CARDIFF UNIVERSITY

School of Computer Science and Informatics

Using social media to observe wildlife distribution in the UK

CM3202 – One Semester Individual Project (40 Credits)
Initial Plan

William Dunn – c1317264
Supervised by Chris B Jones
Moderated by Ralph Martin

1.0   Project Description

The aim of this project is to utilise data taken from geo-tagged images on social media platforms to store wildlife sightings in a data warehouse for research purposes. To begin with selected samples of data will be extracted from social media platforms using available APIs and stored appropriately. The use of wildlife citizen science groups and keyword searches will aid in selecting samples of wildlife species with an ample amount of geo-tagged data. If time allows a larger, more broad data set will be mined and used.

By performing complex geo-spatial queries of the data warehouse historical data will be returned to aid in determining whether certain wildlife species are more commonly spotted in specific parts of the UK during the year. The results of these queries could help determine information relating to animal migrations, hibernation, dwindling population, balance between numbers of predators and prey and much more if interpreted efficiently.

The project will make use of spatial computing methods, such as kernel density estimation [1], to generate significant clusters to derive useful information. Padraig Corcoran and Christopher Jones have completed utilised similar methods to visualise the movement of a swarm of fish in a pond using persistence landscape methods [2]. Time slices will then be taken and used in the persistence landscape methods to evaluate how these clusters compare seasonally (be that animal cluster movement due to migration, or a plant cluster disappearing and reappearing with the seasons). The main programming language for the analysis of spatial data will most likely be python.

This project will predominately be focused on research and what can be interpreted from complex geo-spatial queries of collected spatial data. If time constraints allow a simple front end GUI for the data warehouse will be produced to improve ease of interpretation and make the system easily accessible to nature conservationists who could benefit from such a system e.g. A research group called Scottish Squirrel Group researching a cure for a disease called squirrel pox, a leading killer of the native red squirrel [3], could utilise the GUI to locate approximate areas of red squirrels for vaccination. The GUI would include functionality to view and compare data from different species sightings and different time periods, display data in an intuitive raster format, and allow users to view the specific photos that were taken in each location. The GUI would be developed using JavaScript.

2.0   Project Aims and Objectives

The following sections state the main aims and objectives of the project, which are further split into main and desirable sub categories. The aims and objectives are also roughly in order of completion, a more detailed order of completion can be found in section 3.0 of this report entitled work plan.

2.1 Main Aim and Objectives

- Research geo-spatial data analysis techniques
  - Research kernel density estimation and how it can be implemented.
  - Research clustering implementation and other spatial analysis options.

- o Research tools and programming languages that have built-in useful features that suit the requirements detailed in the project aims and objectives.

- Gain a basic understanding of wildlife patterns and behaviour, and citizen science sites to help select popular species to research.
    - o Study migration patterns, hibernation habits and general behaviours of particularly interesting and available species.
    - o Gain understanding of the rough location of particularly interesting animals.
    - o The purpose of this aim is to improve my ability to notice particularly interesting results from my spatial queries.
    - o Locate and analyse citizen science sites and cross reference popular species with image quantity on social media platforms to ensure there is enough data to produce quality conclusions from my analysis.

- Collect and store datasets of geo-tagged data for several wildlife species.
    - o Collect geo-tagged images from social media by accessing the platforms APIs.
    - o Store it in an appropriate format that allows for geo-spatial queries to be executed (For example Oracle or MySql).

- Execute complex spatial queries to uncover interesting species behaviours.
    - o Query the data warehouse to retrieve interesting information about animal behaviour.
    - o Utilise clustering methods, and other spatial database techniques, to locate trends and useful information.
    - o Create time slices of data sets to compare wildlife patterns over a specified period (For example monthly or seasonally).

- Document findings in a detailed final report.
    - o Produce an extensive report detailing the results of my project, including research completed, cycles followed, samples of code, testing and evaluations of my work.

2.2 Desirable Aims and Objectives

- Develop a front-end GUI for the data warehouse.
    - o Plot results of a query on a map of the UK
    - o Offer ability to display data in raster format.
    - o Offer ability to compare results with historical data.

- Expand selective species datasets to a large, broader dataset including many species.
    - o Use data mining techniques to access the social media APIs frequently to extract a large set of data.
    - o Expand data area to from UK to Western Europe.

3.0 Work Plan

After an initial meeting in week 1 with my supervisor we have decided it would be beneficial to the project to meet up for updates every week on a Friday (if our schedules allow). This is to discuss the progress of the previous week and to discuss what the focus of the following week should be. The *Week by Week Plan* will be primarily used for deciding the next week's focus, however physical meetings will help to reorganise any missed goals or aims. Between meetings further communication will be done via email.

Padraig Corcoran may attend some supervisor meetings so to help provide insight into spatial analysis methods and the mathematical processes involved.

In the work plan I have been intentionally vague when describing goals for each week, this is due to some implementation and analysis decision not yet being know until further research has been completed in week 2, and due to the unpredictability of how long each step will take.

I have also included the desirable objective of creating a front-end GUI for the database within the work plan. The work plan below does not consider any setbacks that may arise and so if all is successful the GUI should be achievable.

Pre-Week 1
20/10/2016 - 20/01/2017

- Research a potential project involving the combined fields of spatial databases and nature conservation.
- Meet with and propose initial project ideas with a potential supervisor.

Week 1
23/01/2017 – 29/01/2017

- Meet with supervisor to further discuss the project, decide upon weekly meeting times, and discuss the aims of the next week of the project. *Deliverable*: A list of meeting times with supervisor, and a clear plan for the first few weeks of work.
- *DEADLINE 1*: Complete the Initial Plan.

Week 2
30/01/2017 – 05/02/2017

- Research tools that will aid with identifying valuable data. *Deliverable*: A list of tools including data storage and data analysis programs.
- Research spatial analysis methods and mathematical processes that will aid with identifying valuable data. *Deliverable*: A list of analysis methods.
- Research citizen science groups to decide upon several wildlife species to collect initial data for. Compare citizen science research with amount of geo-tagged images on social media. *Deliverable*: A list of suitable wildlife to study with data to back up reasoning for decision of each species
- Research appropriate social media platforms to retrieve data from. *Deliverable*: A list of appropriate social media platforms to extract geo-tagged info.

Week 3
06/02/2017 – 12/02/2017

- Extract several datasets for different wildlife species from social media platforms including information detailing the geo-reference of the data. *Deliverable:* Several wildlife species data sets.
- Store datasets in a format appropriate for performing geo-spatial analysis. Must include support for complex spatial data types, spatial query operators, and spatial indexing [4]. *Deliverable:* A database containing several datasets of geo-referenced data.

Week 4 – Week 6
13/02/2017 – 05/03/2017

- Analyse stored datasets using an appropriate programming language that can perform spatial operations. *Deliverable:* A document containing each analysis method performed on each data set.
- Establish several time slices for each dataset to view historical changes to the data. *Deliverable:* Document each time slice taken for each wildlife dataset.
- Create several static visualisations of the data found. *Deliverable:* Document several static visualisations of significant results from data sets.

Week 7
06/03/2017 – 12/03/2017

- Determine suitable development tools for implementing a simple front-end GUI tool for the data store. *Deliverable:* A list of potential tools and programming languages to use in order to create a front-end GUI.
- *Deliverable:* Create a list of functional requirements for GUI tool.
- Decide on a development cycle to follow for the GUI tool implementation. *Deliverable:* Create a document comparing different development cycle options with comparisons, and a final decision.
- Research chosen development tools to gain a practical understanding of how it works and how it caters to my requirements. *Deliverable:* Complete online tutorials and classes to sharpen knowledge of selected tool.

Week 8
13/03/2017 – 19/03/2017

- Limited access to computer during week 8 as out of the country.
- Start work on first draft of final report. *Deliverable:* Complete first draft of Introduction, Background, and Approach.

Week 9 – Week 11
20/03/2017 – 09/04/2017

- Develop a simple front-end tool GUI which satisfies functional requirements defined previously. *Deliverable:* A front-end GUI to aid in analysis of wildlife geo-referenced data sets.

Week 12
10/04/2017 - 16/04/2017

- Finish the GUI tools if not completed. If completed evaluate remaining time until deadline 2 and the possibility of adding new functionality to the GUI.
- Utilise the newly created GUI tool to look for further data trends in datasets. *Deliverable:* Documents all research methods completed using GUI tool and any results found.

Week 13 – Week 15
17/04/2017 –07/05/2017

- Continue work on the first draft of the final draft. *Deliverable:* Complete first draft of Implementation, Results and Evaluation, Future work, Conclusions, and Reflection on Learning.
- Complete final draft of Final Report. *Deliverable:* Ensure final report is up to standard and that grammar and spelling are perfect.
- *DEADLINE 2*: Complete the Final Report

4.0   References

[1] O'Sullivan, D. and Unwin, D.J. (2010) *Geographic information analysis*. 2nd edition. Wiley, John & Sons.

[2] Corcoran, P. and Jones, C. (2016 *Spatio-Temporal Modeling of the Topology of Swarm Behavior with Persistence Landscapes*. ACM SIGSPATIAL 2016 International Conference on Advances in Geographic Information Systems.

[3] Commission, F. (2010) UK red squirrel group - Scottish squirrel group. Available at: http://www.forestry.gov.uk/fr/infd-8c8blh (Accessed: 25 January 2017).

[4] Jones, C. (2016 *Spatial Database Management with Object Relational Databases)* Large Scale Databases Lecture, Slide 7.