Learning history through web scraping and Tableau public: Is it possible?

Joko Ade Nursiyono a1,*, Peni Meivita a2

Abstract: Learning history has a central role in cultivating patriotism and nationalism value for students. Various countries have placed historical education as a crucial part of national education. It is due to the belief that historical education materials can develop the nature and character of the nation’s young generation. Technological developments make it easier for us to create innovative media, such as web scraping. Learning history with proper media can foster students’ motivation to understand the content, such as the names and roles of national heroes. Web scraping is used to collect data from sites related to 60 national heroes and their places of birth. The data is then processed and visualized in Tableau Public. The result shows that the learning media can be categorized as good (53%). The respondent stated that the history learning media is very useful (67%). Meanwhile, regarding the effectiveness of the linkage feature between the data in the visualization, all respondents stated is good.

Keywords: history learning; national hero; tableau public; web scraping

1. Introduction

Learning history is crucial for student to develop their nationalism. In history class, the student can evolve their national nature and character (Awgichew & Ademe, 2022; Malazonia et al., 2021; Nordgren, 2016; Stoddard, 2022). When the student, the young generation, becomes the stakeholders and supporting role in running the nation’s government, their national character becomes a strong foundation in carrying out this role. They understand how this nation was born and developed. Students learn the problems faced in the nation’s past and also how to solve these problems. They learn from past experiences to shape future life for the better and based on the nature and character of the nation (Jaskulowski et al., 2018; Qian et al., 2017; Sant, 2021; Stoddard, 2022). Therefore, history education has a strategic function in developing the nation’s character and building a better future.

The process of learning history should be encouraged to create situations that foster historical awareness. History learning strategies should not turn off creativity and force students to memorize only the facts in the textbook. History is a time to be taught differently. However, the monotones of learning often occur due to low creativity in learning history. As a result, student boredom is often the main factor faced by teachers in teaching history and students in learning history (Ikrām et al., 2022; Kuswono & Khaeroni, 2017; Saidillah, 2018; Santosa, 2012; Senoprabowo & Hasjim, 2013). One part of history is the introduction of national heroes and heroes of the revolution as historical figures who contributed to our nation (Ekadiensyah, 2015; Seto et al., 2015; Utomo, 2016; Wibowo et al., 2021).

The loss of patriotism and nationalism values among students is very alarming. Indonesian students today remember fictional figures who are considered heroes rather than knowing Indonesian national heroes (Ekadiensyah, 2015; Mustika & Wahyuningsih, 2015).
A hero is a person who stands out for his courage and sacrifice in defending the truth, a valiant warrior. The national heroes are a title given to an Indonesian citizen who fought against colonialism in the territory of the Republic of Indonesia. They died in defense of the nation or committed acts of heroism and produced outstanding achievements and works for the development and progress of the state of the Republic of Indonesia (Mustika & Wahyuningsih, 2020).

Every citizen has the right to promote, fight for, and obtain equal opportunities in building society, nation, and state to deserve appreciation for the services dedicated to the glory and establishment of the Republic of Indonesia based on Pancasila and the Constitution of the Republic of Indonesia year 1945. Awards for services rendered by the state for nation warriors are in the form of titles, service marks, and honor marks to foster pride, exemplary attitude, fighting spirit, and motivation to increase devotion to the nation and state. It is stated in the law of the Republic of Indonesia number 20 of 2009 (Ministry of Social Affairs, 2009). Nevertheless, people often forget where the heroes come from and what form of struggle and sacrifice they have done to get their title as the national hero. Knowing national heroes also leads us to learn and understand the history and important events that occurred in the past (Ekadiansyah, 2015; Mustika & Wahyuningsih, 2020; Seto et al., 2015; Utomo, 2016; Wibowo et al., 2021). Therefore, it is urgent to encourage innovative history learning.

Along with the development of information technology, many devices are used to access information. Despite that, the information related to Indonesian national heroes is dominated by posters and books. Several studies on designing historical learning support systems have also been carried out. Firdaus et al. (2021) have developed Android-based infographics as a history learning media. Mustika and Wahyuningsih (2020) applied the direct search method to the South Sulawesi hero's history information system based on Android. Some researchers introduced Indonesian hero figures using a game (Arta et al., 2020; Senoprabowo & Hasyim, 2013). Then, the research from Utomo (2016) was producing an application to guess images of heroes and temples in Indonesia using the Linear Congruent Method (LCM). Seto et al. (2015) developed the introduction of Indonesian heroes based on augmented reality with Indonesian money markers. Furthermore, another study by Wibowo et al. (2021) was a face-to-face learning application about national heroes produced using a prototype system. There is also research utilizing audio-visual media to learn history (Ainina, 2014).

Some of these studies created a learning media to support learning about national heroes. However, it’s not easy for teachers to use it. In addition, the results have not accommodated any updated data on national heroes. Therefore, this study combines web scraping with Tableau Public platform to design a learning media related to Indonesian national heroes. Web scraping is using bots to extract content and data from a website (Dogucu & Çetinkaya-Rundel, 2021; Flores et al., 2020; Yondra et al., 2022). The content is related to Indonesian national heroes. While Tableau Public work as a platform to explore, create, and publicly share data visualizations online (Batt et al., 2020; Nolan & Perrett, 2016). Combining both web scraping and Tableau Public can utilize to design engaging learning media. Based on this background, the author hopes to help facilitate the learning of history and provide other Learning media alternatives which are more captivating and fun for the student. So, they can feel motivated to learn, memorize, and understand the meaning of names and the role of Indonesian national heroes.

2. Materials and Methods

This quantitative descriptive research utilizes secondary data and visualizes it to produce an online learning support application. The secondary data is related to the names and birthplaces of Indonesian national heroes. The data research comes from the “zon-referensi.com” site, while the heroes’ photographs are from the “id.wikipedia.org” site. The data was obtained by the web scraping technique and then visualized using Tableau Public. Due to the limited number of heroes described and the availability of heroes’ photographs, the number of visualized heroes in this study is only 60 heroes.
The visualization design of learning media uses two tools, RStudio and Tableau Public. RStudio is used to perform the stages of web scraping data from sites containing national heroes and their birthplaces. Tableau Public is used to build an online face-to-face application for national hero learning. At the final stage of this design, the questionnaire was conducted to measure user response. There are 30 users as samples. In summary, the steps of designing a national hero learning media in this study are described as follows:

1. Build web scraping formulation by using an R programming language with RStudio GUI as follows:
   
   ```
   library(rvest)
   library(xml2)
   library(stringr)
   
   # Read the URL address of the site
   laman <- read_html("https://www.zonareferensi.com/daftar-nama-pahlawan-nasional-indonesia/"")
   
   # Obtaining the name of the hero
   nama <- laman |> html_nodes('tr+ tr td:nth-child(2) , .grippy-host') |> html_text()
   
   # Obtaining the name of regional origin
   asal <- laman |> html_nodes('tr+ tr td:nth-child(5)') |> html_text()
   
   # Obtaining year of birth
   lahir <- laman |> html_nodes('tr+ tr td:nth-child(3)') |> html_text()
   
   # The year of death
   wafat <- laman |> html_nodes('tr+ tr td:nth-child(4)') |> html_text()
   
   # Unify in the data frame
   datapahlawan <- data.frame(nama, asal, lahir, wafat, stringsAsFactors = F)
   
   # Converting birth and death data as numeric
   datapahlawan$lahir <- as.numeric(datapahlawan$lahir)
   datapahlawan$wafat <- as.numeric(datapahlawan$wafat)
   
   # Creating a new variable of hero age
   library(dplyr)
   datapahlawan <- datapahlawan %>%
   mutate(usia = wafat - lahir)
   ```
   
2. Perform web scraping of national hero data from the "zonareferensi.com" site. The data obtained from web scraping are described in Figure 1.
3. Download national heroes’ photographs based on data from web scraping results on the “id.wikipedia.org” site (Figure 2). If it turns out that there is no photograph found on the “id.wikipedia.org” site, then it can be downloaded from another source.

4. Check the suitability of the photographs with the name of the heroes and their birthplaces (Figure 3).

5. Link the data of the hero’s name, photo, and age of his life with the data of his province of birth based on the image ID number and hero name ID

6. Prepare the extension map *shp for spatial visualization needs in Tableau Public.

7. Make visualization using Tableau Public by enabling links and matches.
8. Conduct the evaluation of the designed history learning media by questioning user samples using a short google form questionnaire.

3. Results

The web scraping result on the “zonareferensi.com” site shows that 60 subjects' hero names are determined. Then, the data is processed and visualized in Tableau Public. Visualization as a face-to-face learning material displays only 34.29 percent of existing heroes due to limited visualization space (Figure 4). The result of the dynamic visualization generated using Tableau Public can be shown in Figure 4. Figure 5 shows the link and match function checking.

![Figure 4. The result of dynamic visualization generated using Tableau Public](image1)

![Figure 5. Link and match function checking](image2)
The limitation of the national heroes’ data causes the developed media can not to be covered entirely. So, in the design of history learning media, some heroes are given a white identity. However, territorial scope (province) in the media design covers 52.94 percent of the entire territory of Indonesia. The area covered as the largest sample of this learning media design is West Sumatra province. As many as nine heroes with the distribution of the hero’s life span between 55 years to 92 years with a median life span of 70 years.

Meanwhile, the area with the small coverage consists of Papua Province, East Nusa Tenggara Province, Lampung Province, Central Kalimantan Province, West Kalimantan Province, Greater Jakarta Province (DKI), Bengkulu Province, and Bali Province. There is only one hero for each province. The research findings from the process of visualization of the region with only one number of heroes are displayed in the cut photographs. So, the media only can show the name, the hero’s struggle, the origin of the hero’s region, and the age of the hero’s life (Figure 6). The link and match result between the hero’s name and the photographs of the web scraping results in the visualization process of designing the media showed an accuracy of 98.33 percent. The error link and match visualization design process are only 1.67 percent. It is caused by the similarity of the hero’s name in web scraping results, for example, Abdul Halim and Abdul Halim Majalengka.

Figure 6. The visualization of single hero count display checking

The evaluation result of the quality of the designed history learning media is described in Figure 7. The evaluation of the designed historical learning media was carried out by visualizing the interface to 15 respondents.

Figure 7. The evaluation result of the quality of the designed history learning media
The respondents include 53 percent of undergraduate (S1/Diploma), 27 percent of senior high school graduates (SMA/SMK), 13 percent of post-graduate (S2/S3), and as many as 7 percent of the junior high school graduate (SMP/MTs). The majority of respondents stated that the quality of visualization categorizes as “good” (53 percent). As many as 40 percent of respondents express that the media quality is “very good”, while only seven percent categorize the media as a “quite good” category (Figure 7). Meanwhile, regarding the effectiveness of the linkage feature between the data in the visualization, all respondents stated that it went well.

In terms of usefulness, the majority of respondents stated that the designed history learning media is very useful (67 percent). Then, followed by 20 percent express that media is useful. Meanwhile, as many as 7 percent agree that it was quite useful to use in learning (Figure 8). And there are 6 percent stated that it was less useful. One of the reasons why respondents stated that it was not useful was the lack of additional information to expand the user’s knowledge, for example, information about the heroic period, before independence or after independence.

Figure 8. The evaluation result of the usefulness of the designed history learning media

4. Discussion

The result shows that the designed learning media can be categorized as good (53%). The results show that the designed media can be used in history learning, especially national heroes’ topic. The data obtained from the web scraping process using the RStudio computer program is related to the hero’s name, location, and date of birth. Several studies state that web scraping can facilitate the collection of users’ information needs (Flores et al., 2020; Yondra et al., 2022). This data is then visualized using the Tableau Public platform. The results of this good visualization provide captivating learning media for students. Several previous studies also stated that the more color combinations in a media, the more attractive students would be to use them in learning (Ainina, 2014; Lindner et al., 2019; Samsudin et al., 2019; Setianingrum, 2017). The existence of media that attracts students’ attention can increase their learning motivation (Ainina, 2014). As is known in learning history, many students feel bored with the material (Saidillah, 2018; Senoprabowo & Hasyim, 2013). So, it is crucial to have innovations in learning that utilize technological developments.

The respondent stated that the history learning media is very useful (67%). Meanwhile, regarding the effectiveness of the linkage feature between the data in the visualization, all respondents stated is good. Web scraping techniques make it easier for users to get information about national heroes. Web scraping uses bots to extract content and data from websites. The existence of technological assistance in designing learning media
facilitates the teacher's work (Dogucu & Çetinkaya-Rundel, 2021; Flores et al., 2020).

Meanwhile, Tableau Public works as a platform to explore, create, and share data visualizations online with the public (Batt et al., 2020; Nolan & Perrett, 2016). Combining web scraping and Tableau Public can be used to design attractive learning media. The results of the visualization of the designed history learning media were categorized as “very useful” according to the respondents. It means that designed media can be used for learning in the classroom. Teachers can take advantage of the media developed to teach national heroes topic.

However, this study only designed historical learning media that utilized RStudio for the web scraping process and the Tableau Public platform. The result of the designed media was then assessed by 15 respondents. The results of media design need to be tested both on a small scale and on a large scale to find out whether the media is feasible and reliable. Thus, further research related to the designed media needs to be carried out.

5. Conclusions

The design of history learning media in this research uses two tools, namely RStudio and Tableau Public platform. RStudio is used to perform web scraping data from sites containing national heroes and their birthplace data. There are 60 subjects of hero names that are processed and visualized in Tableau Public. Tableau Public is used to build an online face-to-face application for history learning media. Visualization as a face-to-face learning material displays 34.29 percent of existing heroes due to limited visualization space. However, the territorial scope (province) of the designed media covers 52.94 percent of the entire territory of Indonesia. At the final stage of this design, interviews using a short questionnaire about the quality of the designed media. The result shows that the learning media can be categorized as good (53%). The respondent stated that the history learning media is very useful (67%).

Meanwhile, regarding the effectiveness of the linkage feature between the data in the visualization, all respondents stated is good.

Author Contributions: Methodology, Nursiyono, J.A. and Meivita, P.; validation, Nursiyono, J.A.; analysis Nursiyono, J.A. and Meivita, P.; writing—original draft preparation Nursiyono, J.A. and Meivita, P.; review and editing, Nursiyono, J.A.

Acknowledgments: We would like to thank the Principal of Central Bureau of Statistics of East Java Province for supporting the implementation of this study.

Conflicts of Interest: We declare that there is no conflict of interest.

6. References


Utomo, A. B. (2016). Aplikasi tebak gambar pahlawan dan cendiki Indonesia menggunakan metode Linear Congruent Method (LCM) [Universitas Nusantara PGRI Kediri]. http://simki.unpkediri.ac.id/detail/11.1.03.02.0026