

Beitr. Ent.	Keltern	ISSN 0005 - 805X
55 (2005) 2	S. 463 - 469	27.12.2005

## **Earwigs-online: Vor- und Nachteile einer selbst entwickelten Internetdatenbank**

Mit 3 Figuren

FABIAN HAAS

### **Zusammenfassung**

Die Website [www.earwigs-online.de](http://www.earwigs-online.de) widmet sich ausschließlich den Dermaptera (Ohrwürmer) und basiert auf einer Reihe von Datenbanken, die über die letzten 10 Jahre entwickelt wurden. Die Vernetzung und Programmierung erfolgte entsprechend den Kenntnissen und Bedürfnissen des alleinigen Nutzers, d.h. des Autors. Als weitere wichtige Elemente wurden automatische Verschlagwortung und eine digitale Bibliothek aufgenommen. Derzeit umfassen die Datenbanken rund 16.000 Einträge zur Verbreitung, 4.000 Literaturstellen, 7.200 Schlagworte, 5.000 Taxonnamen und eine digitale Bibliothek mit 1.000 Zusammenfassungen und Artikeln. Bei der Arbeit mit den Datenbanken kristallisierten sich eine Reihe von Vor- und Nachteilen heraus. Nachteilig an einem vereinzelt Projekt sind die Belastungen durch die Programmierung, Datensicherung und Internetanbindung, geringere bzw. keine (finanzielle) Unterstützung durch Projektträger und Arbeitgeber, sowie die Nichtverfügbarkeit der Daten in größeren Portalen. Als vorteilhaft erwiesen sich die Verfügbarkeit aller Daten zu jeder Zeit, die schnelle Veröffentlichbarkeit und die freie Kombinierbarkeit der Daten. Der gewichtigste Vorteil ist sicher die reine Existenz dieser Datenbank: Da Dermapteren in Wissenschaft, Medizin und Landwirtschaft nicht im Focus des Interesses stehen, ist es unwahrscheinlich, dass größere Projekte und Institutionen den Aufbau dieser Datenbanken betrieben hätten. Die Fülle der jetzt verfügbaren Daten wäre sicher nicht zugänglich.

### **Summary**

The website [www.earwigs-online.de](http://www.earwigs-online.de) is dedicated exclusively to the Dermaptera (earwigs) and is based on a series of databases developed during the last ten years. The programming and integration of the databases proceeded together with the developing knowledge and requirements of the sole user, namely the author. Important incorporated features are an automatic facility for selecting key words and a digital library with 1,000 abstracts and articles. At present the databases contain approximately 16,000 entries on distribution, 4,000 literature citations, 7,200 key words and 5,000 names of taxa. As work proceeded, a number of advantages and disadvantages became clear. Disadvantages in a single project are the burdens of programming, safeguarding data and providing a link with the internet, together with small or absent financial support by sponsors and employers, as well as the non-availability of the data in larger portals. It proved advantageous to have the data constantly available, and to be able to publish these quickly in a freely combined form. The most important advantage is certainly the very existence of this database. Earwigs are not the focus of interest in zoology, medicine or agriculture. It is therefore unlikely that larger projects and institutions would have created these databases. A wealth of data would have remained unavailable.

### **Schlagworte**

Dermaptera, Ohrwurm, Datenbank, Earwigs-online, Internetdatenbank, Digitale Bibliothek

### Geschichte, Struktur und Leistungsfähigkeit

Die Ursprünge der Website Earwigs-online liegen etwa 10 Jahre zurück in den eigenständigen Datenbanken zur Verwaltung von Literatur, Sammlung und Diasammlung des Autors.

Ursprünglich wurden für Literatur und andere Daten verschiedene Programme benutzt. Erst mit zunehmender Nutzung wurden die Vorteile der Vernetzung der einzelnen Datenbestände erkannt, weshalb auch die Literaturverwaltung auf FileMaker, das benutzte Datenbankprogramm, umgestellt wurde. Seit dieser Umstellung wurden die Datenbanken immer enger vernetzt, um die einmal eingegebenen Daten an den verschiedensten Stellen in ihrer aktuellen Version vorliegen zu haben. Zum Umfang der aktuellen Version siehe Tabelle 1.

Eine Internetpräsenz wurde erst 2004 etabliert, wobei folgende Konstruktion benutzt wird. Die HTML-Seiten liegen beim Internetprovider (Strato), die eigentlichen Datenbanken aber auf dem Server der Sektion Biosystematische Dokumentation in Ulm, auf dem die Serverversion von FileMaker installiert ist.

Die Dateneingabe ist ausschließlich über den Autor der Website möglich, was neben der Qualitätssicherung noch andere Vorteile aufweist: einfachere Struktur und Zugangsberechtigung und keine Synchronisierungsprobleme. Daten können über Text- und andere Dateiformate importiert werden.

Ziel des Projektes "Earwigs-online" ist die Verfügbarmachung von Informationen zu allen Aspekten der Dermapteren, einschließlich einer Bestimmungshilfe durch umfangreiche Bilddatenbanken. Taxonomischer Katalog und Sammlungsmanagement stehen nicht im Zentrum des Interesses.

Tab. 1: Umfang der Earwigs-online Datenbanken mit Stand April 2005. NHM: The Natural History Museum, London, UK.

Thema	Datensätze Anzahl (Megabyte)
Indexkarten aus NHM	4.500
Einträge zur Verbreitung	15.027
Literaturzitate	3.739
Digitale Bibliothek	942 Zusammenfassungen und 160 Artikel (530)
Photos und Zeichnungen	10.027 (1.300)
Taxonnamen inkl. Synonyme	4.983
Spracheinträge	154 aus 35 Sprachen

Die umfangreiche Datengrundlage (Tab. 1) ermöglicht die sehr effiziente Bearbeitung verschiedenster Probleme aus dem Bereich der Dermapterologie. Literaturzitate liegt aus allen Aspekten der Biologie der Ohrwürmer, einschließlich agrarwissenschaftlicher Aspekte, vor und es kann weitgehende Vollständigkeit behauptet werden. Ebenso können durch die Aufnahme von Literatur- und Originaldaten von den in Museen vorhandenen Tieren zu jedem beliebigen Staat schnell und mit hoher Vollständigkeit, Faunenlisten erstellt werden. Mikkelsen & Bieler (2000) untersuchten diesen Ansatz für die Karibischen Mollusken und strichen dessen Effizienz heraus. Die Weboberfläche ist bewusst in englischer Sprache gehalten, um eine möglichst weltweite Nutzung zu ermöglichen (Fig. 1).

## Technik

Zur Programmierung der Datenbanken wurde ausschließlich das Programm FileMaker verwendet. Die Gründe dafür sind die Verfügbarkeit dieses Programms für Apple Macintosh und Microsoft Windows Betriebssysteme; das Programm ist weit verbreitet, d.h. eine gute Dokumentation (auch von freien Autoren) ist verfügbar; eine große Benutzergemeinde, die auch spezifische Probleme beantworten kann, steht bereit; und unabhängige Hersteller programmieren Zusatzprogramme (sog. ‚Plug-Ins‘). Eine Internetserver Version von FileMaker war von Anfang an verfügbar. Aufgrund der guten Positionierung der Firma, gerade im nichtakademischen Bereich, ist ein Verschwinden von FileMaker wenigstens für die absehbare Zeit unwahrscheinlich. All dies ist bei einigen Insellösungen an verschiedenen Museen nicht gegeben. Derzeit wird die Version FileMaker 6.0 Developer benutzt.

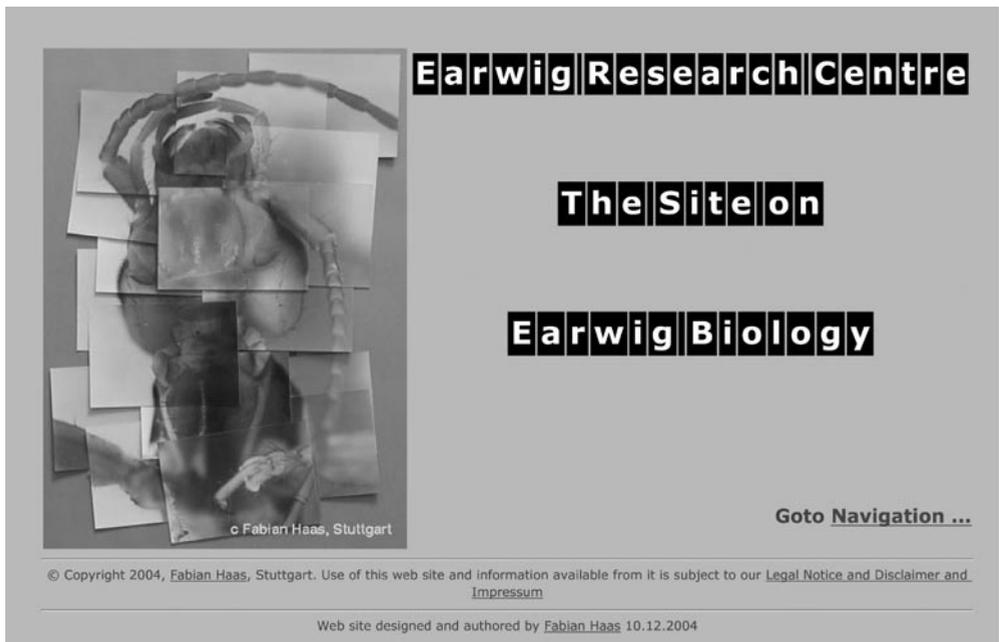


Fig. 1: Eingangsseite der Earwigs-online Website ([www.earwig-online.de](http://www.earwig-online.de))

## Automatische Verschlagwortung

Die Verschlagwortung der Artikel ist eine besonders aufwändige Arbeit, die aufgrund der wünschenswerten Konsistenz der Schlagwortvergabe auch nur schwer delegierbar ist. Neben der weiterhin durchgeführten manuellen Verschlagwortung wurde eine automatische Version entwickelt, die hier kurz, in allgemeiner Form beschrieben werden soll (Fig. 2).

Die automatische Verschlagwortung nimmt an, dass der Titel des Aufsatzes die kürzeste Zusammenfassung des Inhaltes ist, weshalb ausschließlich der Titel zur Analyse herangezogen wird (Zeitschriftentitel sind dafür nicht geeignet).

Der Titel wird in einem ersten Schritt in einzelne Worte zerlegt, und jedes einzelne Wort über eine relationale Verknüpfung mit einem Thesaurus verglichen. Dieser wurde speziell erstellt, und enthält wichtige Worte auch in verschiedenen Sprachen (z.B. Ökologie, ecology, écologie). Ist ein Wort im Thesaurus vorhanden, liefert dieser das entsprechende Schlagwort (in immer nur einer Form), das in ein Feld kopiert wird, wobei der Eintrag gleichzeitig mit "autoKeyword" gekennzeichnet wird. Dies ermöglicht, nach Erweiterung des Thesaurus, die komplette Löschung und Neuerstellung der automatischen Verschlagwortung. Dies ermöglicht auch die Bewertung der Schlagworte und das Erkennen von – selten vorkommender – falscher Vergabe von Schlagworten.

In einem zweiten Schritt wird der Titel in zweier Wortgruppen zerlegt, da bestimmte Inhalte nur durch Wortkombinationen ausgedrückt werden (z.B. "sp. nov.", "French Guyana" etc). Das weitere Vorgehen entspricht dem ersten Schritt.

## Automatische Verschlagwortung

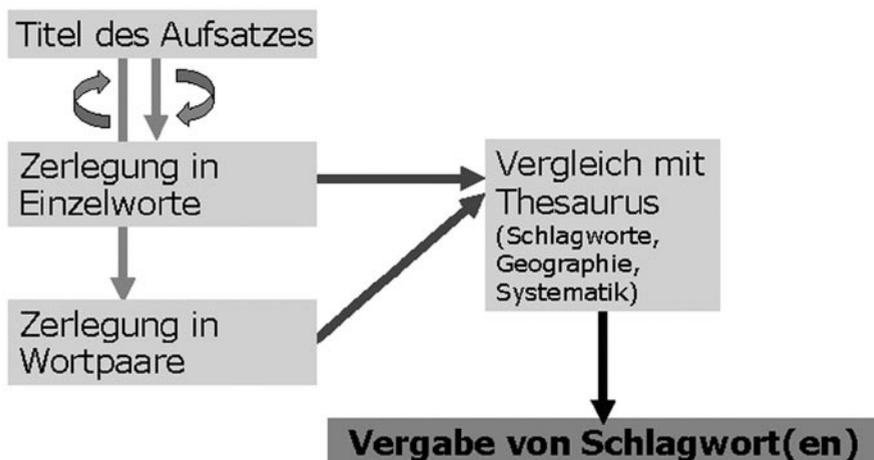


Fig. 2: Schematische Darstellung der Vorgehensweise bei der automatischen Verschlagwortung der Aufsätze und Literaturzitate.

### Digitale Bibliothek

Die Vorzüge einer digitalen Bibliothek, d.h. des Vorliegens der Aufsätze in Form von Dateien, brauchen hier nicht mehr ausgeführt werden, ebenso wenig wie die Problematik betreffend dem Urheberrecht und der Verfügbarmachung über das Internet. Für Earwigs-online wurde folgender Weg beschritten. Wenn immer möglich werden die Literaturstellen durch die frei verfügbaren Zusammenfassungen (z.B. von ScienceDirect, Google Scholar, CHLA, BCA und andere) im entsprechenden Datensatz ergänzt. Damit sind die Zusammenfassungen auch über Earwigs-online durchsuch- und abrufbar. In einem nicht-öffentlichen Bereich von Earwigs-online liegen die Dateien im JPEG-, PDF- oder HTML-Format, mit den entsprechenden Datensätzen verknüpft, auf der lokalen Festplatte vor und können über einen Tastendruck in einem entsprechenden Programm geöffnet werden.

### Zugriffe und Nutzung

Die online verfügbare Version von Earwigs-online erfreut sich trotz ihres spezialisierten Inhaltes zu ausschließlich einem Taxon und des Fehlens der Anbindung an große Portalsysteme, wie GBIF und anderen, erfreulich großer Nutzung (Fig. 3).

Die Website Earwigs-online wurde den Internetsuchmaschinen, wie Google, Yahoo und Excite vom Autor gemeldet, und auch gezielt Kollegen angesprochen, Earwigs-online mit den von ihnen betriebenen relevanten Seiten, wie OSF, DORSA etc, zu verknüpfen. Durch beide Initiativen konnte die Bekanntheit soweit gesteigert werden, dass nun Anfragen von nicht-Systematikern eingehen.

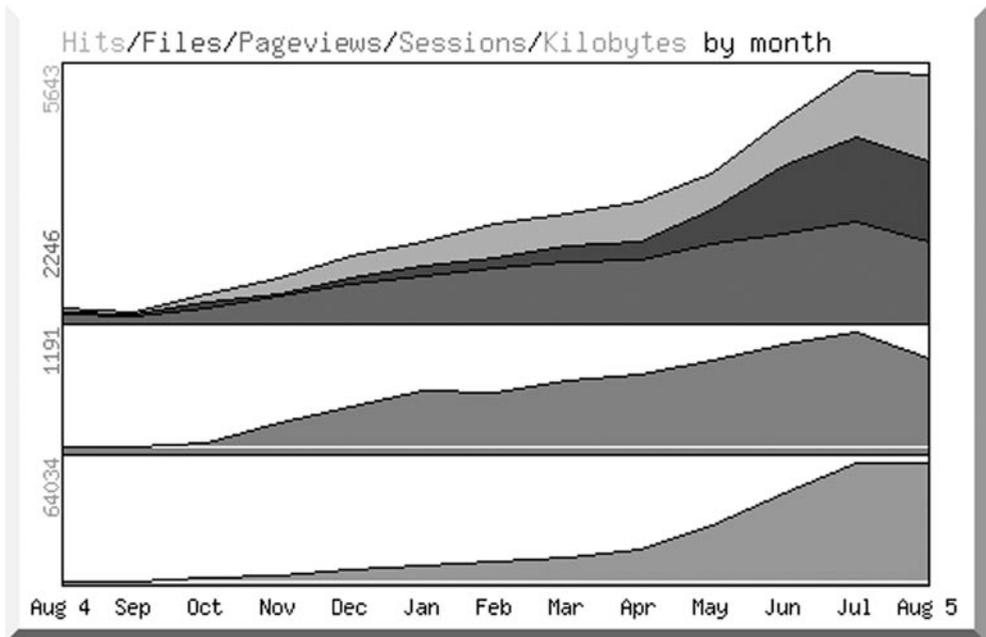


Fig. 3: Nutzung der Earwigs-online Datenbank nach den Serverprotokollen des Providers in den letzten 12 Monaten (Juli 2004 bis Juli 2005). Mit rund 40 Sessions pro Tag (ca. 1200 pro Monat) ist die Nutzung erfreulich hoch.

### Vor- und Nachteile zusammengefasst

Nach 10-jähriger Arbeit haben sich folgende Vor- und Nachteile konkretisiert, die einfach tabellarisch zusammengefasst werden können (Tab. 2). Der gewichtigste Vorteil soll aber hier angeführt werden: Die Dermaptera sind weder in medizinischer, landwirtschaftlicher, ökonomischer noch molekularbiologischer oder systematischer Hinsicht im Brennpunkt des Interesses. Daher ist es zum gegenwärtigen Zeitpunkt extrem unwahrscheinlich, dass dieses Taxon von irgendeiner Institution derart vollständig bearbeitet und verfügbar gemacht worden wäre.

Tab. 2: Vor- und Nachteile der Eigenentwicklung einer Datenbank.

<b>Vorteile</b>
1) alle Daten sind beliebig abrufbar und kombinierbar, z.B. für Regionalfaunen und andere Publikationen
2) Anpassung an eigene (kurzfristige) Bedürfnisse
3) Schnelle und kontrollierte Veröffentlichung der Daten
4) eigene Designmöglichkeiten für die Benutzeroberfläche, auch im Internet
5) eigene IT-Kenntnisse werden aufgebaut
6) die verwendete(n) Systematik(en) sind frei wählbar
7) keine vorgegebene Daten-/Feldstruktur
8) Kooperationsgrundlage für Regionalfaunen und Anträge etc
<b>9) REALISIERUNG DER DATENBANK</b>
<b>Nachteile</b>
1) Eigenverantwortung (und Eigenarbeit) für Datensicherung, Schnittstellen, Kompatibilität, Internet etc
2) „Bus-Test“: ist die Datenbank nach eigenem Ausscheiden durch andere noch nutzbar?
3) Standards, Synonyme und Thesauri müssen selbst eingearbeitet und organisiert werden („Wiedererfindung des Rades“)
4) finanzielle Förderung ist schwerer möglich als in etablierten Projekten: folglich weniger studentische und mehr eigene Arbeit
5) Bekanntheitsgrad ist geringer, daher geringere Nutzung
6) je nach Artenzahl ist ein derartiges Projekt nicht für Einzelpersonen durchführbar; die Obergrenze liegt wohl bei rund 2000 Arten

### Danksagung

Das Projekt der Earwigs-online Datenbank wäre ohne die großzügige Unterstützung verschiedener Personen nicht möglich gewesen. Dieter Walošek danke ich für die ganz wesentliche Unterstützung während meiner Ulmer Zeit, ebenso wie für die Überlassung von 2 GB Serverspace und den Internetdienst. Caro Raap und Andrea Bäuerle danke ich für die unermüdliche Dateneingabe. Daneben ist es mir eine Freude den an den Programmen COBICE, COLPARSYST und SYS-RESOURCE beteiligten Kollegen zu danken, die mir den Zugang zu bedeutenden Ohrwurmsammlungen ermöglichten. Christoph L. Häuser möchte ich für die umfangreiche Unterstützung am SMNS danken.

## Literatur und Websites

MIKKELSEN, P. M. & BIELER, R. 2000: Marine bivalves of the Florida Keys: discovered biodiversity. - In: HARPER, E. M.; TAYLOR, J. D. & CRAME, J. A. (eds): The Evolutionary Biology of the Bivalvia. - Geological Society Special Bulletin, 177: 367-387.

Apple Macintosh	Computer- und Softwarehersteller	<a href="http://www.apple.com/de">www.apple.com/de</a>
BCA	Biologia Centrali-Americana	<a href="http://www.sil.si.edu/digitalcollections/bca/">www.sil.si.edu/digitalcollections/bca/</a>
CHLA	Core Historical Literature of Agriculture	<a href="http://chla.library.cornell.edu/">http://chla.library.cornell.edu/</a>
DORSA	Deutsche Orthopteren Sammlungen Online	<a href="http://www.dorsa.de">www.dorsa.de</a>
Earwigs-online	Website über Dermaptera	<a href="http://www.earwigs-online.de">www.earwigs-online.de</a>
Excite	Internetsuchmaschine	<a href="http://www.excite.de">www.excite.de</a>
FileMaker	Softwarefirma des gleichnamigen Programmes	<a href="http://www.filemaker.de">www.filemaker.de</a>
Gallica	Französische digitale Bibliothek	<a href="http://gallica.bnf.fr">http://gallica.bnf.fr</a>
GBIF	Global Biodiversity Information Facility	<a href="http://www.gbif.org">www.gbif.org</a>
Google	Internetsuchmaschine	<a href="http://www.google.de">www.google.de</a>
Google Scholar	Internetsuchmaschine für Inhalte wissenschaftlicher Zeitschriften	<a href="http://scholar.google.com/">http://scholar.google.com/</a>
Göttinger Digitalisierungszentrum		<a href="http://gdz.sub.uni-goettingen.de/de/index.htm">http://gdz.sub.uni-goettingen.de/de/index.htm</a>
Microsoft Windows	Softwarehersteller	<a href="http://www.microsoft.de">www.microsoft.de</a>
OSF	Orthoptera Species File	<a href="http://osf2x.orthoptera.org">http://osf2x.orthoptera.org</a>
ScienceDirect	Bibliographische kommerzielle Website	<a href="http://www.sciencedirect.com">www.sciencedirect.com</a>
Strato	Internetprovider	<a href="http://www.strato.de">www.strato.de</a>
Yahoo	Internetsuchmaschine	<a href="http://www.yahoo.de">www.yahoo.de</a>

## Anschrift des Verfassers:

Dr. FABIAN HAAS  
 GTI-Kontaktstelle  
 Staatliches Museum für Naturkunde  
 Rosenstein 1  
 D - 70191 Stuttgart  
 e-mail: [haas.smns@naturkundemuseum-bw.de](mailto:haas.smns@naturkundemuseum-bw.de)