Tag2Blog: Narrative Generation from Satellite Tag Data

Kapila Ponnamperuma Advaith Siddharthan Cheng Zeng Chris Mellish
Department of Computing Science
University of Aberdeen
{k.ponnamperuma, advaith, c.zeng, c.mellish}@abdn.ac.uk

René van der Wal
Aberdeen Centre for Environmental Sustainability (ACES)
University of Aberdeen
r.vanderwal@abdn.ac.uk

Abstract

The aim of the Tag2Blog system is to bring satellite tagged wild animals "to life" through narratives that place their movements in an ecological context. Our motivation is to use such automatically generated texts to enhance public engagement with a specific species reintroduction programme, although the protocols developed here can be applied to any animal or other movement study that involves signal data from tags. We are working with one of the largest nature conservation charities in Europe in this regard, focusing on a single species, the red kite. We describe a system that interprets a sequence of locational fixes obtained from a satellite tagged individual, and constructs a story around its use of the landscape.

1 Introduction

We present a system, Tag2Blog, that uses Natural Language Generation (NLG) in bringing up-to-date information about wild animals in their natural environment to nature enthusiasts. We focus on the reintroduction of the red kite to the UK. The red kite, a member of the raptor family, has been persecuted to near extinction in the UK. Since 1989, efforts have been underway to reintroduce the species across the UK with mixed success. Where less successful, illegal activities of humans are partly responsible (Smart et al., 2010).

We are working with the RSPB[1], one of the largest nature conservation charities in Europe, around a reintroduction site where the species struggles to get re-established. We propose to use NLG for public engagement around a small number of satellite tagged individuals. The nature conservation goal is to create a positive perception of the species through informative blogs based on the movements of individual birds. The NLG goal is the generation of these blogs; specifically, to put individual locations of a bird into an ecological context. This paper describes the design and implementation of the system. We are also carrying out concurrent ecological research on red kites that will further inform the NLG component.

2 Related work

There is increasing realisation of the potential of digital approaches, including the use of websites and social media, to increase public engagement with nature conservation issues. For instance, in the UK, the Open Air Laboratories (OPAL) network[2] is a large initiative led by Imperial College, which aims to create and inspire a new generation of nature-lovers by getting people to explore their local environment (Silvertown, 2009). Such initiatives are typically labour and time intensive, and require continual effort to maintain interest through the creation of new content. To date, initiatives such as OPAL have largely focused on biological recording as a public engagement tool, thereby using - for example - standard social networking sites to prompt the collection of species distributional data (Stafford et al., 2010), or web interfaces that use NLG to provide feedback to citizen scientists (Blake et al., 2012).

We propose something altogether different: the use of sensor data as a starting point for public engagement through the delivery of self-updating automatically generated blogs. This application provides fresh challenges for the field of NLG, where typically systems are designed to offer decision support in the workplace (Goldberg et al., 1994; Portet et al., 2009). Decision support requires accuracy and clarity first and foremost. We, on the other hand, aim to generate texts that are suffi-

[1] http://www.rspb.org.uk

[2] http://www.opalexplorenature.org

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 169–174,
Sofia, Bulgaria, August 4-9 2013.

Figure 1: Screenshot of the Tag2Blog system.

ciently fluent and engaging for the general public to be attracted and informed by.

This does not mean that there is no precedent to our work. There are a handful of NLG systems that offer "info-tainment", such as Dial Your Disc (Van Deemter and Odijk, 1997) and Ilex (O'Donnell et al., 2001). Systems that generate sports commentary are particularly relevant, as they contextualise objects spatially and temporally and track the movement of objects as part of the game analysis (André et al., 2000). Rhodes et al. (2010) further explore dramatic narrative generation, to bring emotional content into the texts.

We subscribe to the same goals, adding to these the requirement that texts should be easy to read. For instance, ecological concepts (such as site fidelity) could be communicated by explicitly defining them. However, we would prefer these to be inferred from more engaging narratives, such as that in Fig. 1, which is a screenshot showing sample text generated by our system.

3 System architecture

The aim of the Tag2Blog system is to bring satellite tagged individuals of a species (e.g., the red kite) "to life" by constructing narratives describing their movements. In this regard, we need to interpret a sequence of locational fixes obtained from a tagged bird, and construct a story around its use of the landscape. To facilitate ecological interpretations, it is important to first supplement the locational data with other spatially relevant data; for example, landscape features and

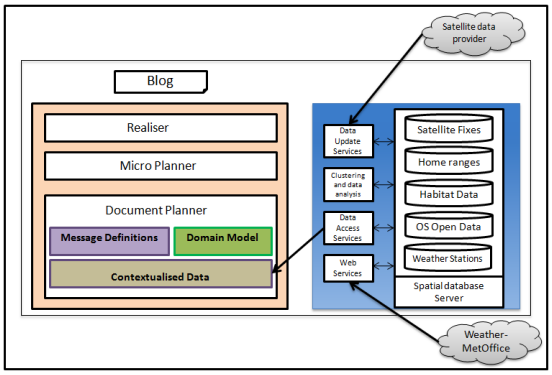


Figure 2: Architecture of the Tag2Blog system

weather. The Tag2Blog system therefore consists of two modules: Data Acquisition and Contextualisation (DAC), described in §3.1 and Natural Language Generation (NLG), described in §3.2.

3.1 Data acquisition and contextualisation

This module is composed of a spatial database and a set of services for updating and accessing data. We start with the information obtained from the satellite tags on the birds, which provide time-stamped locational information. This is augmented with data of associated habitat types, terrain features, place names and weather conditions. Our database thus stores rich information about the locations visited, acquired from a variety of sources summarised below:

Habitats: Land cover maps[3] are used to associate different habitat types (e.g., coniferous woodland, moorland, improved grassland, etc.) to

[3]http://www.ceh.ac.uk

locational fixes.

Terrain features: Ordnance Survey Vector Map data[4] are used to identify features (e.g., lochs, rivers, roads, etc.) in the vicinity of the fixes.

Names: Ordnance Survey Gazetteer data is used to obtain place and feature names.

Weather: The closest weather station to the fix is queried for historical weather data from the time of the fix, using an external web service.

The following services were implemented to update and enrich red kite location fixes:

Data update service: The satellite tags on the red kites have been programmed to transmit up to 5 GPS fixes per day, usually every two hours between 8am and 6pm[5]. The satellite data provider sends a daily email, using which we update the spatial database with red kite locations automatically. We also provide the conservation charity with a user interface, to allow them to censor ecologically sensitive locations (such as nesting sites), as and when required.

Data analysis service: Location data of each individual bird is periodically clustered (i.e., weekly) to identify their temporary *home ranges*. These clusters are spatially represented as ellipses and are stored in the database so that new fixes can be compared against known locational patterns.

Weather web service client: Weather data relevant to the time and location of each red kite locational fix is obtained on demand from a met office web service by providing the date, time, and the closest weather station.

Data access service: Each satellite fix is associated with a Java object (GeoLocation), which encapsulates the enriched data (habitats, place names, features, weather, etc.) for that location. Apart from individual locations, overall fight parameters such as distance from geographic features, displacement from or presence within known home ranges, are also computed and encapsulated into a Java object. These objects are generated on demand and passed onto the NLG module, described next.

3.2 Natural language generation module

The Tag2Blog system follows the NLG architecture proposed by Reiter and Dale (2000) and is composed of three components: a document planner (§3.2.2), a microplanner (§3.2.3) and a surface realiser (§3.2.4). The document planner utilises a domain model (§3.2.1) to populate and order message definitions, which are in turn passed on to the microplanner for creating sentence specifications. The surface realiser then generates text from these sentence.specifications.

3.2.1 Domain model and data analysis

The enriched data, as described above, be used as such to generate narratives of journeys. However in order to make these narratives insightful, an ecological interpretation is needed, and kite behaviours must also be included in the domain model. Siddharthan et al. (2012) has identified key behaviours that can be fruitfully communicated through such narratives. We broadly categorise these behaviours into:

- Site fidelity and exploratory behaviour
- Feeding and roosting behaviour
- Social behaviour (associations with other red kites)

A domain model was developed to infer likely kite behaviours from the enriched data. To build the domain model, we used explicit and implicit knowledge elicitation methods, such as data analysis and interviews, annotations of NLG produced blogs by ecologists, and analysis of hand-written blogs by ecologists from source data.

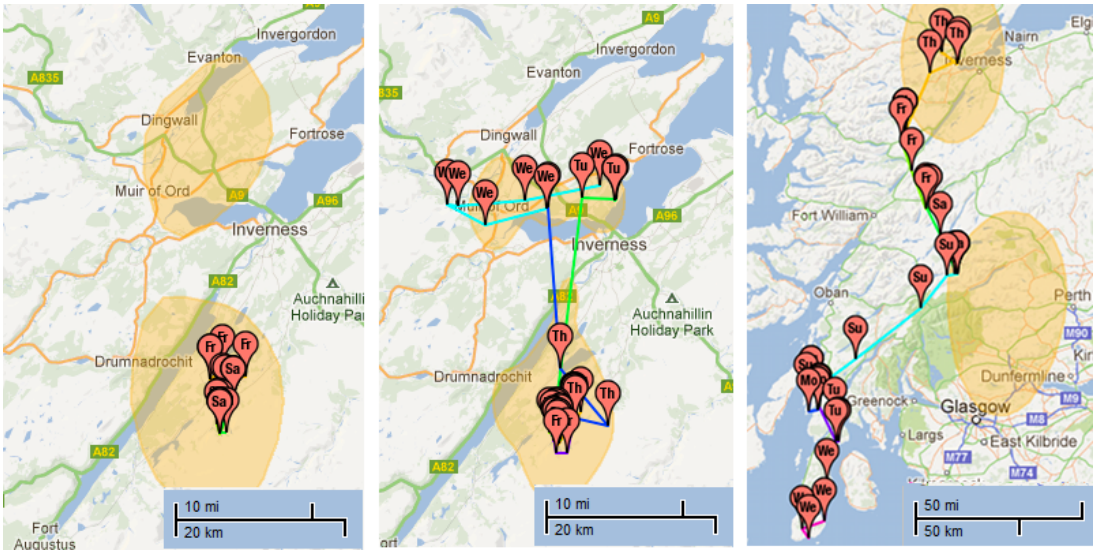
Site fidelity and exploratory behaviour: Historical location data is used to identify clusters (temporary home ranges) for each bird using the ADEHABITATHR[6] package (Calenge, 2006). In order to describe the overall movement pattern during the period, spatial data analysis is carried out and parameters, such as total distance travelled, displacement from clusters, percentage of fixes within each cluster, are calculated. These parameters are then used to identify the overall movement pattern. Fig. 3 shows three such patterns: Stationary, Short circular trip and Long distance movement.

Feeding and roosting behaviours: After conducting structured interviews with ecologists and analysing blogs written by ecologists, a set of rules were created to identify different feeding and roosting behaviours. Likely foraging patterns were defined on the basis of habitat type, season,

[4] http://www.ordnancesurvey.co.uk

[5] The satellite tags are solar powered, and only have power to provide a single fix per day in the winter months.

[6] http://cran.rstudio.com/web/packages/adehabitatHR



(a) (b) (c)

Figure 3: Movement patterns demonstrated in different weeks by different birds: (a) Stationary, staying within the temporary home range, (b) Short circular trip, moving out and returning to the temporary home range, and (c) Long distance movement, ending in a different location. The orange areas represent clusters of locations of the red kite from historical data that model temporary home ranges for the bird.

time of the day and weather conditions. The following extract from a blog written by an ecologist shows how experts can infer a behaviour from data. Note that it is acceptable in our application for such behaviours to be speculative, as long as they have a basis in kite ecology, and are plausible given the data.

> "Early that evening she was seen in farmland near Torness. Here, the rain must have brought up earthworms to the surface snacks well worth catching!"

From this text, we inferred the following rule:

> ***Rule: Feeding on Earthworms***
> IF it previously rained AND habitat is farm land, THEN it is likely that the red kite is feeding on earthworms.

We have expressed a range of such behaviours as JBoss[7] rules.

Social behaviours: Red kites being social birds, there are many social interactions that could be inferred from the type of data we brought together. Associations between red kites are typically inferred by analysing relative locations of different red kites. However, there is one specific behaviour, communal roosting, where a large group of red kites sleeps together in woodland during the winter months, for which we make use of our knowledge of known communal roost locations; i.e., local knowledge provided by ecologists.

3.2.2 Document planner

The document planner carries out content determination and document structuring.

Content determination: There are several types of message definitions, implemented as Java classes, that correspond to different narrative descriptions (flying, feeding, etc.). The message generator infers possible behaviours (feeding, roosting, exploring, etc.) using the domain model and then selects one or more based on content determination rules. For example, the message generator might infer possible behaviours such as feeding and exploring from the analyses described above in §3.2.1. However, the content determination rules would prioritise exploring behaviours over feeding (due to their rarity) and hence generate a EXPLORINGMESSAGE, which contains the information required to generate a description of the exploration journey. Similarly, corresponding messages would be generated for other flying, feeding, and social behaviours.

Document structuring: Our weekly blogs contain an introductory paragraph, which captures the overall movement pattern for the week, followed by a more detailed paragraph, which describes interesting behaviours during that week. Each para-

[7] http://www.jboss.org/drools

graph is internally represented as a schema, which also orders the messages into a document plan.

3.2.3 Microplanner

The document plan generated at the previous stage is passed on to the microplanner for creating text specifications. This includes phrase specifications and their aggregation into sentences. Clauses are combined using discourse cues to express different discourse relations, such as concession, comparison and explanation.

3.2.4 Surface realiser

The role of the surface realiser is to convert the text specification received from the microplanner into text that the user can read and understand. This includes linguistic realisation (converting the sentence specifications into sentences) and structural realisation (structuring the sentences inside the document). Both the linguistic and structural realisations are performed by using functionalities provided by the SIMPLENLG realiser library (Gatt and Reiter, 2009).

4 Utility of blogs in this domain

Until recently, our partner charity was publishing hand-written blogs based on the journeys of these satellite tagged red kites. They have had to close down the site due to resource constraints: Such blogs are difficult, monotonous and time consuming to produce by hand. Tag2Blog will allow the charity to restart this form of public engagement.

We have earlier studied the use of ecological blogs based on satellite tag data (Siddharthan et al., 2012). Using hand-written blogs in a toy domain, we found that readers were willing to anthromorphise the bird, and generally formed a positive perception of it. Additionally, users were able to recall ecological insights communicated in the blog, demonstrating that such blogs are informative as well.

In this paper, we restrict ourselves to reporting a very preliminary evaluation of the quality of the computer generated blogs. We compared three blogs produced from the same data (the movements of one individual red kite during one week):

a) A computer generated blog of a journey, produced without using any domain knowledge as described in §3.2.1, and merely describing spatial movements of the red kite over time.

b) A computer generated blog of a journey with ecological insights, as described in §3.2.1. This is the production version used in Fig. 1.

c) Version (a), which has been post-edited by an ecologist to introduce ecological insights into the narrative. The ecologist was give access to a table containing habitat, terrain and weather information for each satellite fix.

Tab. 1 shows samples from the three versions. All three versions were shown to five human judges, without indication of provenance. They were asked to rate each blog on a scale of 1 (low) to 5 (high) for how readable, informative, engaging and ecologically sound they considered the texts. They were also asked to rate the relevance of each blog to different age groups (primary school children, secondary school children and adults).

We used as judges, a social scientist specialised in human–nature interactions, a public engagement officer at our University who interacts with local schools on a regular basis, a secondary school English teacher, and two school students, aged 14 and 16. Our goal was to obtain a diversity of opinion to inform system design.

Tab. 2 shows the ratings of our five evaluators for different aspects of each blog. The averages show that in most aspects, version (b) is rated higher than version (a) and, rather expectedly, the human edited/annotated version (c) is rated the highest. But, note that the two school students rated the automatically generated blogs highly, and that both felt that version (b) was the best suited for secondary schools. The public engagement officer rated (b) as less readable, and less relevant to schools. She specifically highlighted the use of terminology without introduction (e.g., “roost” and “foraging”) as an issue.

Our focus will now be on improving the language, to address some of the readability and engagingness concerns.

5 Conclusions and Future Work

We have presented an NLG system that can generate ecologically informative and engaging narratives of animal (red kite) movements. Our initial evaluations have shown encouraging results and further evaluations are now planned. The system can be accessed through http://redkite.abdn.ac.uk/blog/.

Acknowledgements

This research is supported by an award made by the RCUK Digital Economy programme to the dot.rural Digital Economy Hub (ref. EP/G066051/1). We also thank our partner organisation, the RSPB (with special thanks to Stuart Benn), who manage the reintroduction project.

Text	First four sentences from each blog
(a)	This week, Millie did not travel far, but was actively exploring a small area. During this week, Millie has been observed on various habitats. However, except Thursday she chose to spend the night at the same woodland near Torness. No doubt Millie was not alone this week as kites Moray and Beauly were also observed often in the vicinity.
(b)	This week, Millie did not travel far, but was actively exploring a small area mainly within her home range. During this week, Millie's foraging patterns have been varied. However, except Thursday she chose to roost in the same woodland near Torness. No doubt Millie had a quite social week as kites Moray and Beauly were also observed often in the vicinity.
(c)	This week Millie did not travel far but was actively exploring a small area north-east of Loch Ness. Friday morning Millie left the woodland where she spend the night to fly to Loch Ruthven amid heavy rain. The poor visibility may have driven her to fly low when searching for food along the water sides. Early that evening she was seen in farmland near Torness.

Table 1: Excerpts of texts in each experimental condition

	Sociologist			Pub. Eng.			Teacher			16yo			14yo			Average		
Blog	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**
Readability	3	3	5	4	3	5	3	2	4	3	4	4	3	4	4	3.2	3.2	4.4
Informativeness	3	4	5	5	5	5	2	1	2	3	4	5	3	3	4	3.2	3.4	4.4
Engagingness	2	4	5	3	3	4	2	1	3	3	4	5	2	4	4	2.4	3.2	4.2
Ecological soundness	4	3	3	4	4	4	5	5	5	3	4	4	3	4	3	3.8	4.0	3.8
Relevance to:																		
Primary Schools	3	4	5	3	2	4	4	4	4	4	4	3	3	2	3	3.4	3.2	3.8
Secondary Schools	3	4	5	4	3	4	2	2	2	4	5	3	3	4	3	3.2	3.6	3.4
Adults	3	4	5	4	4	4	3	1	3	3	4	5	3	4	4	3.2	3.4	4.2

Table 2: Evaluation of Blogs by Experts

References

E. André, K. Binsted, K. Tanaka-Ishii, S. Luke, G. Herzog, and T. Rist. 2000. Three robocup simulation league commentator systems. *AI Magazine*, 21(1):57.

S. Blake, A. Siddharthan, H. Nguyen, N. Sharma, A. Robinson, E. O Mahony, B. Darvill, C. Mellish, and R. van der Wal. 2012. Natural language generation for nature conservation: Automating feedback to help volunteers identify bumblebee species. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 311–324.

C. Calenge. 2006. The package adehabitat for the r software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197:1035.

A. Gatt and E. Reiter. 2009. SimpleNLG: A realisation engine for practical applications. In *Proceedings of the 112th European Workshop on Natural Language Generation (ENLG)*, pages 90–93.

E. Goldberg, N. Driedger, and R.I. Kittredge. 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2):45–53.

M. O'Donnell, C. Mellish, J. Oberlander, and A. Knott. 2001. Ilex: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(3):225–250.

F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7-8):789–816.

E. Reiter and R. Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

M. Rhodes, S. Coupland, and T. Cruickshank. 2010. Enhancing real-time sports commentary generation with dramatic narrative devices. In *Proceedings of the Third Joint Conference on Interactive Digital Storytelling, ICIDS 2010*, Lecture Notes in Computer Science, pages 111–116. Springer.

A. Siddharthan, M. Green, K. van Deemter, C. Mellish, and R. van der Wal. 2012. Blogging birds: Generating narratives about reintroduced species to promote public engagement. In *Proceedings of the Seventh International Natural Language Generation Conference (INLG)*, pages 120–124.

J. Silvertown. 2009. A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9):467–471.

J. Smart, A. Amar, I.M.W. Sim, B. Etheridge, D. Cameron, G. Christie, and J.D. Wilson. 2010. Illegal killing slows population recovery of a reintroduced raptor of high conservation concern–the red kite *milvus milvus*. *Biological Conservation*, 143(5):1278–1286.

R. Stafford, A.G. Hart, L. Collins, C.L. Kirkhope, R.L. Williams, S.G. Rees, J.R. Lloyd, and A.E. Goodenough. 2010. Eu-Social Science: The Role of Internet Social Networks in the Collection of Bee Biodiversity Data. *PloS one*, 5(12):e14381.

K. van Deemter and J. Odijk. 1997. Context modeling and the generation of spoken discourse. *Speech Communication*, 21(1-2):101–121.