

MULTIMODAL PUZZLES IN A NATURAL HISTORY ARCHIVE

Historian Andreas Weber uses AI techniques to understand the secrets of the natural history archive of the *Natuurkundige Commissie* (Natural History Committee). 'For computer scientists, this is highly challenging material.'

By **Reineke Maschhaupt**
Images Collectie Naturalis

In 1820, King William I established the *Natuurkundige Commissie*. Eighteen scientists were sent to the Dutch East Indies to study the local flora and fauna there. That yielded an incredible amount of archive material, which is now stored at Naturalis Biodiversity Center in Leiden. Andreas Weber is assistant professor of Science, Technology and Culture at the University of Twente, where, amongst other things, he is working on the project "Making Sense of Illustrated Handwritten Archives". He has yet to make a spectacular find, such as an extinct animal, says Weber laughing. However, the team did manage to enrich the archive and make it digitally accessible. In collaboration with academic publisher Brill, the entire archive has been available online since the end of November and can be searched by everyone free of charge.

Linking animal to diary

Weber: 'Worldwide, this is the most complete nineteenth-century archive of biodiversity in South-East Asia. It contains 20,000 scanned pages with handwritten texts, illustrations, publications and thousands of objects including stuffed animals, specimens on alcohol, skins, skeletons and eggs.' 'Archives like this are only meaningful if you view them in their multimodality. For example, the nameplate of a stuffed crocodile just states Borneo as its origin. Only after linking the object to the diaries and the field notes, can you discover exactly which river that animal was found in. There once did exist links between the images, text and objects in the archive, but many of those have been lost. We have tried to reconstruct those links with the help of learning and semantic computer techniques.'

Interaction

The project made use of MONK, a handwriting recognition system that actively learns. 'The system divides a text into lines. Then it starts to label the words, interpret these and to give meaning with levels of probability. Twelve volunteers helped to train the system. At a certain point, a self-learning process is initiated. However, people need to continuously point the system in the right direction in that visual labyrinth.'

The team added metadata previously recorded by Naturalis to this handwriting recognition technique. For each scan, somebody has once documented what was on a drawing and who had produced it. Finally, a semantic system was linked to this procedure. Now, if you come across a scientific animal name, the system can link this to the overarching group.

Collaboration

Collaboration between the three disciplines of computer science, biology and history is vital to smartly digitalise such a diverse archive. 'The computer scientists, for example, found it difficult to deal with the fact that scientific names for animals changed over the centuries. Then you need biologists who can look very carefully at these images to identify which animal it is.'

Archives like these still contain a vast amount of hidden information. Together with colleagues, Weber is now working on a large new grant proposal for AI and digital heritage. 'These techniques can also play an important role for other collections. For example, you can link objects in museums with handwritten documents or with ships' passenger lists. We want to make further use of this knowledge.'



Red-throated barbet, drawn by **Pieter van Oort** in Buitenzorg (Java) in 1827.

The archive is available via:
dh.brill.com/nco/