Technology, Universitas Tarumanagara

* viny@untar.ac.id

Abstract. YouTube is one of social media that allows people to upload, view, and share videos through website. YouTube's mission statement is to give everyone a voice and show them the world. Through YouTube, people can express their voices or show their creativity effortlessly. YouTube provides service for users to give opinions through comment section. The comments of users can be used to decide either the video has good ratings or not. Web Scraping can be used to take the comments of users from comment section. Web Scraping is a method to take some information from website by extracting data through tags in html. By inspecting code the YouTube website, the information of comments is contained in 'ytd-comment-thread-renderer'. Information of comments is divided into reviewer and content of comments. Reviewer's information is contained in id 'author-text' and the content is in 'content-text'.

1. Introduction

In the new era of globalization, business's competitions are likely increased, both in national and international. It affects the company to think about how to promote products or services in creative and innovative way. One of the ways to promote products or services is promoting some contents through social media.

YouTube is one of the social media that is popular nowadays. Lembaga riset pasar statistika predicts that the users of YouTube will reach around 1.8 billion in 2021 [1]. YouTube's mission statement is to give everyone a voice and show them the world. Through YouTube, people can express their voices or show their creativity effortlessly. People can involve indirectly in YouTube by viewing the videos, giving comments, and sharing the videos to another people. By the advantages of YouTube, company decide to do promotion through YouTube. The promotions are shared by video that has a good content inside.

Company can get the feedback for users by seeing the comments of users in the comment section. The comments can be used to decide either the content of video has good ratings of not. Checking the comments one by one in the comment section will take a long time, so there is a need to scrap all of the comments and use it to do some predictions. Scraping the comments in YouTube can be done by web scraping method. Web Scraping is a method to take some information from website by extracting data through tags in HTML. Web Scraping can be done by inspecting the code of YouTube and see position of the information that will be taken from it. The reviewer and the comments from You Tube will be extracted from web.

Anand, Kedar, and Shweta [2] use web scraping method to mine information from different and unstructured websites and transform it into a comprehensible structure like spreadsheets, database, or



comma-separated values (CSV) file. The data such as item pricing, stock pricing, reports, market pricing, and product details can be used to take effective decisions in business process.

Ahmat, Leon, and Suryayusra [3] use web scraping to make an application for searching scientific article. The application is run by inputting the keywords and the machine will search the information through some portals, such as Garuda, ISJD, and Google Scholar. The machine will show the information if the keywords are matching with the data in database. If the machine did not find the keywords then it will show “the information are not found”

Susan Li [4] uses web scraping to get hotel reviews in Trip Advisor. The reviews can be used to do a sentiment analysis. Sentiment analysis can decide the voice of customer materials that comes from reviews or survey responses. Mahek, Ricky, and Nishil [5] uses web scraping to get reviews of different mobile company such as Apple, Samsung, and Oneplus. The reviews are likely consist of “not Satisfied and satisfied” reviews. The reviews are classified based on keywords, ratings, and emotions that decide the final review score of the products.

In this research we build a system that is used dynamically to scrap and clean text data from the YouTube automatically. The system is equipped with a validation function that can be used as a form to carry out the comment validation process. The result of this application is to prepare comment data for analysis in the opinion mining process.

2. Methodology

The method in this research is the waterfall method in the System Development Life Cycle (SDLC) which is one of the methods of making application programs. SDLC consists of planning, system analysis, system design, system development, system testing by researchers and maintenance. The stages used in this research are the stages in text processing from web that pay attention to the web structure in crawling and scraping.

2.1. YouTube

YouTube is a media for uploading videos and sharing information in the form of videos [6]. YouTube provides a commentary column for comments on shared videos. YouTube is often used as a place to market products and get opinions through the number of likes or comments written by readers. In the Figure 1 we can see video about Samsung note 10 and at figure 2 we can see review from people.



Figure 1. Video about product



Figure 2. Review product from reader

YouTube service provides a column for writing comments that can contain opinions. But sometimes, the word is not an opinion but only meaningless term. In the comments column, other readers can reply the comment to indicate whether they agree or not, or add other comments. The sentence in this comment can also not be related and not mutually supportive. Language inconsistency, inaccuracy in the form of the comment sentence requires validation from humans so that the data entered for analysis is the right data in the form of comments rather than others

In principle, opinions on YouTube are almost similar to opinions on blogs. Where on the blog there is also a comment column that can be used by users to write opinions about the writings listed by the author. On YouTube the comment service is limited in character length. An opinion can be considered as one large document that contains many contents or as a small document that contain one content [7].

Unlike collections from news sources where each document is written briefly and contains topics related to the writing, a comment post is usually slightly improved and tends to be irregular. In the use of comment words for example the word "sedap" is written as "sedaaap", so that the detection of opinion words becomes more difficult. In addition the comments writing page often contains comments that can extend the writing without adding text related to the topic being discussed. Comments collection is also vulnerable to spam content in the form of spam blogs (splogs) and irrelevant comments [7].

2.2. Web HTML Structure

HTML is an abbreviation of Hypertext Markup Language is a script for compiling Web documents [16]. HTML documents are stored in regular text format and contain tags that instruct the web browser to execute the specified commands. The basic structure of an HTML document can be seen in Figure 3. The structure HTML consists of Tag, Element and Attribute.

```
<html>
<head>
<title>Di sini Judul Dokumen HTML</title>
</head>
<body>
  Disini penulisan informasi Web
</body>
</html>
```

Figure 3. HTML Structure

On YouTube it also consists of existing structures in HTML. In figure 4, we can see comments that are visible on the web. But when the source is seen the form consists of html structure. So we must create application that only takes comments without taking other data that is not needed. Some of comments not only consist of letters but can contain images. The sentence "ya ampun senyumnya nic itu sumpah ya bikin meleleh....😄😄😄😄😄" consist of emoticon so the comment must be cleaned up before it can be processed in opinion mining.

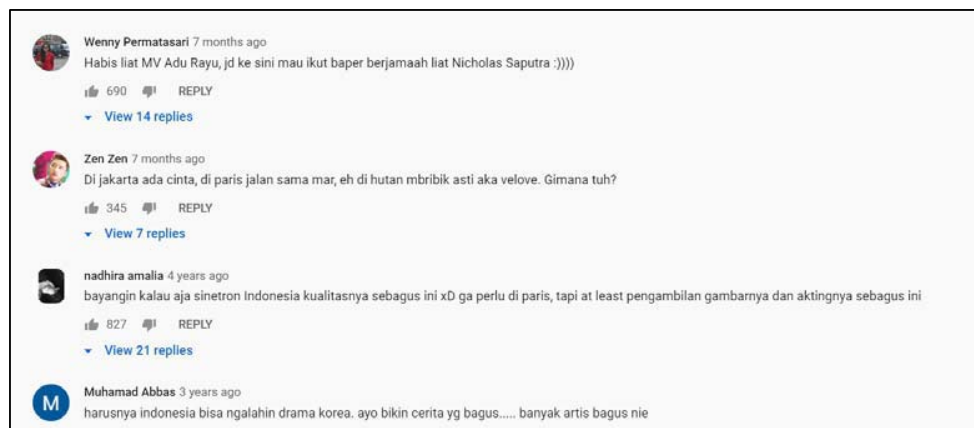


Figure 4. Example of YouTube comment

3. Result

This research makes a comment scraping program on YouTube by using the PHP and HTML programming languages. This name of this application is “Rukomindo” (Pengeruk Komentar Indonesia) equipped with a validation function. The goal of this application is scrap the comment, clean the comment, validation the comment and prepare comment data for analysis. The program consists of 5 main modules namely scraping module, comment results, validation process, and summary of comments validation.

At figure 5, we can see interface of scraping module. At scraping module, we only copy paste the link of YouTube video that we will scrap. We can see the video and the application will automatically take out the comment. User can arrange what they want to scrap, comment only or comment with all replies.



Figure 5. Interface Scraping Module

After put the YouTube link, next step is see the comment result. We can see the comment result at figure 6. There are lists comment that already represent in table format. We can validation the comment and prepare for analysis step. We only save the comment without the name ID who wrote of comment.

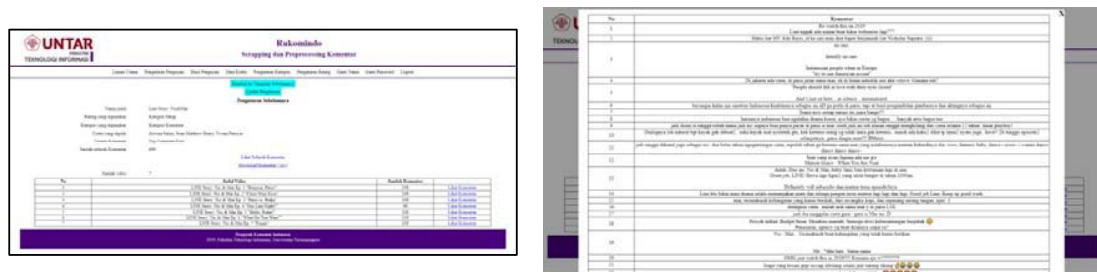


Figure 6. Interface Comment Result

At figure 7 we can see interface of validation process. We can arrange how many validators will give voting for each comment. In the end user will make decision is the comment is opinion or not. User will prepare the data separate the sentence that not consists of statement or opinion. Only opinion will be used for sentiment analysis.

Laman Utama Pengaturan Pengujian Hasil Pengujian Data Koder Pengaturan Kategori Pengaturan Rating Ganti Nama Ganti Password Logout						
1 Kembali ke Page Sebelumnya						
No	Komentar	Pilihan Coder		Valid		
		Ya	Tidak			
1	Re-watch this on 2019 Line nggak ada matan buat bikin webseries lag???	0	3	<input type="radio"/> Ya <input type="radio"/> Tidak		
2	Habis liat MV Adu Rayu, jd ke sini mau ikut baper berjamaah liat Nicholas Saputra :)))	2	1	<input type="radio"/> Ya <input type="radio"/> Tidak		
3	Di jakarta ada cinta, di paris jalan sama mar, eh di hutan mbrbak asti aka velove. Gimana tuh?	0	3	<input type="radio"/> Ya <input type="radio"/> Tidak		
4	bayangin kalau aja sinetron Indonesia kualitasnya sebgus ini xD ga perlu di paris, tapi atleast pengambilan gambarnya dan aktingnya sebgus ini	3	0	<input type="radio"/> Ya <input type="radio"/> Tidak		
5	harusnya indonesia bisa ngalahin drama korea. ayo bikin cerita yg bagus..... banyak artis bagus nie	1	2	<input type="radio"/> Ya <input type="radio"/> Tidak		
6	Suara nico setiap narasi itu juara bange!!! "People should fall in love with their eyes closed"	2	1	<input type="radio"/> Ya <input type="radio"/> Tidak		
7	And I just sit here... in silence... mesmerized.	3	0	<input type="radio"/> Ya <input type="radio"/> Tidak		
8	jadi disini si rangga rubah nama jadi nic supaya bisa punya pacar di paris si mar. oohh jadi ini toh alasan rangga menghilang dari cinta selama 12 tahun. dasar play-by!	3	0	<input type="radio"/> Ya <input type="radio"/> Tidak		
9	Dialognya loh natural bgt kayak gak dibuat2. suka kayak asal nyeleak gbu, kek ketemu orang yg udah lama gak ketemu... masih ada kako2 dlat tp lama2 nyatu juga. kece!! Di tunggu episo-de2 selanjutnya. paris dingin men!!! Ebbnrrr	2	1	<input type="radio"/> Ya <input type="radio"/> Tidak		
10	jadi rangga dikenal juga sebagai nic. dia belas tahun ngegarhngin cinta, sepuluh tahun ga ketemu sama mar yang notabenernya maritan kekasihnya dia. wow, fanbase baby, dance-- ooo--i wanna dance dance dance dance--	2	1	<input type="radio"/> Ya <input type="radio"/> Tidak		

7 [Submit dan Selesai](#) [Submit dan Lanjutkan](#) [Cancel](#)

Figure 7. Interface validation process

After validation process, we can see at figure 8 the summary of validation process. We can download recapitulation from validation process. So we can prepare the data for another process like sentiment analysis or mining process.

UNSTAR
FAKULTAS TEKNOLOGI INFORMASI

Rukomindo
Scrapping dan Preprocessing Komentar

Laman Utama Pengaturan Pengujian Hasil Pengujian Data Koder Pengaturan Kategori Pengaturan Rating Ganti Nama Ganti Password Logout

1 [Kembali ke Tampilan Sebelumnya](#)

Line Story: Nic&Mar

Progress Coder
Total Seluruh Komentar: 699

No	Nama Coder	Jumlah telah divalidasi	Jumlah telah dikategori	Jumlah telah dirating	
1	Alvina Salas	0	0	0	Lihat Hasil Coder
2	Sean Matthew Henry	0	0	0	Lihat Hasil Coder
3	Vivian Patricia	699	170	0	Lihat Hasil Coder

Hasil coder adalah rekap hasil dari semua coder (yang telah menginput data)
Data mungkin belum lengkap
File dalam format csv, yang dapat dibuka dengan microsoft excel.
[Cetak Ekspor Hasil Coder \(.csv\)](#)
Belum selesai semua, tidak bisa proses kesimpulan.

Pengantar Komentar Indonesia
2019, Fakultas Teknologi Informasi, Universitas Turunamagara

Figure 8. Interface Summary of Comment Validation

4. Conclusion

In this research a scraping application has been built to obtain comment data on YouTube. This application can be used to pull comments from a video and provide validation for these comments. So the results of comments that have been collected can be used for mining or further analysis. This application helps users in taking comments from YouTube movies just by filling in the links from the desired YouTube.

This application has been used by the management team who wants to process the film data of Nic and Mar 7 episodes and obtained 699 comments from the video. The research team from the management faculty has also tried to provide validation for the comments in the film.

For further research we will try to create scraping for another social media such as Twitter, Facebook, Instagram and etc. After create scraping, this application must develop for pre-processing step, like separate reply and real comment, separate which negative and positive comment or save the account id to evaluate hoax or non-hoax comment.

5. Acknowledgments

We would like to gratitude to DPPM Untar, for funding this research. Thank you to research assistant that carries out this research. And we would like to thank to Dr. Cokki from Economics Faculty Untar who used our application for their research.

6. References

- [1] Translation from Praditya, Diaz. (2018). 3 Fakta menarik dari Riset Google tentang Perkembangan Youtube di Indonesia, <https://id.techinasia.com/fakta-perkembangan-youtube-di-indonesia>, Accessed on November 2019, 16th.
- [2] Saurkar, Anand V; Pathare, Kedar G; Gode Shweta A. (2018). An Overview On Web Scraping Techniques and Tools. International Journal on Future Revolution in Computer Science & Communication Engineering Volume 4.
- [3] Translation from Josi, Ahmat; Abdillah, Leon A.; Suryayusra. (2014). Penerapan Web Scraping pada Mesin Pencari Ilmiah. Bina Darma University.
- [4] Susan Li. (2018). Web Scraping TripAdvisor, Text Mining and Sentiment Analysis for Hotel Reviews. <https://towardsdatascience.com/scraping-tripadvisor-text-mining-and-sentiment-analysis-for-hotel-reviews-cc4e20aef333>. Accessed on November 2019, 16th.
- [5] Merchant, Mahek; Shah, Nishil; P.Boominathan. (2016). Sentiment Analysis of Web Scraped Product Reviews using Hadoop. IJRASET Volume 4 Issue VIII
- [6] Hopkins, Jim. "Surprise! There's a third YouTube co-founder." *USA Today* 11, no. 10 (2006).
- [7] Elsas, Jonathan L., Jaime Arguello, Jamie Callan, and Jaime G. Carbonell. "Retrieval and feedback models for blog feed search." In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 347-354. ACM, 2008.