The Design of Automatic Bird Data Capture Systems

Chia-Yuan Lin[1], Yi-Chun Lin[2], Che-Jui Chang[2], Chi-Feng Chen[3] and Ming-Fong Tsai[4, *]

[1]Department of Information Engineering and Computer Science, Feng Chia University, Taiwan
[2]Information and Communications Research Laboratories, Industrial Technology Research Institute, Taiwan
[3]Industrial Ph.D. Program of Internet of Things, Feng Chia University, Taiwan
[4]Department of Electronic Engineering, National United University, Taiwan
*mingfongtsai@gmail.com

***Abstract*—This paper implements an automatic bird data capture system which uploads data captured from social network websites to the server and stores them in the database. When using the Facebook API to capture data of interest, the system needs to filter the captured data to avoid false data storage. This paper takes building an automatic system that includes reinforcing the training database of the bird community as a top priority.**

Keywords: Capture System; Social Network;

I. INTRODUCTION

Preparatory work for bird picture collection is a necessary step for training machine learning or deep learning models. Thus, we first adopt existing websites with photographic guides to birds to capture and classify the data to build an open-source picture database. However, the above-mentioned websites with photographic guides are still insufficient to provide an efficient inquiry for bird pictures. Generally, manpower is required for further recognition, so this study automatically captures related keywords and bird pictures involved in these posts from social network websites and provides them to an open-source picture database to proceed with machine learning and deep learning so as to obtain a more and more precise bird picture database. Using machine learning and deep learning to screen information found by users is an effective method to enhance overall satisfaction with the system. This study regards building a system that takes reinforcing the training database of the bird community as a top priority. This paper is organized as follows: Section II describes the design of the automatic bird data capture system. Section III discusses the system implementation and results. Section IV presents the conclusions of this paper.

II. DESIGN OF AUTOMATIC BIRD DATA CAPTURE SYSTEM

This system uses the biggest social network website in Taiwan to capture data. First, we write a crawler program that can read data on webpages and combine it with an API provided by Facebook to capture posts on fan pages. After the data captured from websites with a picture guide is classified, it will then be stored in the server end and the database. Figure 1 shows that Facebook is the biggest social network website, with massive posts and data updates every day, and meanwhile, it has a lot of fan pages which can be treated as a source of big data. This social network website also provides a variety of development software APIs for developers. Web Data in Figure 1 is searching for classified websites to capture data. It focuses on online encyclopedias or websites with photographic guides to rebuild a database to store data captured from Facebook. The Web Crawler in Figure 1 is developed using Python and meanwhile it uses the API provided by Facebook to capture posts on fan pages, and finally data are classified and stored in the server according to Web Data. The Data Server in Figure 1 is used to store all the images and manage the database, classifying captured images in the database and managing image storage. Web in Figure 1 manages the network platform through Tomcat and meanwhile constructs an Internet platform that can display the captured images to other users.

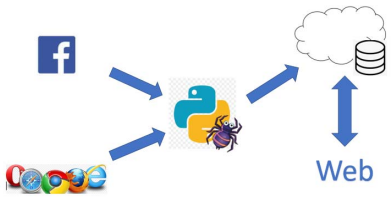

Figure 1: System Framework

III. SYSTEM IMPLEMENTATION AND RESULTS

The system implementation takes fan pages of birds on Facebook as the data resource through the API provided by Facebook to do Python program writing. When using the Facebook API, we need to register an APP robot provided to software developers by Facebook. This robot has access to Facebook API functions and we can use the key obtained after registration to get an access token. When capturing posts on Facebook, a large amount of feedback might be obtained, e.g. written posts, uploaded pictures, times, etc. The Python program first uses the time to differentiate, writing a program to capture data on Facebook fan pages at a fixed time every day. After the time is set, this should also exclude unnecessary posts. Because the times of posts on Facebook are not sorted

by time, we add an end value to prevent the program from an infinite loop. When executing data capture, the program will capture data page by page and classify it into categories according to posts and messages. It uses posts where the content is shared directly by users and then extracts messages which may sometimes be too long or contain massive data from posts. Then data in the posts is matched with stored bird data in the database. If it matches the data in the database, it will be uploaded to the database and stored. When doing the match, it will also screen typos, and typos which are recognizable are considered passed. The system implementation adopts websites with photographic guides to birds to build the database. Through classifying this photographic data and uploading it to the database, bird-related data is displayed in various forms, including according to the bird name, family/genus/species (taxonomy), nicknames and so on. When capturing data, it also divides it according to tags to obtain correct data and these data are captured separately and uploaded to the database. This system implementation uses Linux to construct the Server, and classifies and then stores the captured data, including images and general bird-related data, and launches a Web service to enable users to watch image data as shown in the Web displays in Figures 2 and 3.

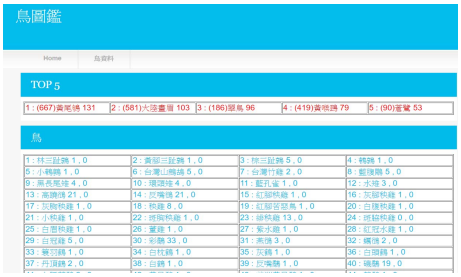

Figure 2: Web Display TOP 5

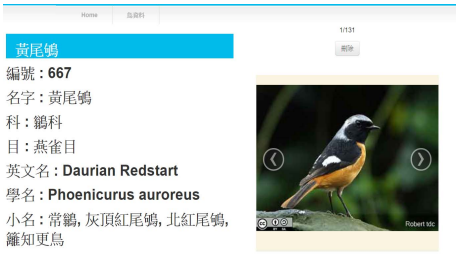

Figure 3: Web Display Bird Data

The system is divided into three main parts. The first part is the crawler process. Firstly, we collect the images and data from the users uploaded to the social network. The first part is the elementary classification of the data filter, through which we get to the special format information for convenient analysis and application of the data. If the filter confirms that the data is to be retained, the database is uploaded, and then classification is done. The second part is web application, the main part of which is to look and artificially review. Artificial review using the human eye is needed because machine learning samples do not have enough status, and cannot be used to do machine learning to train a powerful recognition module. So first filtering is done by artificial review and results in some rewriting to remove error data information. The system flowchart is shown in Figure 4.

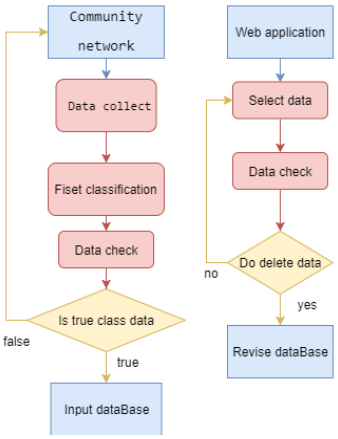

Figure 4: System flowchart

IV. CONCLUSION

For database integration, we need to analyze the entire storage framework for big data analysis in the future. In the future, data in the database will be analyzed and applied, for example, using image recognition and neural networks to train a more powerful identification network which can constantly expand the sample database via capturing programs. In this way, we can create a neural network that can not only identify birds but also identify all kinds of objects.

ACKNOWLEDGMENT

We thank the Ministry of Science and Technology of Taiwan for supports of this project under grant number MOST106-2221-E-239-036. We thank co-authors and reviewers for their valuable opinions.

REFERENCE

[1] R. Roberts, P. Donald and I. Fisher, Wordbirds: Developing a Web-based Data Collection System for the Global Monitoring of Bird Distribution and Abundance, Biodiversity and Conservation, Vol. 14, No. 1, pp. 2807-2820, 2005.

[2] E. Martin and G. Ballard, Data Management Best Practices and Standards for Biodiversity Data Applicable to Bird Monitoring Data, U.S. North American Bird Conservation Initiative Monitoring Subcommittee, 2010.

[3] J. Gustavo, J. Klarevas-Irby, A. Maldonado-Chaparro, I. Mueller and D. Farine, An Automated Barcode Tracking System for Behavioural Studies in Birds, Cold Spring Harbor Laboratory, bioRxiv, 2017.

[4] R. Yoshihashi, R. Kawakami, M. Iida and T. Naemura, Construction of a Bird Image Dataset for Ecological Investigations, IEEE International Conference on Image Processing, pp. 1-5, 2015.

[5] M. Barrueto, A. Clevenger, B. Dorsey and A. Ford, A Better Solution for Photo-classification, Automatic Storage and Data Input of Camera Data from Wildlife Crossing Structures, International Conference on Ecology and Transportation, pp. 1-11, 2013.