RESEARCH ARTICLE   OPEN ACCESS

Data Extraction Using The Web Crawler As A Media For Information On The Popularity Of Lampung Province Tourism For The Development Of Rides And Tourism Facilities

Ade Dwi Putra[1], Temi Ardiansyah[2], Sukadiningrat[3], Dede Latipah[4], Sarif Hidayat[5]

1(Departmentof Information Technology, Budi Luhur University, Jakarta Email: adedwiputra.tekno@gmail.com)
2(Department of Information Technology, Budi Luhur University, Jakarta Email: temiardiansah@gmail.com)
3(Department of Information Technology, Budi Luhur University, Jakarta Email: ningratdoang@gmail.com)
4(Department of Information Technology, Budi Luhur University, Jakarta Email: dedelatifah0906@gmail.com)
5(Department of Information Technology, Budi Luhur University, Jakarta Email: sarifhid@gmail.com)

--***********************------------------------------------

Abstract:

Referring to the government regulation of the Republic of Indonesia number 50 of 2011 concerning the 2010 - 2025 national tourism development master plan in point number 4 states that the tourism destination area hereinafter referred to as a Tourism Destination is a geographical area within one or more administrative regions in which there is Power Tourism, Public Facilities, Tourism, Accessibility Facilities, and communities that are interrelated and complement the realization of tourism. One of the tourism facilities stated in the government regulation is tourism, so it is necessary to know what tourism is the trend in tourism places, especially in Lampung. ways to find out the facilities that are in trend in tourism places including visiting tourism places, asking directly to the tourism office or seeing tourism facilities that are currently on social media trends, such as Facebook, Twitter and Instagram. This means that Lampung province has tourism destinations that can be developed by Lampung or private provinces. A suitable method for carrying out this process is the Web Crawler that will produce clusters. Web Crawler, also known as Web Spiders, Web Walcers is a program or automatic script that searches the World Wide Web automatically to find relevant information. From the results of research that has been done, web crawlers can display information about the tourism vehicle searched through the search page. . Data search is based on the tags typed by the user, where the typed tags will be adjusted to the existing tags on Instragram social media that have been uploaded by Instagram users. The data displayed will be grouped per tourism vehicle in graphical form, so that users can see market segmentation about tourism facilities that are in demand easily.

Keywords —**Tourism, Web Crawler, Lampung Tourism**
--***********************------------------------------------

I. INTRODUCTION

Lampung Province is the easternmost region of the island of Sumatra and makes it the closest province outside Java that is closest to the capital city of Jakarta. Besides being the entrance to the economy of Sumatra Island, Lampung Province also has a lot of tourism potential. The beauty of the natural bay of Lampung, Bukit Wan abdurahman, Putuk Stone Nature Tourism, Butterfly Park, Market Island as Mina Politan, Taman Dipangga, Putri Well, Traditional House, Enggal Art Market, Souvenir Center, Culinary Center, Souvenir Center that becomes icons for each region and other historical relics and the availability of infrastructure and tourism services can be the main attraction and target of foreign tourists and tourists visiting Lampung Province.

Lampung Province has quite a lot of tourist objects which are scattered in 13 districts, which of course not all have been recorded properly. In addition to tourism objects that need to be developed, there are also tourism supporting potentials that need to be considered, namely the potential of human resources, mental attitudes and cultural values related to Sapta Enchantment in Tourism. Finally, tourism potential in

Lampung, especially non-superior tourism needs to be well developed, because in addition to increasing the country's foreign exchange, expanding employment, and introducing and preserving a variety of beautiful Indonesian culture and nature (Rostiyati, 2013).

Referring to the government regulation of the Republic of Indonesia number 50 of 2011 concerning the national tourism development master plan for 2010 - 2025 in point number 4 states that the tourism destination or tourism destination is a geographical area located within one or more administrative regions, within the region must have Tourism Attractions, Public Facilities, Tourism, Accessibility Facilities, and communities that are interrelated and complement the realization of tourism. Based on data from the tourism office of Lampung Province there are approximately 42 tourism places managed by the tourism service. This means that Lampung province has tourism destinations that can be developed by Lampung or private provinces. According to the results of the interview with Dra.Agustinawati, MM as Head of Tourism Marketing Development. One of the tourism facilities stated in the government regulation is tourism, so it is necessary to know what tourism is the trend in tourism places, especially in Lampung. ways to find out the facilities that are in trend in tourism places including visiting tourism places, asking directly to the tourism office or seeing tourism facilities that are currently on social media trends, such as Facebook, Twitter and Instagram.

Instagram is one of the Social Media which has more than 600 million active users, seeing many users, the government or owners and managers of tourist attractions can make Instagram social media as a tool to determine market segmentation of tourism that will be aimed at sustainable tourism. Segmentation must be done by the owner or manager of a tourist place that provides tourism facilities, where segmentation is the process of dividing customers, or potential customers, into markets into different groups, or segments (McDonald and Dunbar 2012).

Segmentation can bring or create sustainable tourism. To be able to determine the segmentation of Instagram, it takes the process of taking pictures and locations obtained from tags. A suitable method for carrying out this process is a web crawler that will produce clusters. Web Crawler, also known as Web Spiders, Web Walcers is an automated program or script that automatically searches the World Wide Web to find relevant information (Schrenk 2012).

The Web Crawler uses an algorithm that narrows its search so that it can find out relevant information. This Web Crawler is used to collect data obtained from text taken from existing social media, the public who writes or updates the status in the social media will be information that can be used, the method is used as data presentation to find out tourism segmentation in Lampung (Porter, 1998).

Based On The Theoretical Study Of The Web Crawler, Grouping Data By Displaying The Graphics Produced By Retrieving Data From Instagram, And Then Creating A Segmentation That Will Be Used To Find Out The Advantages Of Tourism, Then See The Many Existing Tourism Triggers The Development Of Sustainable Tourism.

II. THEORY

A. Web Crawler

Web Crawler, also known as Web crawlers, Spiders, and web walkers a program or automated script which browses the World Wide Web by automatically to search for relevant information. The primary function of a web crawler is to perform exploration and retrieval of web pages available on the Internet. the results of the collection of the web site will be indexed by the search engines so that it makes it easy to search information on the Internet [5]. Web Crawler using algorithms that narrow your search so you can find out relevant information. The relevance of information is determined by the algorithm that is used by the crawler by considering factors such as the frequency and location of keywords on a web page. Crawlers also perform the function of taking on new websites and index them.

B. Market Segmentation

Market segmentation is Dividing the market into smaller segments with different needs, characteristics, or behavior that may require a separate marketing strategy or a mixture

C. Competitive Advantage

Competitive advantage is the ability of companies to create advantages over competitors. This causes consumers choose the products of the company for the activities of consumption, competitive advantage is the scope of activity that seeks an accomplished business entity generating three competing strategies that are useful for creating a secure position the company, as in the picture 1
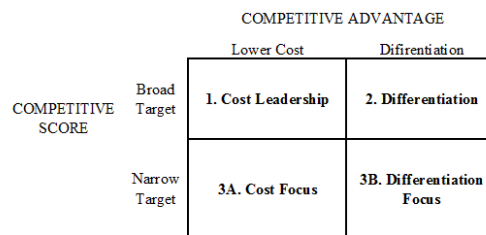
COMPETITIVE ADVANTAGE

COMPETITIVE SCORE		Lower Cost	Difirentiation
	Broad Target	1. Cost Leadership	2. Differentiation
	Narrow Target	3A. Cost Focus	3B. Differentiation Focus

Figure 1*Three Generic Strategie*

III. METHODOLOGY

Research on application of writing a Web Crawler as the development tools market segmentation to achieve regional supremacy tujaun tours have an overview of the research include: Instagram is social networking as a test object that you want to research analysis of tags and pictures of her with a Web Crawler, the process of extraction and analysis of text and images from social media, only deals with the retrieval of tags and supporting data for the presentation of data serving market segmentation, data generated from text analysis – the text is done, presented in the form of graphs based on market segmentation.

Method of data collection is done by studying literature, documentation and interviews. The testing program carried out with the required respondents who will do the filling the questionnaire, respondents will be asked to test the program with the approach of Black Box Testing, i.e. testing the functionality of the program. In this study the measurement scale used is the Likert scale to a positive statement. Likert scale is a scale designed to assess the degree to which respondents agree or not with the statement on a scale with the order as follows:

Jawaban	Skor
SangatSetuju	5
Setuju	4
Ragu – Ragu	3
TidakSetuju	2
SangatTidakSetuju	1

Table 1 Likert Scale

With Likert scale, then the variable will be elaborated indicators measured variables. Then the indicator be used as the starting point for the drafting of the instrument – the instrument in the form of a statement.

In making Crawling social media being used, the entire process has to go through several stages. As for measures on the stages of research can be seen in the picture below:
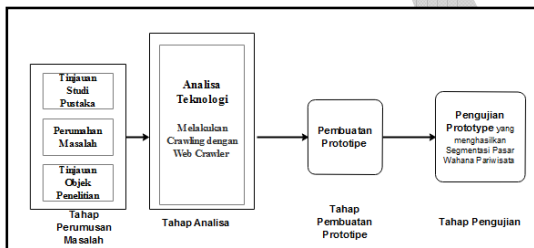

Figure 2 Steps-Steps Research

IV.RESULT AND DISCUSSION

A. Business Process Analysis

Marketing conducted by the tourism agency can become a reference for the selection of tourist attractions in the province of lampung, because with the development of technology, according to the Dra. Agustinawati, MM as the head of the Development Department, Tourism Marketing Pariwisasta also already using social media and the internet as a medium of promotion. However, according to Rostiyati (2013) tourism should be developed, bearing in mind the tourists would love to experience adventure and challenge. But lampung should market the entire tourism potential in Lampung in particular non-seeded tours need to be well developed, because in addition to adding to the country's foreign exchange, expanding employment, and introduce and preserve the diverse culture and nature Indonesia. So the tourists get a reference for determining the interest even though the flagship tour is not a land tour.

B. Needs Analysis System

Sorting images in Crawling from Instargram will be done to find out what seduh method favoured community. Then after sorting the pictures done. It will be created a graph that can be selected in accordance with the region that issued the photo to Instagram. For that it needs needs analysis to be used in the making of this research.

a. Functional
Prototype developed must have the following requirements:
1. User Friendly, a prototype should be able to allow a user in penginputan the products you wish to search, both in terms of user interface as well as the ease of using the prototype.
2. Prototypes should be able to display the search results that the user inputted earlier in the search user interface.
3. The prototype must be able to present the graphs of clustering market segmentation that is selected by the user.
4. Prototype should be able to do a search reset when user want to look for a different product.
5.

b. Non Functional
Non-functional requirements in setting up and running the prototype is as follows:

1. The needs of Software used
   a. Language PHP Programmers
   b. the operating system that can run a Web Browser
   c. Mysql Database
   d. Local Host Server
2. The needs of Hardware used
   a. Your PC with minimal Processor Core i3 sepesifikasi
   b. Minimum 2 GB Ram
   c. 512 GB SATA Harddrive
   d. Internet access with a minimum of 3 g Network

C. System Architecture
Prototype made do by looking at the architecture and also the draft have been made before. The process of shooting for later done segmentation can be seen in figure 3 below:
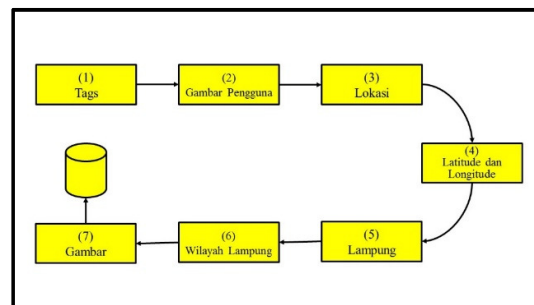

Figure 3 the process of Shooting

The process of taking a picture with a Web Crawler via Instagram is done by means of:

1. **Select The Tags That Will Be Used For The Search Process.**
   The tag is the tag specified earlier Tag-related tours Lampung Province, there are 5 Tags specified waterparklampung, kolamrenanglampung, alamlampung, wisatalautlampung and outboundlampung.

2. **Image Capture**
   Collected images will be taken on his user id. Your user ID will be used to determine the location in the picture.

3. **Site Selection**
   This is processed will be chosen image that has a location. If there are images that don't have the location of the image will not be selected.

4. **Retrieval of Latitude and Longitude**
   Then after the location obtained then the next step is taking the Latitude and Longitude of the location. Latitude and Longitude are taken by way of a notice from the meta data contained in a web document that is in the picture that is already selected.

5. **Location of Lampung**
   Next up is the process of mapping the location of Lampung and the determination of the regions to be taken pictures and the region. Latitude and Longitude are processed to likened to a location in the world for will be in doing the selection, if the Latitude and Longitude that is in the territory of Indonesia then pictures will be selected, if not go to the region Indonesia then the Image will not be selected.

6. **The Location of The Region of Lampung**
   The next step is mapping the region of Lampung. The images are already determined its location there in Lampung will be made the last choice of location, where part of the Lampung region the image came from. This process is determined by dividing into 15 locations in the region of Lampung. 15 counties and 1 region to be determined if the pictures out of the specified area. After the process is finished then the images will be shown to later be used for segmentation of tourism in Lampung.

7. **Display Pictures**
   After the process is finished then the image will be displayed and the disimpanuntuk will be used for segmentation of tourism in Lampung.

D. Architectural Prototype
   The architecture of the Prototype was made to describe the process of retrieval of tags needed to gather information about the location on Instagram. Location obtained will be mapped and customized with a map of Lampung. To be able to see a prototype of the system architecture that will be created can be seen in the picture below:
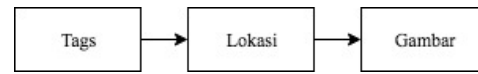

Figure 4 Architectural Prototype

The first specified tags to be used for the search process, and then after that the image of such a tag will be processed. Later the location of found images, After the location was obtained by then taken Latitude and Longitude of the location.

E. User Interface
The system is made to be used by the community, pembisnis, even the service of the province of Lampung. This step explains how the program works as a tool in the search for tourism which is being demanded by the people of Lampung. The program is created using a Web Crawler, the following is a rather simplistic explanation of each component of the program.

a. Search Page
The search page is the page that is used for the menginputkan keyword to search for, i.e. segmentation of tourism by the user. As for the look of the search page look in Figure 5 as follows:

Figure 5 Display Search Menu

b. The Search Results Page
The search results page is the page that displays search results has been done and has been in Crawling by keyword selected. As for the look of the search results page looks in picture 6.
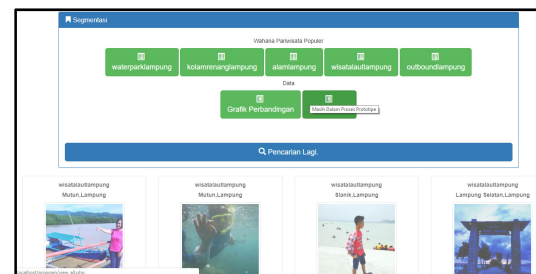

Figure 6 Menu Display Search Results

c. Page Clustering Segmentation Tourism
Page grouping of data segmentation of tourism is the page that is used to present the data in the form of graph appropriate options from the user, rendering the graph using Graph pie divided 2 i.e. the overall tourism comparison chart Lampung and graphs pie per tourism. As for the page view graph comparison seen in Figure 7.
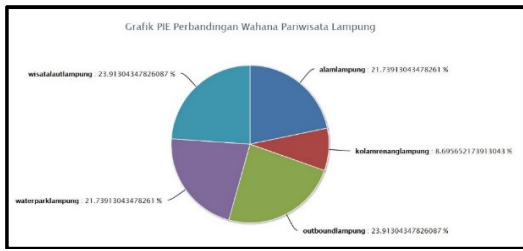
Figure 7Pie Charts Display Comparison Of The Entire Tourism

F. Prototype Test Results

Acceptance testing is performed using the Tourism Web Crawler Technology Acceptance Model (TAM) who will be tested based on the perception of usefulness and ease of use are specified to be some indicator that is speeding up the work, improve performance, improve productivity, effectiveness, ease of work, useful, easy to learn, can be controlled, clear and intelligible, flexible, easy to be skilled, and easy to use. Stages of acceptance testing is a Web Crawler:

a. Create a Questionnaire

The questionnaires are made to do this testing is based on 12 indicators of acceptance of technology that made one grain for each charge indicators will question. Grain question obtained based on the source of Davis (1989) which is tailored to the research

b. Dissemination of The Questionnaire And Data Collection

At this stage the author uses a Simple Random Sampling method, according to Husaini and Purnomo (2018:183) the main characteristic of this sampling is any element of the overall population have an equal chance to be selected. The next step is determining the sampling of some Parties at the University of Indonesia, with a total Technocrat respondent as much as 71 respondents with details as follows: students as many as 30 people, as many as 31 people, lecturer and Assistant Professor 10 person. A questionnaire given to respondents with 12 questions that should be answered by the respondents, 6 questions a question that leads to the perception of usefulness, and 6 the question is a question that leads to the perception of ease of use.

A percentage score based on the actual data table responses respondents based on the perception of usefulness, namely:

$$\%Actual\ Score = \frac{Actual\ Score}{Ideal\ Score}\ x\ 100\%$$
$$\%Actual\ Score = \frac{2552}{2982}\ x\ 100\%$$
$$\%Actual\ Score = 0.86\ x\ 100\%$$
$$\%Actual\ Score = \mathbf{86}\%$$

Data from the responses of the respondents on the basis of the perception of usefulness, the actual total score of 2552 be in range the ideal score very much agree that is the ideal score and agree 2982 i.e. 2556, so it can be said that 86% of respondents agreed that Tourism Web Crawler having a usability or helpful.

A percentage score based on the actual data table responses respondents based on perception of ease of use, namely:

$$\%Actual\ Score = \frac{Actual\ Score}{Ideal\ Score}\ x\ 100\%$$
$$\%Actual\ Score = \frac{2611}{2982}\ x\ 100\%$$
$$\%Actual\ Score = 0.88\ x\ 100\%$$
$$\%Actual\ Score = \mathbf{88}\%$$

Data from the responses of the respondents based on perception of ease of use, the actual total score of 2611 is at the ideal score range could not agree i.e. 2898 and score ideal agree IE 2556, so it can be said that 88% of respondents agreed that the Web Crawler Tourism is easy to use.

V. CONCLUSION

Prototype Web Crawler for tours Lampung province is designed to help users in terms of knowing the classification and any facilities which are much sought after by tourists. By using this Web Crawler users asked to route the search by entering keywords in the form of Tags related to tourism in the province of Lampung. The next Web Crawlers will find data that corresponds to the tag and will perform some image search results in the form of photos, your user name, comment and also the location of the photo. Based on the description, the authors can conclude that:

1. The process of determining kliasifikasi tours Lampung province is to check for tags that have been made by the user previously Instagram. The election of this most tags relied upon the determination of the classification of tourism in the province of Lampung. Information on the classification of the existing tours Lampung province such as: pool, waterpark, nature, sea and outbound obtained based on the process of Crawling on a hashtag related tours Lampung province through social media Instagram. Based on the results of the classification used to pull the data based on the tags that users have uploaded Instagram.
2. By using the Web Crawler tours Lampung users get information on tourist facilities that are the trend in the province of Lampung based on postings from social media Instagram.

In this study the author also did testing against a user response to Prototype Web Crawler for tours Lampung province through the questionnaire. A questionnaire given to respondents with a question that leads to the perception of usefulness and the perception of ease of use. From the responses of the respondents on the basis of the perception of usefulness, the total score is at the ideal score range, so it can be said to be 86% of respondents agreed that the Web Crawler for tours Lampung province have usefulness or benefit. Based on the perception of ease of use, the total score is at the ideal score range, so it can be said to be 88% of the respondents agree that the Prototype Web Crawler for tours Lampung province is easy to use.

REFERENCES

[1]   Ahuja, M. S., Bal, D. J. S., andVarnica,*Web Crawler* : *Extracting The Web* Data. *InternationalEngineering Research* Journal (IERJ), 1(8), 629–632, 2015.

[2]   DeLone, W. H., and McLean, E. R,*TheDeLone and McLeanModel of Information Systems* Success : A Ten-Year Update, 19(4), 9–30, 2003.

[3]   Kolter, P. and Armstrong, G, *Principes of Marketing* .New Jersey : Prentice Hall, 2012.

[4]   McDonald, M., and Dunbar, I, Market *Segmentation :How to do it and how to profit from it. Southern Gate,United Kingdom:* John Wiley & Sons Ltd, 2012.

[5]   Porter, M. E,*Competitive Advantage Creating and Sustaining Superior Performance*. newyork: Simon & Schuster Inc, 1998.

[6]   Rajmohammod, S. S., Arun, P. P., andDnyandeo, G. P,*Web Crawler* : *Extracting The Web* Data. *International Engineering Research* Journal (IERJ), 1(8), 629–632, 2015.

[7]   Rossa A.S , M. Shalahuddin, *RekayasaPerangkatLunak*. Bandung : Modula, 2014.

[8]   Schrenk, M,*Webbots, Spiders, and Screen Scrapers a Guide to Developing Internet Agents with* PHP/CURL. *No Starch Press*. San Francisco: *No Starch Press, Inc*, 2012.

[9]   Sevilla, Consuelo G,*Research Methods. Rex Printing Company*. Quezon City, 2007.

[10]   Singh, A. V., and Mishra, A,*A Review of Web Crawler Algorithms*, 5(5), 6689–6691, 2014.

[11]   Usman, H and Akbar P.S,*PengantarStatistika*. Jakarta : PT BumiAksara, 2008.