Sigfrid INGRISCH, Klaus RIEDE, Karl-Heinz LAMPE & Christian DIETRICH

DORSA - A “Virtual Museum” of German Orthoptera Collections

Abstract - DORSA (German Orthoptera Collections) is a specimen based database with internet access to the Orthoptera held in German museum collections including geographic information on a world-wide basis as well as media data like illustrations of type specimens and sound recordings ("Virtual Museum"). A rapid assessment tool is developed for automated song recognition.

Riassunto - DORSA, un "Museo Virtuale" delle collezioni di Ortotteri conservate in Germania. DORSA (German Orthoptera Collections) è un database con accesso via Internet, che consente di conoscere gli esemplari di Ortotteri conservati nei Musei tedeschi; esso include informazioni geografiche su base mondiale ed ulteriori dati, quali illustrazione dei tipi e registrazioni delle stridulazioni ("Museo Virtuale"). E' inoltre sviluppato uno strumento di rapido uso per il riconoscimento dei suoni.

Key words: Orthoptera, specimen database, type catalogue, photographic documentation, insect sounds, virtual museum, automated song recognition.

INTRODUCTION

The importance of digitised worldwide available data for the future development of taxonomy was just recently emphasized (Godfray 2002a, b). The Orthoptera are one of the few insect groups for which complete digital registers of valid taxa exist (Orthoptera Species File Online: http://OSF2.orthoptera.org/basic/HomePage.asp). Nevertheless, much of the information stored within museum collections is still not readily accessible, because most of it is not digitised (Lampe & Riede 2001).

DORSA (abbreviation for Deutsche Orthopteren-Sammlungen = German Orthoptera Collections http://www.dorsa.de) is part of EDIS, the Entomological Data and Information System (http://www.insects-online.de) launched by the German Federal Ministry of Education and Research. EDIS is intended to provide access to relevant information about insects and other arthropods for research and the general public and to promote digitisation of the data scattered over various entomological collections in Germany in a comprehensive database, the SYSTAX database at the University Ulm (http://www.biologie.uni-ulm.de/systax).

The aim of DORSA is the database access to important Orthoptera (grasshoppers, katydids and crickets) specimens in German research collections including multimedia information as pictures and sounds, and an internet-link to the global species database ‘Orthoptera Species File’ (OSF).

All information in DORSA is specimen based. It is thus possible to reproduce where single specimens have been collected, how they looked like, what was their stridula-
tion etc. DORSA is thus complementary to the Orthoptera Species File where the taxonomic background and general information on the species can be found.

In more detail, DORSA covers the following points:

1. **Registration of Type Specimens.** Registration of type specimens was done in the major German natural history museums that hold large Orthoptera collections (ZMHB Berlin, DEI Eberswalde, SMTD Dresden, ZSMH Hamburg, MLUH Halle, ZFMK Bonn, SMFD Frankfurt, SMNS Stuttgart, ZSM Munich). The Zoological Museum of the Humboldt-University Berlin houses by far the greatest number of Orthoptera types.

   The aim of specimen registration is a type catalogue of the Orthoptera housed in German collections and to verify the corresponding information in the Orthoptera Species File Online (above) which is the taxonomic backbone for our database. Compiling the specimen information in the museums (Tab. 1) covered all primary types, most of the paratypes, additional specimens on which taxonomic revisions were based and voucher specimens for which sound recordings are available. We started with the information given in the Orthoptera Species File (OSF) [version April 2000] which contains information on primary types compiled from published data (presumption in Tab. 1). With this information we checked what is actually housed in the museum collections (museum data). It followed a validity check. In case of disagreement between the information in OSF and the actual museum data, it was necessary to consult the original publications. Finally the databases were updated as was done with the specimen information in the museums when necessary. After the validity check, the inventory data are imported into the all-biology database SYSTAX. In this database, interested users can find the specimen information together with specimen based multimedia data (images and sounds). Moreover, DORSA is linked on a species to species level with OSF2 thus that additional taxonomic information, synonyms, references, will be available by a link from DORSA to OSF, while when searching OSF it will be possible to find specimens and their media data in DORSA.

2. **Photographic and Acoustic Documentation.** The aim is to create a "virtual museum" that documents the type specimens and helps to reduce the inquiries about and the loan of type material. Approximately 25000 photographs have been taken. Photographic documentation includes specimen labels, general habitus of the specimens, details of diagnostic characters, and also genital structures if this did not require manipulation of the specimens. The images are provided with size scales. All media data are linked to a specific specimen. Processing of the media data is time consuming. For most species, full documentation is already available in SYSTAX.

   Sound recordings of Orthoptera were usually done by private researchers and were not previously compiled in a public sound archive. DORSA establishes the first comprehensive public sound library for Orthoptera sounds in Germany, and makes them available on the internet. More than 4000 recordings are so far compiled that are linked with known individual specimens. They were contributed by K.-G. Heller, S. Ingrisch, F. Nischk, K. Riede, and G.H. Schmidt; others will follow. The sound library is the base for the development of software for automatic song recognition.
Table 1. Registration of type specimens of Orthoptera without Tridactyloidea in German museum collections by December 2002.

<table>
<thead>
<tr>
<th>museum</th>
<th>presumption taxed with primary types in OSF*</th>
<th>taxa with primary types checked</th>
<th>taxa with types including paratypes</th>
<th>number of type specimens</th>
<th>total recorded species</th>
<th>total recorded specimens</th>
<th>type data in OF not confirmed in museum</th>
<th>primary types lost: taxa</th>
<th>primary types not listed in OF</th>
<th>no taxon entry found in OF</th>
<th>unlabeled, newly recognized prim. types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berlin</td>
<td>1093</td>
<td>1423</td>
<td>1725</td>
<td>3646</td>
<td>1937</td>
<td>4192</td>
<td>15</td>
<td>330</td>
<td>68</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Eberswalde</td>
<td>45</td>
<td>79</td>
<td>136</td>
<td>325</td>
<td>177</td>
<td>386</td>
<td>6</td>
<td>1</td>
<td>34</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Dresden</td>
<td>66</td>
<td>117</td>
<td>170</td>
<td>521</td>
<td>290</td>
<td>878</td>
<td>5</td>
<td>55</td>
<td>10</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Hamburg</td>
<td>142</td>
<td>135</td>
<td>260</td>
<td>435</td>
<td>314</td>
<td>549</td>
<td>4</td>
<td>55</td>
<td>43</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Halle</td>
<td>38</td>
<td>55</td>
<td>55</td>
<td>75</td>
<td>91</td>
<td>151</td>
<td>4</td>
<td>25</td>
<td>4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Bonn</td>
<td>6</td>
<td>44</td>
<td>48</td>
<td>152</td>
<td>50</td>
<td>156</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Frankfurt</td>
<td>82</td>
<td>151</td>
<td>176</td>
<td>332</td>
<td>263</td>
<td>540</td>
<td>69</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Stuttgart</td>
<td>47</td>
<td>112</td>
<td>129</td>
<td>505</td>
<td>153</td>
<td>590</td>
<td>65</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Munich</td>
<td>27</td>
<td>54</td>
<td>80</td>
<td>214</td>
<td>94</td>
<td>358</td>
<td>1</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1546</td>
<td>2170</td>
<td>2779</td>
<td>6205</td>
<td>3369</td>
<td>7800</td>
<td>30</td>
<td>57</td>
<td>654</td>
<td>107</td>
<td>39</td>
</tr>
</tbody>
</table>

(*) OSF version of April 2000 at the beginning of the DORSA project. - italics = incomplete data.
Fig. 1. Characters most suitable for the classification of cricket songs are the distances between pulses, the duration of pulses, and the frequency course within pulses. The feature vectors are classified individually and the results of the single classifiers are combined through classifier fusion.

3. AUTOMATIC SONG RECOGNITION. Based on the DORSA song database, a rapid assessment tool is developed for automated song recognition (Dietrich et al. 2002).

In principle, automatic song recognition in crickets (Grylloidea) and katydids (Tettigonioidae) follows the scheme:
1. pre-processing and filtering of the raw time signals (fig. 1),
2. segmentation of the filtered signals into single pulses (fig. 1),
3. extraction of typical features of pulses or pulse series,
4. automated classification in local time windows (fig. 2),
5. application of classifier fusion:
   a) to combine the classifier outputs of different features,
   b) for temporal fusion of classifier outputs (fig. 2).
To train the neural network architecture successfully a data set of known taxa is needed. For that as many as possible songs of different individuals of the same species should be used. At present the tool for the automatic sound recognition runs properly with a set of 215 songs from 137 individuals of 30 cricket species (details in Dietrich et al. 2001a, b, 2002).

4. Geographic information. It is often difficult to find the exact locality for classical specimens as old types. The information given on labels may be either not precise or the names of places or their affiliation to countries has changed since the time of collection. Fortunately there are several expedition reports to allow tracing the localities with sufficient precision. For all specimens with the locality information sufficiently reliable, we geo-referenced the collection sites with latitude/longitude coordinates. This will help future users to find this information easily, and it allows mapping by a geographical information system (GIS). In the GIS the distribution of specimens or taxa can be overlaid with additional information such as environmental data.

A web-based GIS mapping tool is already available at our homepage (http://www.dorsa.de > link: News > DORSA-interactive map server). It is based on approximately 2,000 localities of katydid (Tettigonioidea) sound records made by K.-G. Heller. The interface allows geographic queries, retrieval and mapping of species data.
Fig. 3. GIS analysis with geographical coordinates from the DORSA database. Interface of the DORSA-interactive map server showing a map of collection sites of voucher specimens for sound recordings by K.-G. Heller using Platycleis s. str. as an example.
A: Map with specimen data from 5 species on a European scale. Each species is shown with a different symbol.
B: On a map zoomed to regional scale, individual points are attributed with the species names.
With this map server, either a single or several species can be selected; after refreshing the map, the localities of the specimens in the database are drawn on the map (fig. 3A). The map can be zoomed from worldwide to local level. At local level, localities are labelled with species names thus easily allowing to compare the localities of different taxa (fig. 3B).

NOTE ADDED IN PROOF

The actual numbers of digitised specimens and multimedia data can be found on the DORSA homepage http://www.dorsa.de

Meanwhile all specimen data are included in the SYSTAX database http://www.biologie.uni-ulm.de/systax and fully linked with the Orthoptera Species File Online: http://OSF2.orthoptera.org/basic/HomePage.asp

ACKNOWLEDGEMENTS

From April 2000 to March 2003, DORSA was funded by the German Federal Ministry of Education and Research. Our special thanks go to the curators of the museums mentioned above for their cooperation when digitising the specimen data. D. Eades (Illinois Natural History Survey) co-operated with for linking DORSA to OSF.

REFERENCES


Authors' addresses:
E-mail: sigfrid.ingrisch@planet-interkom.de;
k.lampe.zfmk@uni-bonn.de;
k.riede.zfmk@uni-bonn.de
C. Dietrich, Department of Neural Information Processing, University Ulm, Oberer Eselsberg, D-89069 Ulm, Germany. E-mail: christian.dietrich@neuro.informatik.uni-ulm.de