



GoldenGATE Imagine XML Markup Editor

This is work in progress

**Introduction and Manual for the Generation
of XML-based Legacy Literature
Documents using the GoldenGATE
Imagine Editor**

DRAFT 20141124

**Jeremy Miller, Donat Agosti, Guido Sautter,
Terry Catapano & Christiana Klingenberg,**

([klingenberg\[at\]plazi.org](mailto:klingenberg[at]plazi.org); [sautter\[at\]ira.uka.de](mailto:sautter[at]ira.uka.de); [agosti\[at\]plazi.org](mailto:agosti[at]plazi.org); [catapano\[at\]plazi.org](mailto:catapano[at]plazi.org))

The GoldenGATE release, demo files and detailed instructions are available at:

<http://plazi.org>



This version is the preparation for the GoldenGATE Imagine software.

This version should be used to teach colleagues to contribute colleagues to enter data towards the 1 Million Treatment Goal.

It should have a basic workflow, and advanced section, each element we use in the markup process and the 1MT should be defined, the Code given and as well as illustrated with a Code example.

Basic editing steps will be explained.

As an introduction, please consult the [GG-Pro-iBiosphere](#) manual.

Every Monday morning a new version is being uploaded and available at
<http://plazi.cs.umb.edu/GgServer/Downloads/GgImagine-Default.imagine.zip>
Currently there is no autoupdate in place so the entire new version has to be downloaded. To keep your cache, please copy the cache directory





Version Control

20141124: Establishing of manual

20141203: Clean up, removing all GG manual parts that are not part of Imagine

To be done:

1. add link to test pdf set:

https://drive.google.com/a/plazi.org/folderview?id=0B_yrOwn4yBySbXRRZWtMaGxqUGM&usp=sharing

Questions:

How to edit tokens?

How to change sequence of tokens?

Can we cut the images and export them? Export them into Zenodo?

notes

open GG

zoom to original resolution

tools ->detect document structure

images: different in digital born and scanned pdf



Content

[Setup GoldenGATE-Imagine](#)

[Links](#)

[Installation requirement](#)

[Getting GoldenGate to work](#)

[Select the Default.imagine configuration](#)

[Start menu](#)

[2. The markup process](#)

[Opening a PDF](#)

-

[Detect Document Structure](#)

[Mark Taxon Names](#)

[Parse Bibliography](#)

[How to split merged bibliographic references](#)

[Document Meta Data](#)

[Mark Treatments](#)

[Treatment Structure](#)

[Mark Materials Citations](#)

[Parse Materials Citations](#)

[Edit Attributes](#)

[Editing and replacing wrong fonts and symbols](#)

[Opening a scanned document](#)

[Document structure](#)

[Annotating blocks of texts \(names, treatments, etc\)](#)

[Annotate taxonomic names](#)

[Export treatment of document to SRS](#)

[Edit document metadata](#)

[Detect treatment structure](#)

[Link to external resources](#)

[Scientific names](#)

[Export](#)

[Minimal requirement for export to GG SRS](#)

[Trouble Shooting](#)

[File Format not valid](#)

[Loading of the file stalls](#)

[Opening a scanned PDF with underlying OCR as born digital](#)

[Editing and menu choices](#)

[Glossary](#)

[GoldenGate Expressions](#)

[TaxonX Expressions](#)

[Element Descriptions and Examples](#)

[GoldenGate Internal Markup Elements](#)



[Acknowledgments](#)

[Notes](#)

[Editing steps to be added](#)

[GoldenGate List of Functions](#)





Setup GoldenGATE-Imagine

Links

- [Introduction and manual to GoldenGATE](#) This document includes an extensive coverage of the editing steps, analysers, vocabulary, and trouble shooting.
- Introduction to GG-Imagine
 - [Github](#)
 - [PPT](#) introduction to Imagine File Format and markup
 - [Test files](#)
- Download [test version](#). Currently, each Monday a new version of GG-Imagine is posted. Previous version will not be autoupdated.

Installation requirement

REQUIRED JAVA VERSION

The program requires Java 1.6 and above, and uses the Oracle Java run time environment.

To determine your computers current Java version, do the following.

Windows: go to “Command prompt” and type in “java -version”.

```
C:\Users\Donat>java -version
java version "1.8.0_31"
Java(TM) SE Runtime Environment (build 1.8.0_31-b13)
Java HotSpot(TM) Client VM (build 25.31-b07, mixed mode)
```

Mac: From terminal type “java /version”

The program requires Java 1.6 and above, and uses the Oracle Java run time environment.



Linux:

GoldenGate is built on Java, a free, cross-platform software development environment. Java comes in a number of versions, so it is important that you install the version that GoldenGate requires: Java 7 or 8. GoldenGate can run with both the Java "JRE" (Java Runtime Environment). For instructions about running Java on your operating system, see Oracle's [Installing Java](#) page.

Linux users: Unfortunately, Oracle's own pages (such as the link above) focus on commercial Linux distributions. Users of other distributions are better served by distribution-specific instructions.

Special note: It is recommended to use the Oracle Java VM, this is the safest choice. The OpenJDK 7 is a good open source alternative for the Oracle JVM.

Details of the Java VM can be obtained via the following command (two results displayed):

```
java -version
java version "1.7.0_55"
Java(TM) SE Runtime Environment (build 1.7.0_55-b13)
Java HotSpot(TM) 64-Bit Server VM (build 24.55-b03, mixed mode)
```

It is NOT recommended to use "OpenJDK6" (IcedTea) and "GNU Compiler for Java (GCJ) which are shipped by several Linux distributions.

REQUIRED GOLDENGATE VERSION

After you installed GoldenGATE, the program will be updated automatically (online access necessary).

LICENSE

Copyright (c) 2006-2014, Guido Sautter, IPD Boehm, Universität Karlsruhe (TH)
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- * Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- * Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- * Neither the name of the Universität Karlsruhe (TH) nor the names of its contributors may be used to endorse or promote products



derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS 'AS IS' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

CHANGE LOG

Changes to this configuration in reverse chronological order

NEW IN VERSION 2015.05.17.14.53

- small-caps problems in PDF font decoding alleviated
- implicit space text obfuscation problems fixed
- landscape page problems fixed
- added editable XML view of document and annotations
- added option to decode PDFs in slave JVM to alleviate memory problems (deactivated in config by default, though)

NEW IN VERSION 2015.05.09.01.21

- PDF facilities can decode TrueType fonts now
- further improvements to PDF font decoding facilities
- extended font correction functionality for born-digital PDFs
- sorting issues in Java 1.7 and above sorted out
- citations, footnotes, and tables hidden now on treatment and materials citation markup
- many minor fixes and improvements

NEW IN VERSION 2015.03.25.15.07



- improved page image extraction for scanned PDFs (working much the same as for figures now)
- revised distinction of scanned from born-digital PDFs
- fixed decoding of sans-serif fonts
- added MacOS version of OCR engine
- fixed splitting of blocks into paragraphs
- failing gracefully now if Java refuses to render some character
- added manual assignment of captions to images and tables

NEW IN VERSION 2015.03.19.09.31

- added automated title and pagination recognition from document to metadata editor
- added document head annotation of document metadata attributes
- document metadata editor also shows in Tools menu
- extraction buttons for already extracted document metadata attributes reveal themselves with green borders
- restricted emphasis annotations to within paragraph boundaries
- added manual correction of assignment of caption targets (tables and figures) to captions
- updates to caption targets automatically write through to captions, as well as to figureCitations and tableCitations
- improved PDF glyph decoding, including correction of double mappings
- fixed occasional loading and tool application hangups
- many minor UI improvements, including fast scrolling
- added negative caching to CoL and GBIF lookups
- added handling of diacritics to bibliographic reference parser

NEW IN VERSION 2015.03.04.14.27

- added document ID extraction to document metadata extraction
- activated RefBank lookup for document meta data
- fixed table bounds generation on table block merging

NEW IN VERSION 2015.02.26.13.40

- filtering vernacular/common names from CoL lookup results

NEW IN VERSION 2015.02.26.12.31

- center alignment detection works on columns now
- added heading hierarchy assessment to document structure detection



NEW IN VERSION 2015.02.13.11.46

- parallelized catalog lookups in taxonomic name detection
- improved authority handling in taxonomic name detection
- added learning from nomenclature acts to taxonomic name detection
- added block merging functionality

NEW IN VERSION 2015.02.09.10.27

- parallelized document structure detection

NEW IN VERSION 2015.02.04.13.10

- taxonomic name detection bugfix

NEW IN VERSION 2015.02.03.04.30

- bugfix in PDF font decoding
- extended document structure detection to connecting tables
- extended document structure detection to marking emphases and headings
- added marking and handling of images, as well as assignment of images to captions
- completely revised taxonomic name detection

NEW IN VERSION 2015.01.19.07.00

- accelerated glyph transcription in embedded fonts of born-digital PDFs
- added glyph transcription editing for embedded fonts in born-digital PDFs
- extended document structure detection to recognizing and marking tables

NEW IN VERSION 2014.12.09.22.00

- finally got intended version of exporter to GoldenGATE Server

NEW IN VERSION 2014.12.09.18.36

- removed generic upload to GoldenGATE Server (the exporter takes its place now)



- improved object listing
- fixed document normalization issue in treatment markup and structuring

NEW IN VERSION 2014.12.09.12.08

- update to data exporter to GoldenGATE Server

NEW IN VERSION 2014.11.28.10.19

- added list view for taxonomic names

NEW IN VERSION 2014.11.28.09.41

- added document structure detector
- added list views for Image Markup Objects

NEW IN VERSION 2014.11.14.00.52

- all-round update
- added editor for document meta data
- added data exporter to GoldenGATE Server

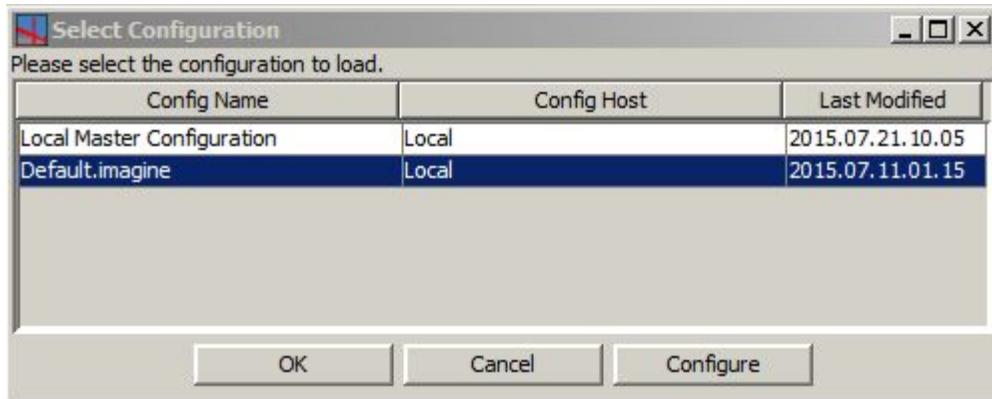
Getting GoldenGate to work

Download [Imagine](#),
extract the files into a directory that is not the root directory
Start Imagine using the GgImagine.exe file

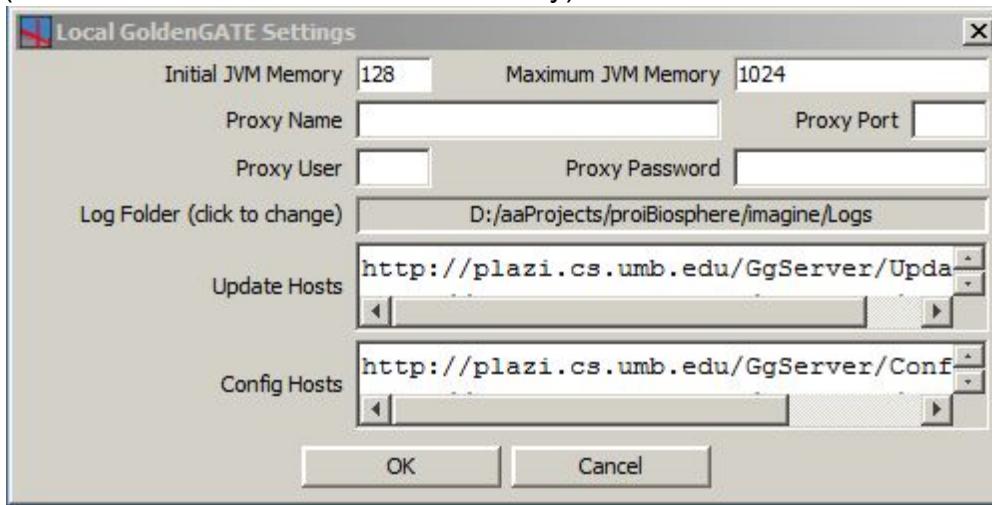


Name	Date modified	Type	Size
Configurations	7/21/2015 9:56 AM	File folder	
Data	7/21/2015 9:56 AM	File folder	
ImageMagick	7/21/2015 10:01 AM	File folder	
Logs	7/21/2015 10:01 AM	File folder	
OcrEngine	7/21/2015 10:01 AM	File folder	
PageImages	7/21/2015 10:01 AM	File folder	
Update	7/21/2015 10:01 AM	File folder	
CleanupLogFiles.bat	11/24/2007 4:00 PM	Windows Batch File	1 KB
CleanupOldFiles.bat	11/19/2007 7:14 PM	Windows Batch File	1 KB
ConfigHosts.cnfg	9/16/2013 7:33 PM	CNFG File	1 KB
EasyIO.jar	4/13/2015 6:10 AM	Executable Jar File	336 KB
Gamta.jar	4/13/2015 6:10 AM	Executable Jar File	1,359 KB
GamtaImagingAPI.jar	3/28/2015 1:55 PM	Executable Jar File	205 KB
GgImagine.bat	10/20/2014 9:06 PM	Windows Batch File	1 KB
GgImagine.cnfg	7/21/2015 10:03 AM	CNFG File	3 KB
GgImagine.cnfg.2015-07-21-10-03-13.old	7/21/2015 10:03 AM	OLD File	3 KB
GgImagine.exe	11/13/2014 10:08 PM	Application	78 KB
GgImagine.jar	7/11/2015 12:00 AM	Executable Jar File	263 KB
GgImagine.sh	11/13/2014 10:09 PM	SH File	1 KB
GgImagine-Default.imagine.zip	7/21/2015 9:56 AM	Compressed (zippe...)	34,337 KB
GgImagineStarter.jar	7/11/2015 12:00 AM	Executable Jar File	35 KB
GoldenGATE.jar	3/28/2015 1:56 PM	Executable Jar File	1,302 KB
HtmlXmlUtil.jar	4/13/2015 6:10 AM	Executable Jar File	283 KB
icepdf-core.jar	4/9/2013 4:58 PM	Executable Jar File	766 KB
ImageMarkup.jar	7/10/2015 11:58 PM	Executable Jar File	710 KB
ImageMarkupOCR.bin.jar	3/25/2015 3:43 PM	Executable Jar File	11,037 KB
ImageMarkupOCR.jar	7/10/2015 11:58 PM	Executable Jar File	52 KB
ImageMarkupPDF.bin.jar	3/25/2015 3:43 PM	Executable Jar File	2,305 KB
ImageMarkupPDF.jar	7/10/2015 11:58 PM	Executable Jar File	4,693 KB
License.txt	7/13/2008 6:32 PM	TXT File	2 KB

Select the Default.imagine configuration



Memory and other technical settings can be edited through the Configure button (recommended for advanced users only)



Start menu

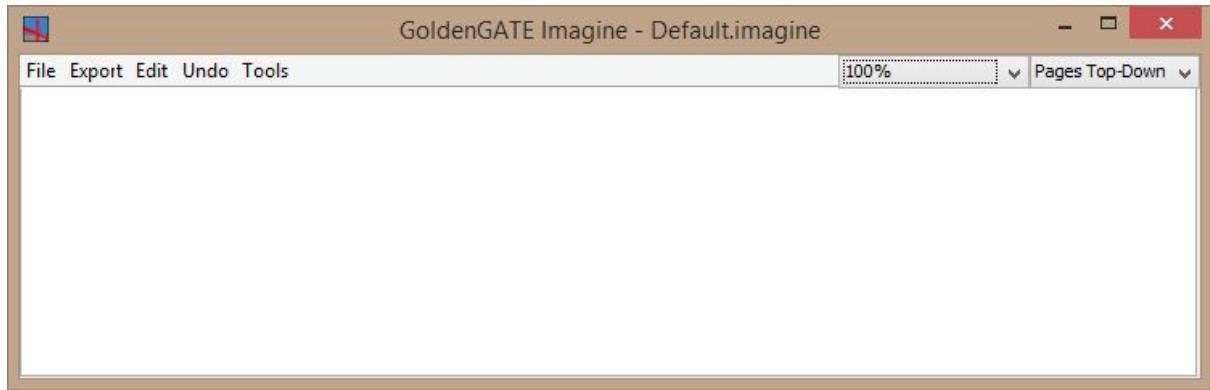
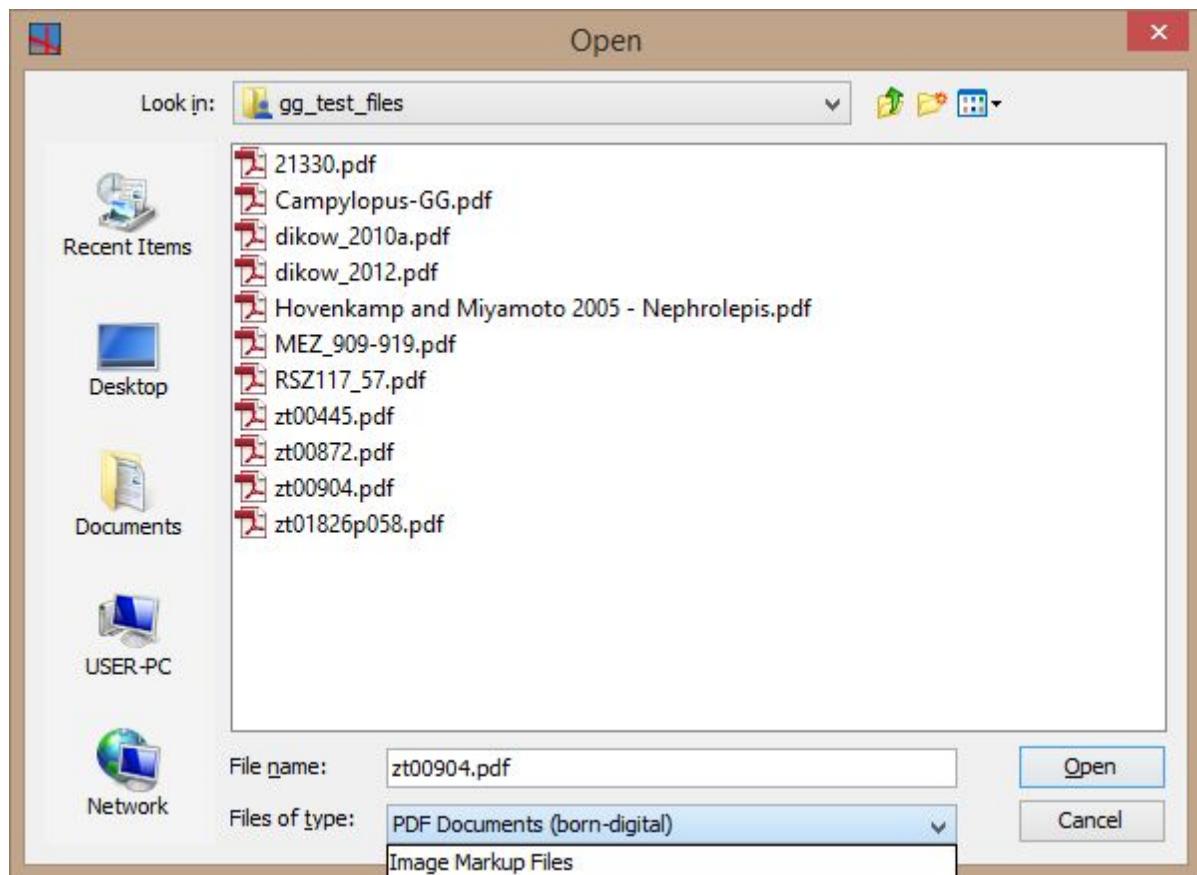


Fig XX: Start menu: 1 Functions; 2 Display zoom; 3 tiling of subwindows

2. The markup process

Opening a PDF

Open a file from the “File” menu, or drag and drop the selected file into the start menu



- Select Image Markup Files to continue with an existing project
- Select PDF Document (born-digital) for a PDF file that has been converted from a text document. In case for choosing this version for a scanned PDF, the editor will close. In case of opening a PDF with OCR underlying, see troubleshooting.
- Select PDF Documents (scanned) for a PDF file that has been made from scanned images of the text
- Select PDF Document in case the origin is not known. GGImagine will then decide how to proceed.

For comparison on an average PC (8GB RAM, 64-bit, Intel Core i7, 2.80Ghz. MS Windows 8.1 Pro) opening the test file [zt00904.pdf](#) will take 15 seconds.

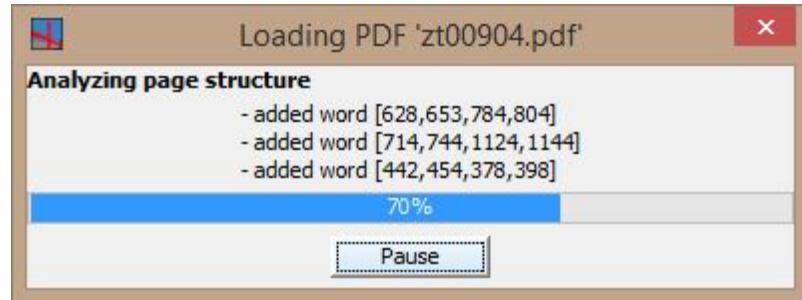


Fig XX Status of loading progress of a pdf

Run

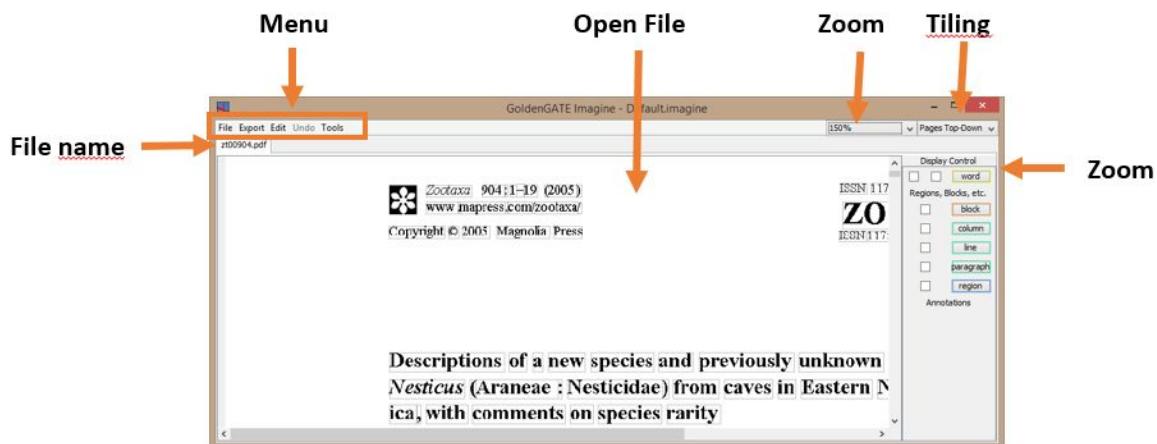


Fig. XX Main window. Display control on the right hand; Zoom control; Page display options; Menu; Tabs showing open files. Overview of menu with an opened pdf file.

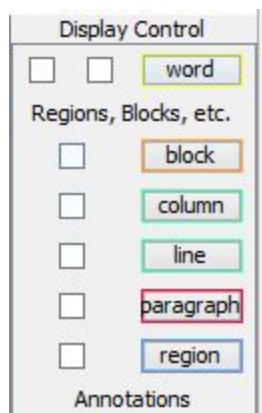




Fig XX Display control. word =, block = , column =; line =, paragraph = , region

Adjust the zoom to the extend that the letters can be properly read.

Generally, all the tools can be used in either sequence.

Detect Document Structure

From the “Tools” menu, run “Detect document structure”

This is necessary in order to assure that the page numbers are assigned.

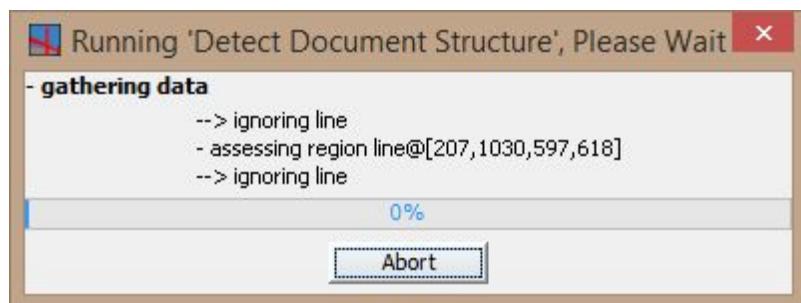


Fig XX Status window for the “Detect Document Structure” analyzer



Introduction

The spider genus *Nesticus* Thorell, 1869 (family Nesticidae) is taxonomically diverse in the southern Appalachian mountains of eastern North America, with at least 30 species distributed over a geographic area extending from southern West Virginia to central Alabama (Gertsch 1984; Coyle & McGarity 1992; Hedin 1997a). Appalachian *Nesticus* are habitat specialists, reflecting apparently strict physiological constraints that limit these spiders to cool, moist microhabitats. These constraints, in combination with habitat discontinuity in both space and over time, have promoted tremendous species diversification and endemism (see speciation models of Wiens 2004a, 2004b). This fine-scale diversification

Accepted by P. Jäger, 23 Feb. 2005; published 18 Mar. 2005

Hide Page
Annotate ▾
Annotate All ▾
Mark Region
Mark Table
Mark Caption
Mark Footnote
Mark Page Header
Mark Parenthesis
Mark Artifact
Mark Image
Merge Words
Make Stream

Fig. XX Modify document structure. The automated document structure detector misidentified a footnote and page number as being part of the document body text. This is indicated by the gray line (revealed by checking the “Word” box on the right panel) that traces document word order. Draw box around the footnote and page number, and select “Mark Page Header.” The word order will change to exclude the footnote and page number.

At this stage, there are several possible tools that can be used (order is not important). These are: Mark taxon names, Parse bibliography, and Document metadata

Mark Taxon Names

From the “Tools” menu, select “Mark Taxon Names”

Imagine will detect taxonomic names throughout the document and find higher rank taxa. To view/edit taxonomic name attributes, click on a name and select “Edit taxonomicName Attributes”. This tool finds associated higher taxonomic names by querying external online databases, so be aware of this in case you will be using Imagine off line.

In most cases, it is not necessary to use the “Parse Taxonomic Names” tool.

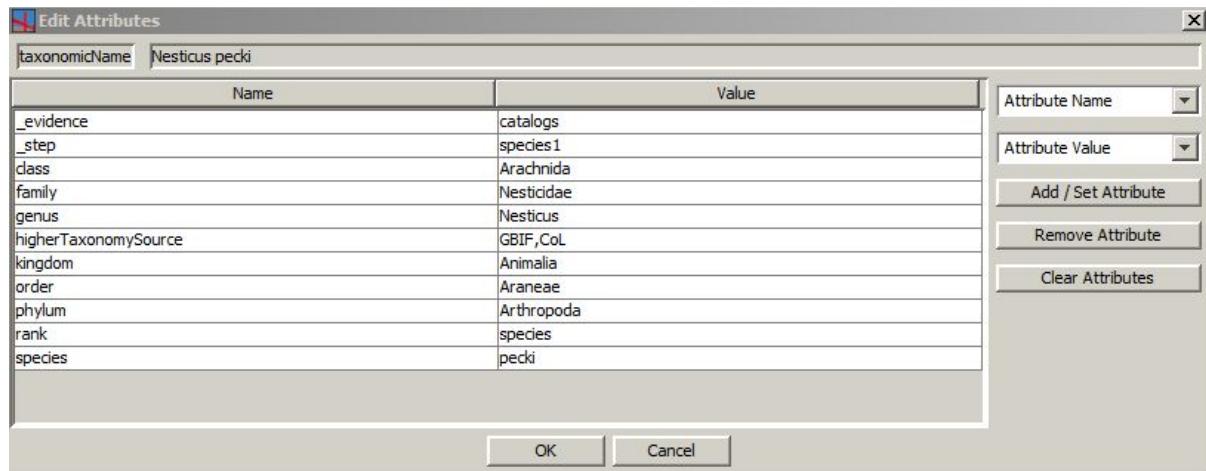


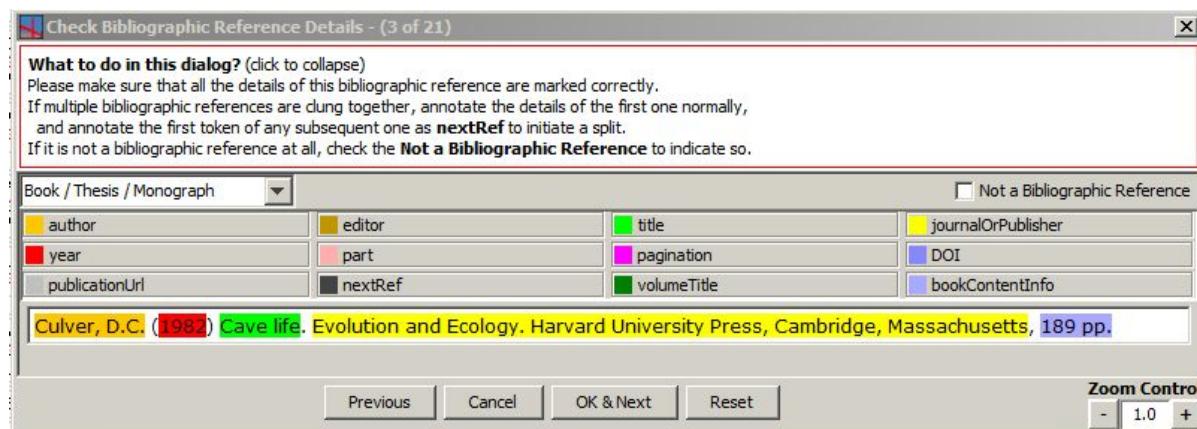
Fig XX Taxonomic name attributes

Parse Bibliography

Before running this tool, it is advisable to check the paragraph structure of the bibliography. Be sure that “Paragraph” is checked in the Display Control. Each reference should be within a single paragraph (by default, indicated by a red rectangle), unless it spans a page break.

If there are any errors, these can be corrected. Draw a box using the click and drag function. Paragraphs can be removed, split, and created (Mark region/paragraph).

When the paragraphs are correct, select “Parse Bibliography” from the “Tools” menu. Check the suggestions for assigning elements of the reference to the appropriate field.





Check Bibliographic Reference Details - (3 of 21)

What to do in this dialog? (click to collapse)
Please make sure that all the details of this bibliographic reference are marked correctly.
If multiple bibliographic references are clung together, annotate the details of the first one normally,
and annotate the first token of any subsequent one as **nextRef** to initiate a split.
If it is not a bibliographic reference at all, check the **Not a Bibliographic Reference** to indicate so.

Not a Bibliographic Reference

Book / Thesis / Monograph

author	editor	title	journalOrPublisher
year	part	pagination	DOI
publicationUrl	nextRef	volumeTitle	bookContentInfo

Culver, D.C. (1982) Cave life. Evolution and Ecology. Harvard University Press, Cambridge, Massachusetts, 189 pp.

Remove 'journalOrPublisher' annotation

Previous Cancel OK & Next Reset Zoom Control 1.0

Check Bibliographic Reference Details - (3 of 21)

What to do in this dialog? (click to collapse)
Please make sure that all the details of this bibliographic reference are marked correctly.
If multiple bibliographic references are clung together, annotate the details of the first one normally,
and annotate the first token of any subsequent one as **nextRef** to initiate a split.
If it is not a bibliographic reference at all, check the **Not a Bibliographic Reference** to indicate so.

Not a Bibliographic Reference

Book / Thesis / Monograph

author	editor	title	journalOrPublisher
year	part	pagination	DOI
publicationUrl	nextRef	volumeTitle	bookContentInfo

Culver, D.C. (1982) Cave life. Evolution and Ecology. Harvard University Press, Cambridge, Massachusetts, 189 pp.

Extend 'title' annotation
Remove 'title' annotation

Previous Cancel OK & Next Reset Zoom Control 1.0

Check Bibliographic Reference Details - (3 of 21)

What to do in this dialog? (click to collapse)
Please make sure that all the details of this bibliographic reference are marked correctly.
If multiple bibliographic references are clung together, annotate the details of the first one normally,
and annotate the first token of any subsequent one as **nextRef** to initiate a split.
If it is not a bibliographic reference at all, check the **Not a Bibliographic Reference** to indicate so.

Not a Bibliographic Reference

Book / Thesis / Monograph

author	editor	title	journalOrPublisher
year	part	pagination	DOI
publicationUrl	nextRef	volumeTitle	bookContentInfo

Culver, D.C. (1982) Cave life. Evolution and Ecology. Harvard University Press, Cambridge, Massachusetts, 189 pp.

Previous Cancel OK & Next Reset Zoom Control 1.0

Fig XX Correcting bibliographic reference attributes. Selected text can be assigned to a field by clicking on the appropriate label above. Note that text can belong to only one field, so any interfering highlights must be removed first. To do this, highlight the problematic text and select “Remove...” from the context menu.



Literature Cited

Barr, T.C. Jr. (1961) <i>Caves of Tennessee</i> . Bulletin 84, Tennessee Department of Conservation and Commerce, Division of Geology, Nashville, Tennessee, 567 pp.	Annotate
Coyle, F.A. & McGarity, A.C. (1992) Two new species of <i>Nesticus</i> spiders from the southern Appalachians (Araneae, Nesticidae). <i>Journal of Arachnology</i> , 19, 161.	Annotate All
Culver, D.C. (1982) <i>Cave life, Evolution and Ecology</i> . Harvard University Press, Cambridge, Massachusetts, 189 pp.	Edit Word Attributes
Culver, D.C., Master, L.L., Christman, M.C. & Hobbs, H.H. III (2000) The cave fauna of the 48 contiguous United States. <i>Conservation Biology</i> , 14, 386-401.	Edit Annotation Attributes ...
Emerton, J.H. (1875) Notes of spiders from caves in Kentucky, Virginia, and West Virginia. <i>American Naturalist</i> , 9, 278-281.	- bibRef' Culver ... '
Gertsch, W.J. (1984) The spider family Nesticidae (Araneae) in North America and the West Indies. <i>Texas Memorial Museum Bulletin</i> , 31, 1-91.	- journalOrPublisher 'Harvard University Press, Cambridge, Massachusetts'
Gibert, J. & Deharveng, L. (2002) Subterranean ecosystems: a truncated science. <i>Bioscience</i> , 52, 473-481.	Remove Annotation ...
	Remove All Annotations ...
	Change Annotation Type ...
	Start Annotation
	Split bibRef Before
	Split journalOrPublisher Before
	Split bibRef After
	Split journalOrPublisher After
	Copy Annotation Text ...
	Copy Annotation XML ...

Culver, D.C. (1982) *Cave life, Evolution and Ecology*. Harvard University Press, Cambridge, Massachusetts, 189 pp.

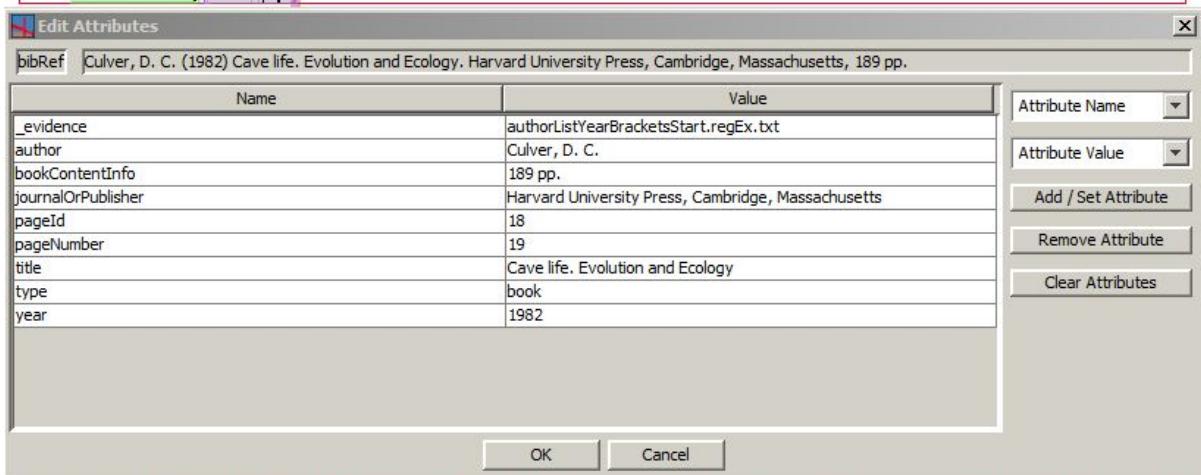
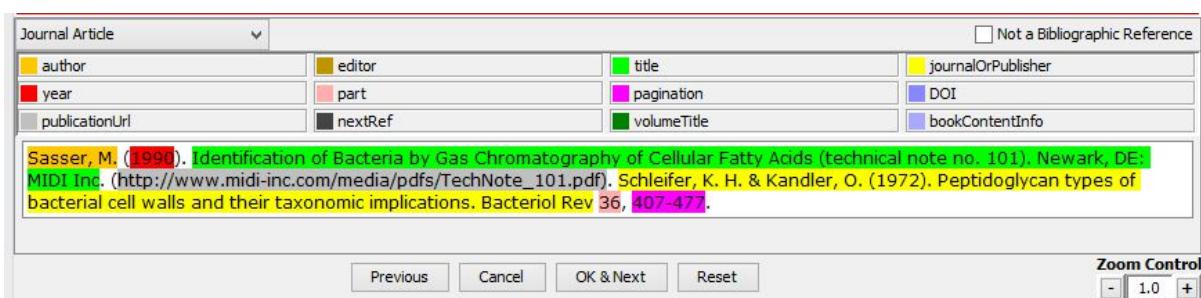


Fig XX Check bibliographic reference attributes, select “bibRef...” Edit attributes as needed.

How to split merged bibliographic references



Finish analyser (and remember the reference)



in main display close all highlights of tags and open only bibRef

GoldenGATE Imagine - Default.imagine

File Export Edit Undo Tools Help
ps.02898-0.pdf

Page 5 / 5 (Nr. 225) 150% Pages Top-Down

Display Control

word

Regions, Blocks, etc.

Show All Hide All

block

column

image

line

paragraph

region

table

tableCell

tableCol

tableRow

Annotations

Show All Hide All

bibRef

caption

emphasis

figureCitation

heading

journalOrPublisher

pageNumber

pageTitle

pagination

nces

D. & Jones, D. (1981). Distribution of isoprenoid quinone types in bacteria and their taxonomic implication. *Rev 45*, 316–354.

D. Phrouz, T., Goodfellow, M. & Minnikin, D. E. (1977). Distribution of menaquinones in actinomycetes and corynebacteria. *Int J Bacteriol* **100**, 221–230.

D. Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez, I., Garcia, P., Cai, J., Hippe, H. & Farrow, J. A. E. (1996). The phylogeny of the genus *Clostridium*: proposal of five new species combinations. *Int J Syst Bacteriol* **46**.

G. (1983). A least squares algorithm for fitting additive models to rank data. *Psychometrika* **48**, 621–626.

T. J. (1993). PHYLIP (phylogeny inference package), version 3.6. Department of Genetics, University of Washington, Seattle.

Schumann, P., Weiss, N., Martin, K. & Rainey, F. A. (1995). *Paracoccus jannaschii* gen. nov., sp. nov., a new genus of bacteria with diaminobutyric acid in the cell wall. *Int J Syst Bacteriol* **45**, 234–239.

(1986). Genus *Erysipelothrix*. In *Bergey's Manual of Bacteriology*, pp. 1245–1249. Edited by P. H. A. Sneath, R. M. F. Mair, M. E. Sharpe & J. G. Holt. Baltimore: Williams & Wilkins.

Schleifer, K. H. & Kandler, O. (1972). Peptidoglycan in bacterial cell walls and their taxonomic implications. *Bact Rev* **36**, 407–477.

Schubert, K. & Fiedler, F. (2001). Structural investigation of the cell surface of *Erysipelothrix rhusiopathiae*. *Syst Appl Microbiol* **26**, 30.

Sherman, W. B. D., McGowan, V. & Sneath, P. H. A. (1981). Approved lists of bacterial names. *Int J Syst Bacteriol* **31**.

Silburt, R. M. & Krieg, N. R. (1994). Phenotypic characterization methods for general and molecular bacteriology, pp. 607–624. Edited by P. Gerhardt, R. G. E. Murray, W. A. Wood & N. W. Washington, DC: American Society for Microbiology.

Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic notes for DNA-DNA reassociation and 16S rRNA sequence data in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 845–849.

Takemoto, T., Fujisawa, T., Benno, Y., Tamura, Y., Suzuki, S., Muramatsu, M. & Mitsuoka, T. (1987). *Erysipelothrix tonsillarum* sp. nov., isolated from tonsils of apparently healthy children. *Int J Syst Bacteriol* **37**, 166–168.

Takahashi, T., Fujisawa, T., Tamura, Y., Suzuki, S., Muramatsu, M., Benno, Y. & Mitsuoka, T. (1992). DNA homology among *Erysipelothrix rhusiopathiae* strains representing three serovars and *Erysipelothrix tonsillarum*. *Int J Syst Bacteriol* **42**, 469–473.

Takeshi, K., Makino, S., Ikeda, T. & 7 other authors (1996). Direct and rapid detection by PCR of *Erysipelothrix rhusiopathiae* prepared from bacterial strains and animal tissues. *J Clin Microbiol* **34**, 103–106.

< >

A red circle highlights the reference to *Erysipelothrix tonsillarum* sp. nov. in the list of publications.

Each bibRef is highlighted with the begin and end of the annotation in bolder shade.

Highlite the begin of the omitted reference



GoldenGATE Imagine - Default.imagine

File Export Edit Undo Tools Help

Page 5 / 5 (Nr. 225) 150% Pages Top-Down

js.0.02898-0.pdf

nces

D. & Jones, D. (1981). Distribution of isoprenoid quinone types in bacteria and their taxonomic implication. *Rev 45*, 316–351.

D., Pirouz, T., Goodfellow, M. & Minnikin, D. E. (1977). On menaquinones in actinomycetes and corynebacteria. *Cytobiol 100*, 221–230.

D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-L., J., Garcia, P., Cai, J., Flippé, I. L. & Farrow, J. A. E. (1986). The phylogeny of the genus *Clostridium*: proposal of five new species combinations. *Int J Syst Bacteriol 36*, 845.

G. (1983). A least squares algorithm for fitting additive covariates to data. *Psychometrika 48*, 621–626.

J. (1993). PHYLIP (phylogeny inference package) version 3.6. Department of Genetics, University of Washington, Seattle.

Schumann, P., Weiss, N., Martin, K. & Rainey, P. A. (1995). *Mycobacteriaceae* gen. nov., sp. nov., a new genus of bacteria with diaminobutyric acid in the cell wall. *Int J Syst 6*, 234–239.

(1986). Genus *Erysipelothrix*. In *Bergey's Manual of Bacteriology*, pp. 1245–1249. Edited by P. H. A. Sneath, R. M. F. Sharp & J. G. Holt. Baltimore: Williams & Wilkins.

Schleifer, K. H. & Kandler, O. (1972). Peptidoglycan bacterial cell walls and their taxonomic implications. *B 36*, 407–477.

Schubert, K. & Fiedler, F. (2001). Structural investigation of the cell surface of *Erysipelothrix rhusiopathiae*. *Syst Appl Microbiol 26*, 39.

Skerman, V. B. D., McGowan, V. & Sneath, P. H. A. (1980). Approved lists of bacterial names. *Int J Syst Bacteriol 30*.

Smibert, R. C. (1960). The Smibert-McFaddin test for the differentiation of enteric bacteria. *Int J Syst Bacteriol 10*, 845.

Takemoto, K. (1987). *Int J Syst Bacteriol 37*, 107–110.

Tamura, Y., T., T. (1987). *Int J Syst Bacteriol 37*, 107–110.

Uzzuki, S., Murata, T. (1992). DNA sequence representing the *Erysipelothrix rhusiopathiae* genome. *Int J Syst Bacteriol 42*, 107–110.

Parse Reference

Parse Materials Citations

Edit Font

List bibRef Annotations

Mark Caption

Mark Footnote

Display Control

- word
- Regions, Blocks, etc.
- Show All Hide All
- block
- column
- image
- line
- paragraph
- region
- table
- tableCell
- tableCol
- tableRow

Annotations

- Show All Hide All
- author
- bibRef
- caption
- emphasis
- figureCitation
- heading
- journalOrPublisher
- pageNumber
- pageTitle
- pagination

Select “Split bibRef Before”

Hower over the reference with the mouse, click and select “Parse Reference”

Document Meta Data

From the Edit menu, select Edit Document Meta Data (or alternatively, from the Tools menu, select Add Document Meta Data). There are three ways to enter the document meta data: Extract, Search, and manual entry. Values entered using the first two methods can be modified using manual entry.



Get Meta Data for Document zt00904.pdf

Publication Type: Book / Thesis / Monograph

Authors (use '&' to separate):

Title: Descriptions of a new species and previously unknown males of *Nesticus* (Araneae: Nes

Year: 1994 Pagination: 1-19

Journal:

Part Designators: volume: issue: numero:

Publisher:

Location:

Editors (use '&' to separate):

Volume Title:

Publication Url:

DOI Identifier:

Handle Identifier:

ISBN Identifier:

ISSN Identifier:

ZooBank Identifier:

HNS-Pub Identifier:

Extract Search Validate OK Cancel

Fig XX The Edit Document Meta Data window

The “Extract” method displays the first several lines of document text (expandable by using the “more” button). Highlight text corresponding to the fields outlined in red. Press the field button, which captures the selected text and turns the outline green.

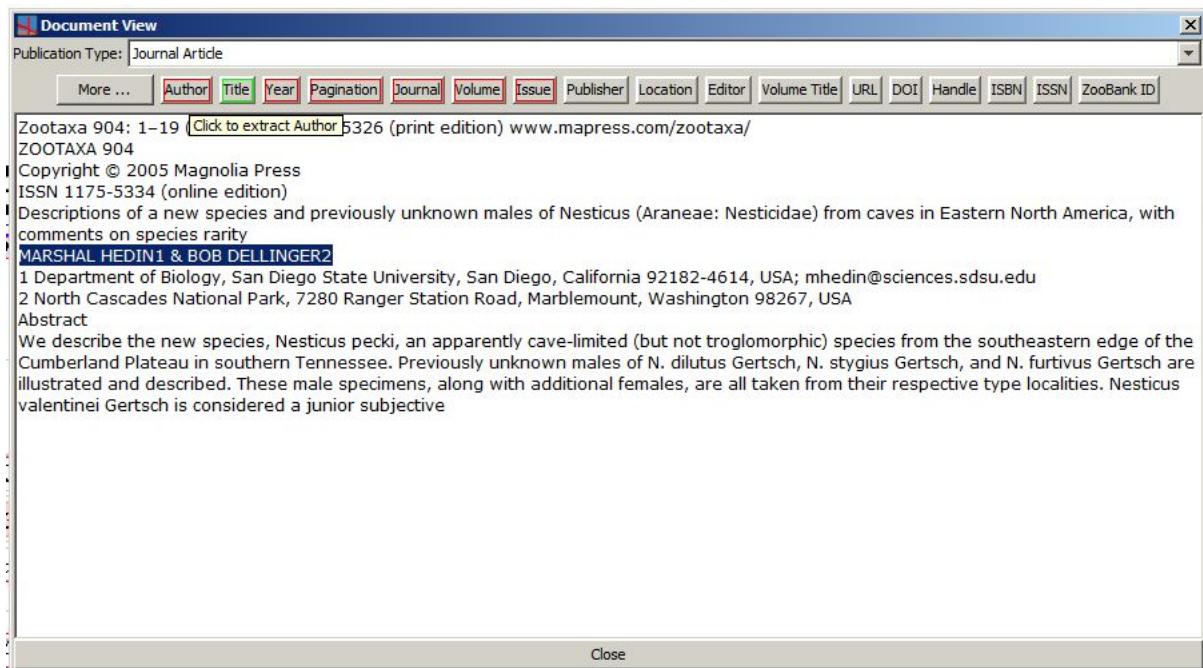


Fig XX Extracting bibliographic meta data. Highlight the appropriate text before clicking on the appropriate field.



Get Meta Data for Document zt00904.pdf

Publication Type:	Journal Article		
Authors (use '&' to separate):	Marshal Hedin1 & Bob Dellinger2		
Title:	eae: Nesticidae) from caves in Eastern North America, with comments on species rarity		
Year:	2005	Pagination:	1-19
Journal:	Zootaxa		
Part Designators:	volume: 904	issue:	numero:
Publisher:			
Location:			
Editors (use '&' to separate):			
Volume Title:			
Publication Url:			
DOI Identifier:			
Handle Identifier:			
ISBN Identifier:			
ISSN Identifier:			
ZooBank Identifier:			
HNS-Pub Identifier:			
<input type="button" value="Extract"/> <input type="button" value="Search"/> <input type="button" value="Validate"/> <input type="button" value="OK"/> <input type="button" value="Cancel"/>			

Fig XX Bibliographic meta data values found using the Extract method. The numbers following the author names can be removed by simply editing the text field.

The “Search” method checks the database for similar documents and offers these meta data.

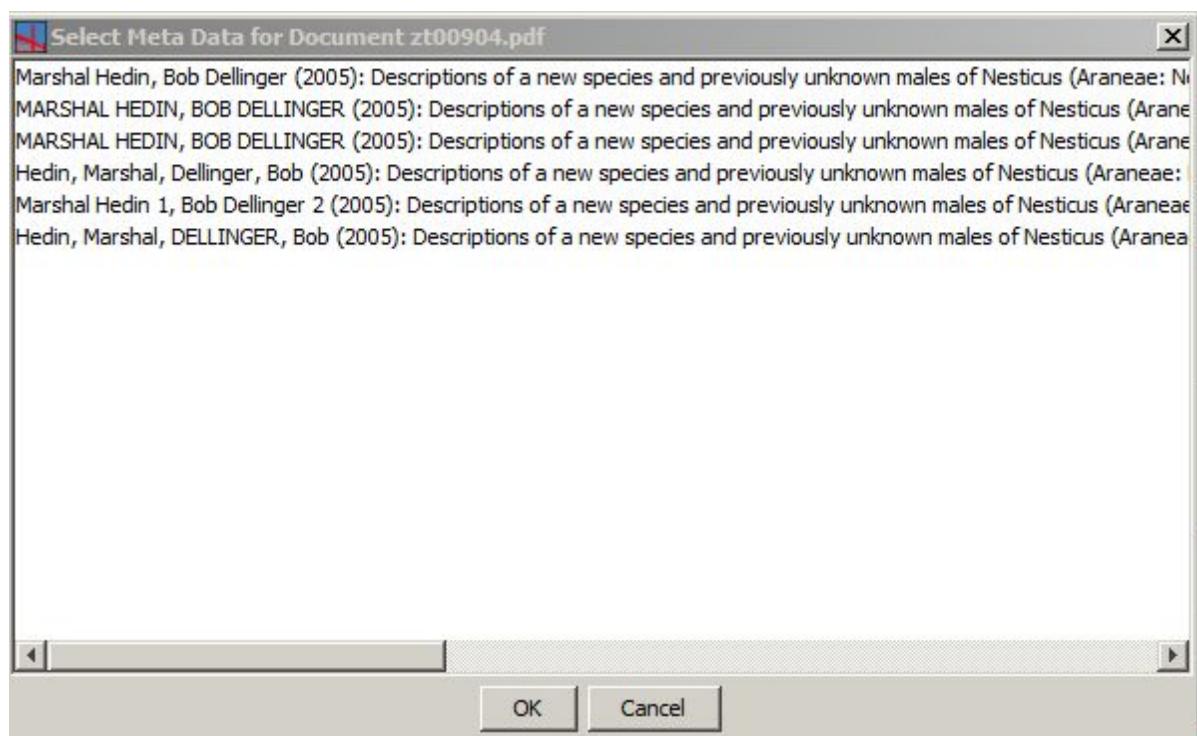


Fig XX Suggested document meta data based on parsed references available to Imagine. Select the best match and click OK



Get Meta Data for Document zt00904.pdf

Publication Type:	Journal Article		
Authors (use '&' to separate):	MARSHAL HEDIN & BOB DELLINGER		
Title:	eeae: Nesticidae) from caves in Eastern North America, with comments on species rarity		
Year:	2005	Pagination:	1-19
Journal:	Zootaxa		
Part Designators:	volume: 904	issue:	numero:
Publisher:			
Location:			
Editors (use '&' to separate):			
Volume Title:			
Publication Url:	http://www.mapress.com/zootaxa/2005f/zt00904.pdf		
DOI Identifier:			
Handle Identifier:			
ISBN Identifier:			
ISSN Identifier:			
ZooBank Identifier:			
HNS-Pub Identifier:			
<input type="button" value="Extract"/> <input type="button" value="Search"/> <input type="button" value="Validate"/> <input type="button" value="OK"/> <input type="button" value="Cancel"/>			

Fig XX Document meta data fields populated using the “Search” method

To refine the data, manually edit. In this example, we have changed the author format (Lastname, Firstname with individuals separated by “&” is preferred) and added a DOI. Use the Validate button to check that all necessary meta data fields have content.



Get Meta Data for Document zt00904.pdf

Publication Type:	Journal Article		
Authors (use '&' to separate):	Hedin, Marshal & Dellinger, Bob		
Title:	eae: Nesticidae) from caves in Eastern North America, with comments on species rarity		
Year:	2005	Pagination:	1-19
Journal:	Zootaxa		
Part Designators:	volume: 904	issue:	numero:
Publisher:			
Location:			
Editors (use '&' to separate):			
Volume Title:			
Publication Url:	http://www.mapress.com/zootaxa/2005f/zt00904.pdf		
DOI Identifier:	http://dx.doi.org/10.11646/zootaxa.904.1.1		
Handle Identifier:			
ISBN Identifier:			
ISSN Identifier:			
ZooBank Identifier:			
HNS-Pub Identifier:			
<input type="button" value="Extract"/> <input type="button" value="Search"/> <input type="button" value="Validate"/> <input type="button" value="OK"/> <input type="button" value="Cancel"/>			

Fig XX Final document metadata.

Mark Treatments

Imagine offers two alternative methods for marking treatment boundaries: the Mark Treatments Tool, and manual method.

From the Tools menu, select “Mark Treatments”. This brings up a dialog box that can be stepped through using the “OK & Next” button. In addition to treatments, several other domains can be marked

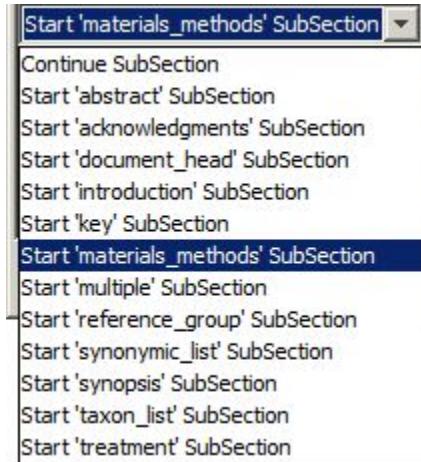


Fig XX The Mark Treatments Tool also allows users to designate several other document SubSections. Use ‘multiple’ for sections that do not fit with any other available category

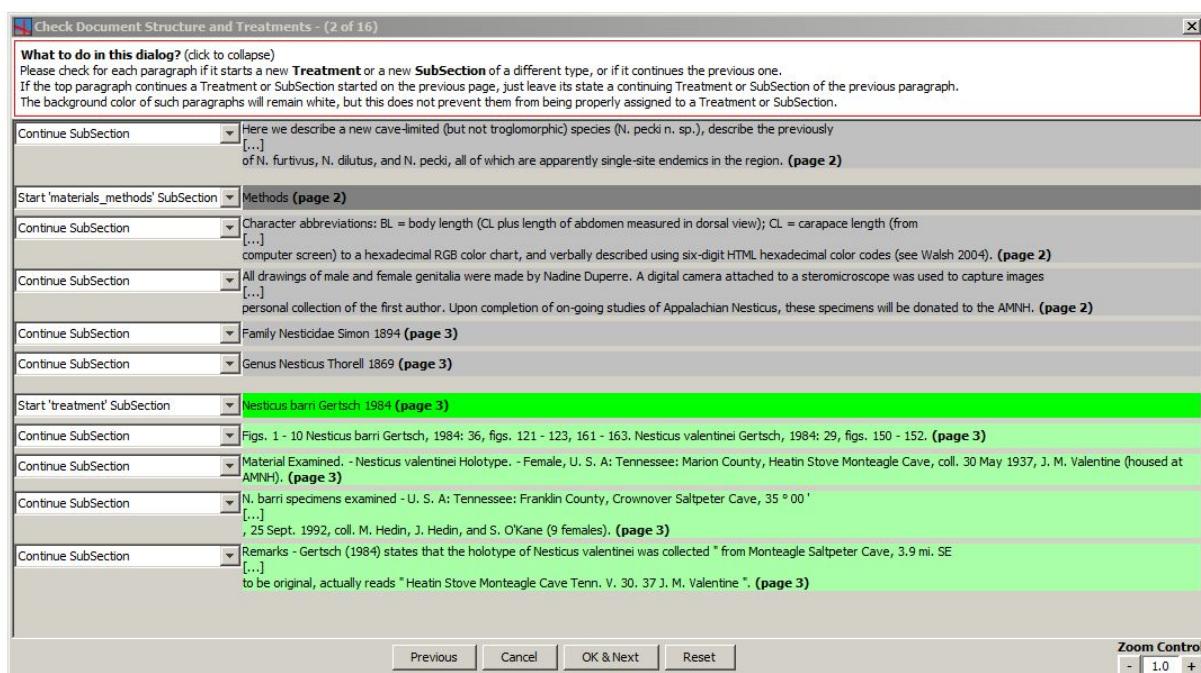


Fig XX Mark treatment boundaries using the Mark Treatments Tool

Alternatively, treatments can be marked manually. Select the first word of the treatment, and select “Start Annotation” from the context menu. A red banner will appear across the top of the screen indicating that an annotation has been started. Scroll down to the end of the treatment. Select the end of the treatment (typically a period, so select carefully; cursor will change from standard arrow to “pencil” icon when selection is possible). A dialog box will prompt the user to enter the type of annotation; select “treatment”.



Family | Nesticidae Simon 1894

Genus | *Nesticus* Thorell 1869

Nesticus Linné Cottsch 1624

Figs

21–123, 161–163
igs, 150–152

Start Annotation

valentinei Holotype — Female, U.S.A; Tennessee;
Monteagle Cave, coll. 30 May 1937, J.M. Valentine

U.S.A; Tennessee; Franklin County, Crownover
V, 21 August 1999, coll. W. Reeves (1 female); Salt
Alabama, 34°59' N, 85°58' W, elev. 210 m, 24 March

females, 2 males), Lost Cove Cave, ~4.7 km NNE of
v, 240 m, 23 Sept. 1992, coll. M. Hedin, J. Hedin, and
th Cave, 35°07' N, 86°00' W, elev. 320 m, 24 March

females, 1 male), Custard Hollow Cave, 24 April 1960,
Spring Cave, 6.4 km NNW S. Pittsburg, 28 August

Cave, on Tennessee-Alabama Line, 13 January 1957,

rion County, Nickajack Cave, July 29 1967, coll. S.

cutt Cave, 7 mi. NW S. Pittsburg, 28 August 1968, S.

SE of Monteagle, 35°10' N, 85°48' W, elev. ~200 m,
Hedin, R. Keith, J. Starrett and S. Thomas (2 females,
1 male) — Alabama; Jackson County, Guess Creek Cave, ~4.8 km E Trenton, 34°45' N,

Set Next Word Relation

Cut Stream Before

Cut Stream After

Click Predecessor

Click Successor

Make Stream

Show XML View of taxonomicName

Peck

Peck

15 A

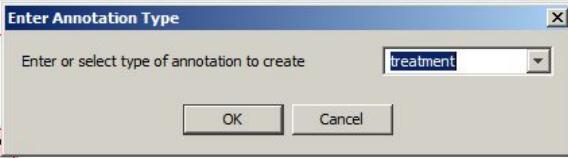
Fig XX Start manually marking treatment boundaries by selecting the first word of the treatment, and selecting “Start Annotation” from the context menu



Gertsch 1984, figs. 121–123. Add annotation starting from 'Nesticus' across populations (Figs. 9 & 10), but the condition seen in the Tate Springs male is very close to that illustrated by Gertsch (figs. 121–123). Again, we interpret this as geographic variation in a single, dispersal-limited, species.

Although *N. valentini* has page priority, we prefer to merge the single *N. valentini* population with the many known *N. barri* populations. This decision is consistent with Article 24.2 (determination by first reviser) of the International Code of Zoological Nomenclature (ICZN, 4th edition).

DNA sequences. — Mitochondrial 16S (Genbank accession numbers AF004596–99) and nuclear ITS (AF003769–70) DNA sequences for several populations of *N. barri* were reported in Hedin (1997a).



Nesticus stygius Gertsch 1984
Figs. 1, 11–12

Nesticus stygius Gertsch 1984: 36, figs. 170–172.

Material Examined; Holotype. — Female, U.S.A; Tennessee; Overton County, Obe Lee Cave, coll. 21 December 1958, T.C. Barr, F. Breeding, & C. Kacsur (housed at AMNH).

Additional Type Locality Material Examined. — Male and female, U.S.A; Tennessee; Overton County, Obe Lee Cave, 36°12'23" N, 85°15'08" W, elev. ~400 meters, coll. 11 October 1993, M. Hedin & C. Phillips (MCII male specimen #1882; MCII female specimen #1885).

Other Material Examined. — U.S.A; Tennessee; Overton County, Raven Bluff Cave, NW of Allons, 36°29'33" N, 85°21'36" W, elev. ~275 meters, coll. 1 October 1991.

Fig XX Finish manually marking treatment boundaries by selecting the last token of the treatment, and selecting “treatment” from the Enter Annotation Type dialog box

Once treatments are marked, it is advisable to check the paragraph boundaries. Be sure both paragraph and block are selected from the Display Control. In this example, separate lines of text at the beginning of the treatment have been erroneously placed within the same paragraph. Click within the paragraph and select “Revise Block Paragraphs” from the context menu. Select appropriate options as necessary; in this example, “Make each line a separate paragraph”.

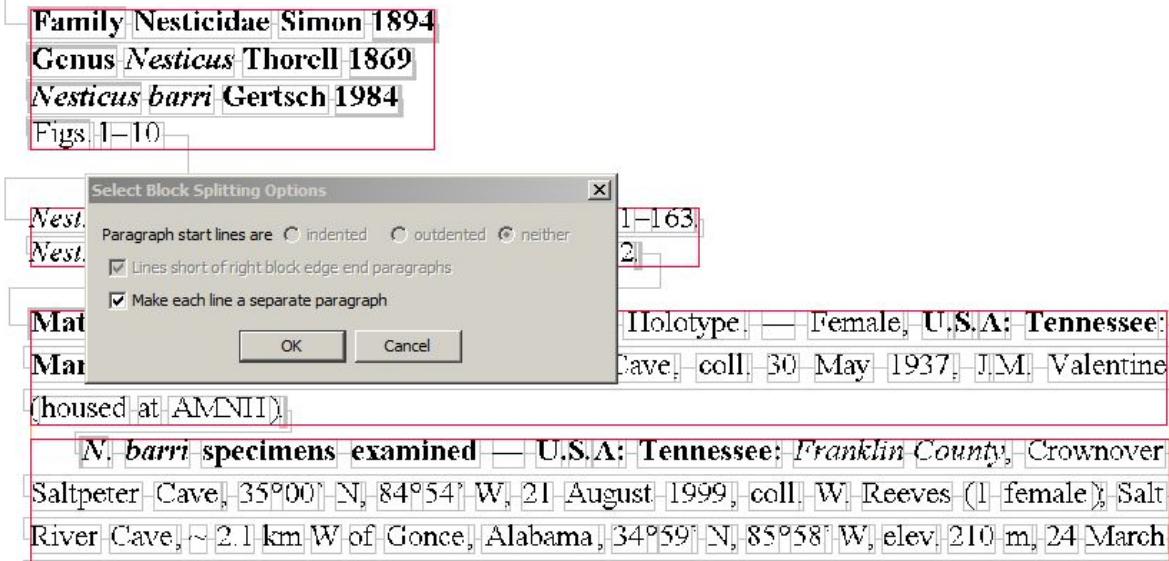


Fig XX Use Revise Block Paragraphs to efficiently split each line into its own paragraph

Treatment Structure

From the Tools menu, select “Treatment Structure”. This brings up a dialog box that can be stepped through using the “OK & Next” button.

The first line is typically the name of the taxon that this treatment concerns; this should be marked as “nomenclature”. References to previous literature on this species are marked as “reference_group”. Other typical treatment sections include diagnosis, description, materials_examined, etymology, biology_ecology; additional text can be categorized as discussion.

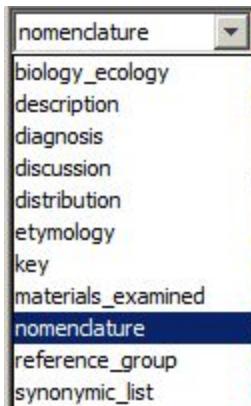


Fig XX The Treatment Structure Tool also allows users to designate treatment paragraphs according to the categories shown.



Check Substructure of Treatments - (2 of 5)

What to do in this dialog? (click to collapse)
Please select to which data domain (e.g. nomenclature or description) of the treatment these paragraphs belong.

nomenclature	Nesticus stygius Gertsch 1984
description	Figs. 1, 11-12
reference_group	Nesticus stygius Gertsch 1984: 36, figs. 170 -172.
materials_examined	Material Examined: Holotype. - Female, U.S.A: Tennessee: Overton County, Obe Lee Cave, coll. 21 December 1958, T.C. Barr, F. Breeding, & C. Kacsur (housed at AMNH).
materials_examined	Additional Type Locality Material Examined. - Male and female, U.S.A: Tennessee: Overton County, Obe Lee Cave, 36°12'23" N, 85°15'08" W, elev. ~ 400 meters, coll. 11 October 1993, M. Hedin & C. Phillips (MCH male specimen #1882; MCH female specimen #1885).
materials_examined	Other Material Examined. - U.S.A: Tennessee: Overton County, Raven Bluff Cave, NW of Allons, 36°29'33" N, 85°21'36" W, elev. ~ 275 meters, coll. 1 October 1991, M. Hedin, K. Crandall & A. Gerber (MCH female specimen #1018; MCH male specimens #1019, #1020).
discussion	Remarks. - Gertsch (1984) lists a female from Webb Cave, Putnam County, Tennessee. Webb Cave is actually in Overton County.
diagnosis	Diagnosis. - The troglomorphic features of <i>N. stygius</i> readily distinguish this species from smaller, surface-dwelling <i>Nesticus</i> found further east and north in the Appalachians. Compared to other large-bodied Appalachian troglomorphs, males of this species are distinguished by the unique shape of the dorsal paracymbial process (Fig. 11), and the distinctive thickened and chisel-like regular apophysis (Fig. 12). The narrow median septum and bulging posterolateral edges of the epigynum distinguish females (Gertsch 1984).
description	Description of male from type locality (MCH male specimen #1882) (Figs. 11 & 12; Table 1). - Color of appendages and cephalothorax uniform in coloration, approximating HTML color "dark orange" (#FF8C00). Abdomen darker, more flesh-colored, between "dark salmon" (#E9967A) and "gray" (#808080). Lenses of secondary eyes evident, without pigmentation. No external evidence of median eyes. Leg formula 1423. Legs very long, leg I over 11 times longer than carapace width. Paracymbium of male palpus with translucent dorsal process, bowed dorsally, forward, wide at base then narrowing towards tip, shaped like a duck's head. Small paradistal process, distal paracymbium with fine serrations. Ventral paracymbium without modification. Regular apophysis of palp well-sclerotized, thickened, blade-like, curving anteriorly. Small cuticular process arising near posterior origin of regular apophysis. Median apophysis semi-translucent, approximately triangular, anterioromedial edge thickened.
description	Redescription of female from type locality (MCH female specimen #1885) (Table 1). - Gertsch (1984) describes the holotype female as a "pale, eyeless, long-legged troglobite", with "cephalothorax and appendages bright yellow, without pattern", "eyes obsolete or nearly so". The redescribed female is mostly consistent with this description, although coloration is not "bright yellow". Color of cephalothorax and legs similar to male, but with lighter, unpatterned abdomen. Lateral eyes only apparent as small patches of pearl-colored cuticle. No external evidence of median eyes. Leg formula 1423. Leg I almost 11 times longer than carapace width. Epigynum as illustrated by Gertsch (1984), with narrow median septum, pear-shaped anterior lateral fovea, posterolateral edges thickened, rounded, and bulging conspicuously.
description	Variation. - Two males from Raven Bluff Cave, Overton County, Tennessee with paradistal process of paracymbium more pronounced, drawn into three-sided sharp point, otherwise similar to described male.
biology_ecology	Natural History. - Both Obe Lee and Raven Bluff caves are moist caves, Raven Bluff with a small (permanent?) stream, Obe Lee with a wet weather stream that flows into the mouth. Spiders were found beyond the twilight zone in both caves, occupying vertically-oriented webs in relatively high humidity crevices along cave walls. A female collected on 1 October, 1991 from Raven Bluff Cave carried an eggsac containing 22 eggs.
distribution	Distribution. - This species is known only from four caves on the western margin of the Cumberland Plateau in Overton county, north-central Tennessee (Fig. 1).

Previous | Cancel | OK & Next | Reset | Zoom Control - 1.0 +

Fig XX Example of treatment structure annotated using the Treatment Structure tool

Mark Materials Citations

From the Tools menu, select “Mark Materials Citations”. This brings up a dialog box that can be stepped through using the “OK & Next” button. This is a complex, multi-part tool. Note that the materials_examined sections marked in the treatment structure tool (above) are not utilized here.

The first stage of the Mark Materials Citations tool looks for text patterns consistent with collection codes and prompts to user to indicate whether or not they are (check the box if the bold text is a collection code).

The second stage refines which paragraphs contain materials citations and annotates these with country and major region names. Use the “Exclude paragraph” check box to indicate that the paragraph does not contain materials citations.

The third stage establishes boundaries between specimen records. Each record is represented as a line of text. Highlight the start of the first record and use “Split materialsCitation...” from the context menu; select text that is not materialsCitation and select “Remove ‘materialsCitation’ annotation” from the context menu. Use “Merge ‘materialsCitation’ annotations” to join parts of the same record spread across multiple lines.

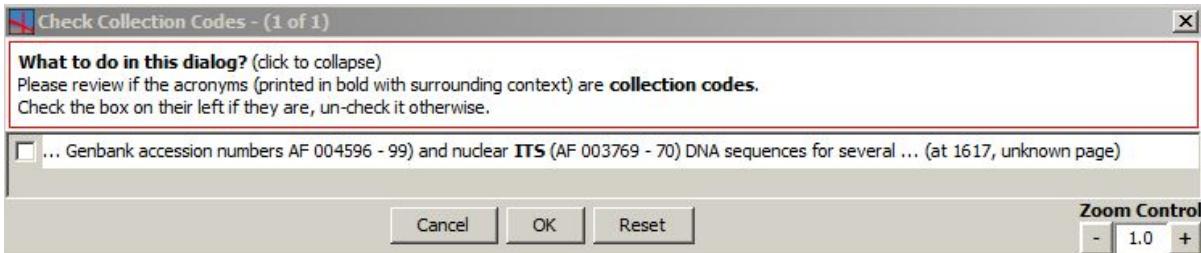


Fig XX The Mark Materials Citations tool first looks for collection codes

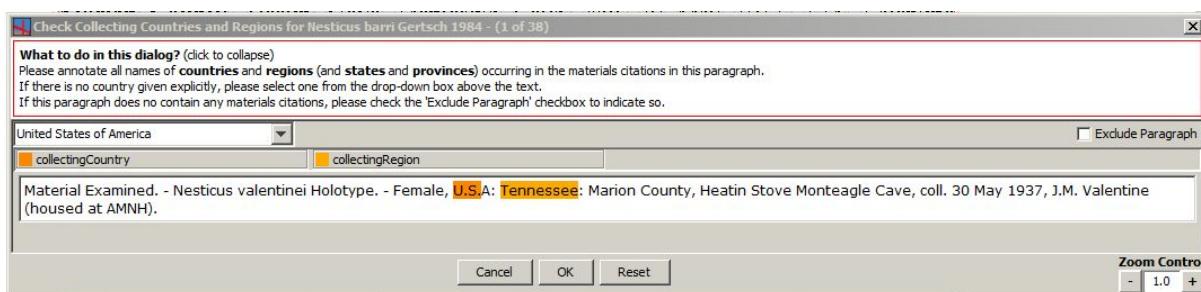


Fig XX The second stage of the Mark Materials Citations tool is to identify collecting countries and major regions

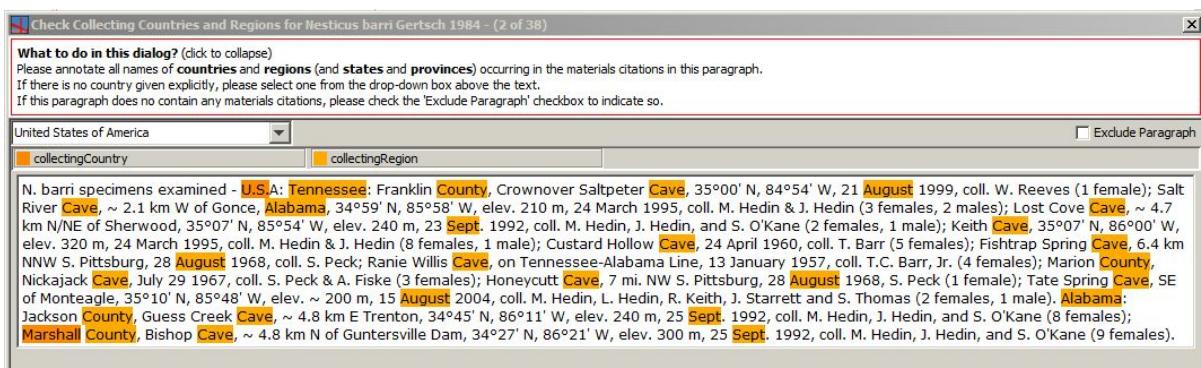


Fig XX The second stage of the Mark Materials Citations tool is to identify collecting countries and major regions. Select the country from the drop down menu, and highlight the appropriate text for the country (collectingCountry) and and region (collectingRegion). Imagine suggests country and region annotations; remove erroneously marked text by highlighting (one or more) and selecting “Remove...” from the context menu

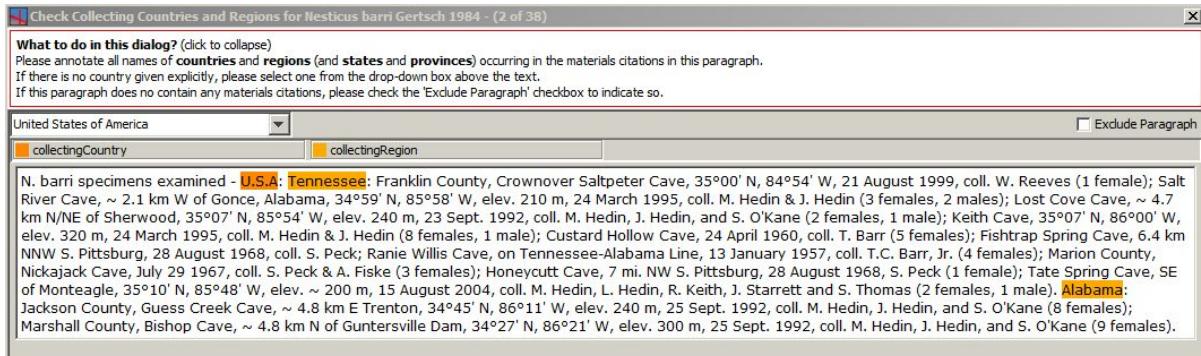


Fig XX Collecting country and region after manually removing erroneously marked text



Fig XX automated parsing of materialsCitations into individual records. Modify so that each yellow highlighted line represents one record.

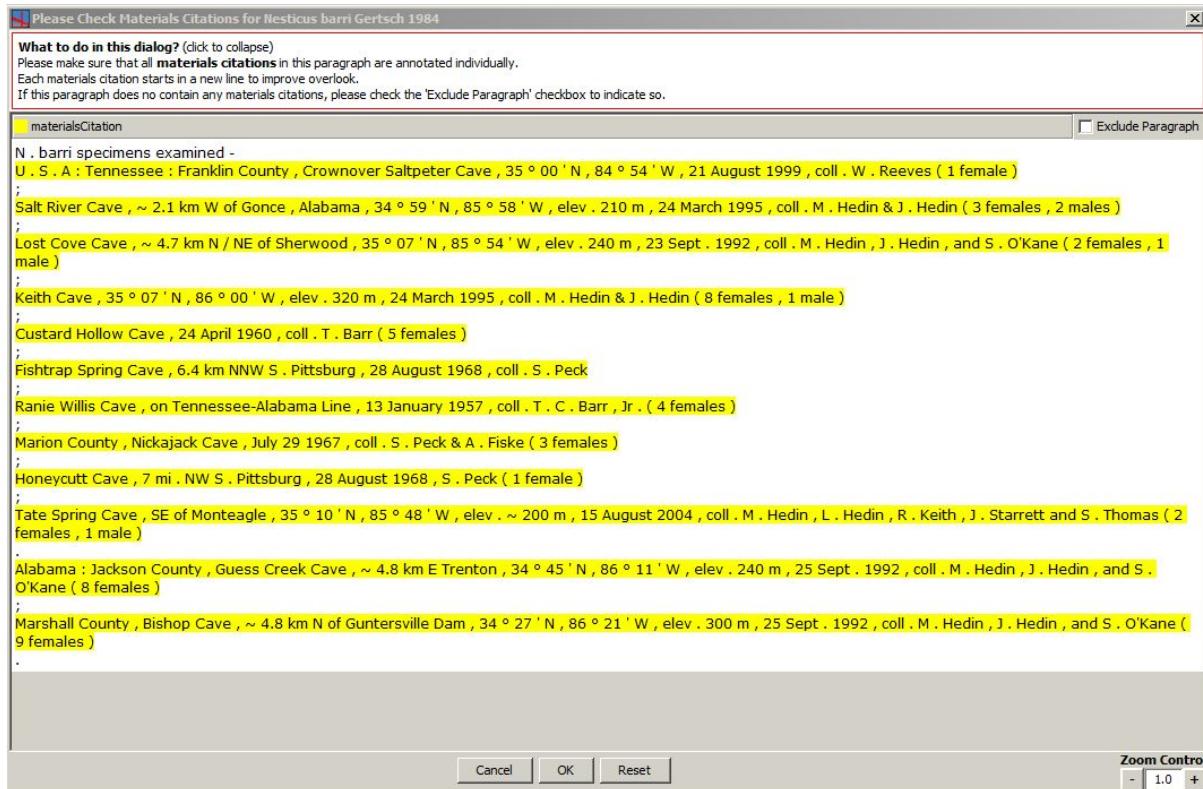


Fig XX manually corrected materialsCitations records

Parse Materials Citations

From the Tools menu, select “Parse Materials Citations”. This tool works much like the Parse Bibliography tool. Check the suggestions for assigning elements of the record to the appropriate field. Selected text can be assigned to a field by clicking on the appropriate label above. Note that text can belong to only one field, so any interfering highlights must be removed first. To do this, highlight the problematic text and select “Remove...” from the context menu.

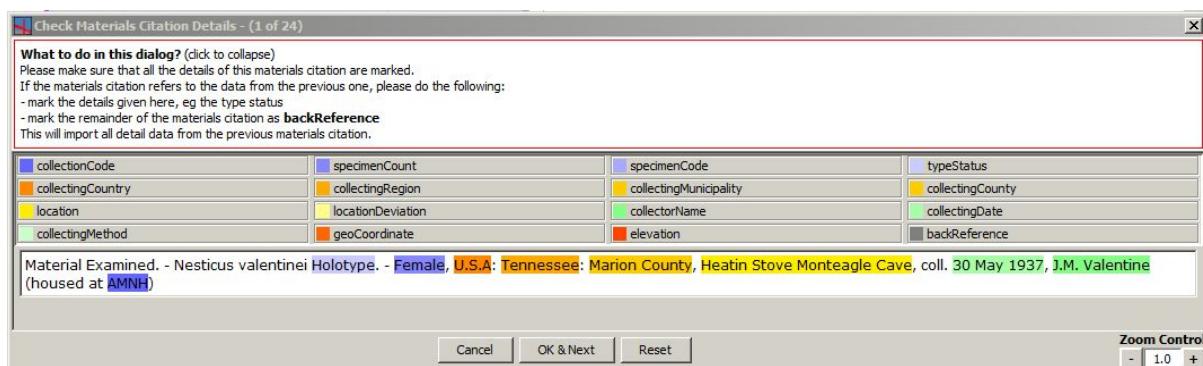
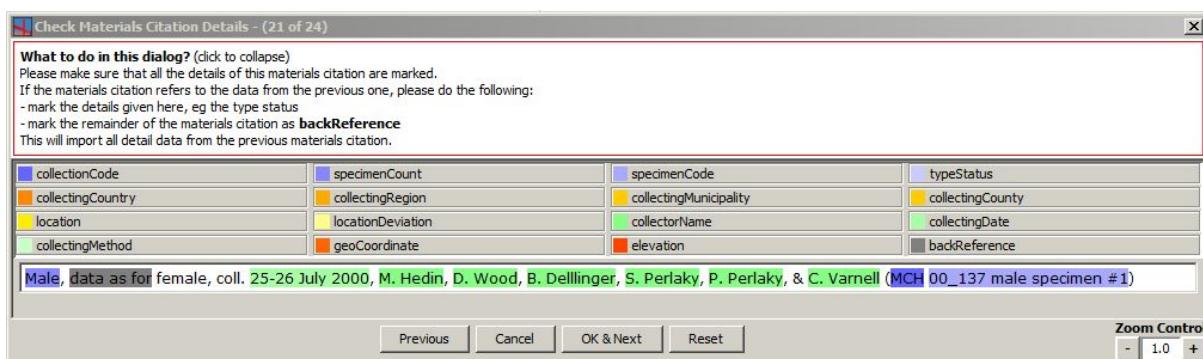




Fig XX Parse materials citations



Fix XX use backReference for materialsCitations that duplicate data in previous records. Double check such records later using Edit Attributes and correct as necessary

Edit Attributes

Edit attributes is an advanced function for refining the details of mark up. This tool is especially useful in the context of materials citations, and also bibliographic references, but can also be used to inspect or edit virtually any data element in the document.

There are two main ways to access the Edit Attributes dialog: in standard Imagine interface, and through the XML view.

Editing and replacing wrong fonts and symbols

Select from the Edit menu “Edit embedded fonts”. Each changed symbol will be changed through the entire document, for example in cases where male and female symbols have been misapplied.

The ten most commonly changed symbols keep listed on the left hand menu.



Opening a scanned document

To view the OCR, make sure that the slider in the upper right corner is on the right side. If problems with the OCR are visible, then move the slider to left to see the underlying page image.

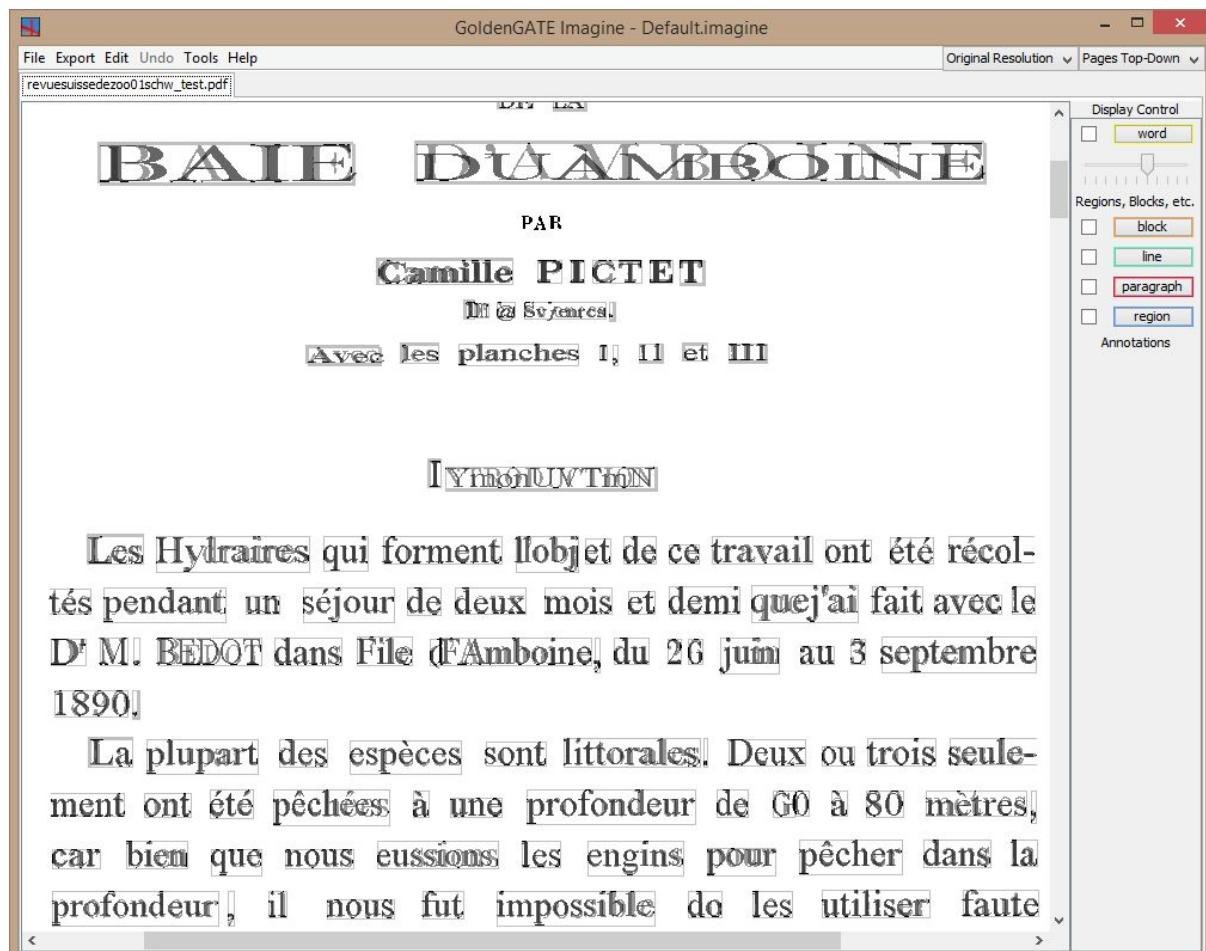


Fig. XX upper right: Slider to allows to regulate the transparency of the OCR result.

In case of OCR errors, click on the word

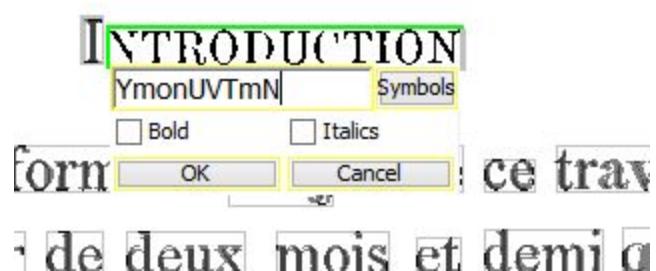


Fig. XX

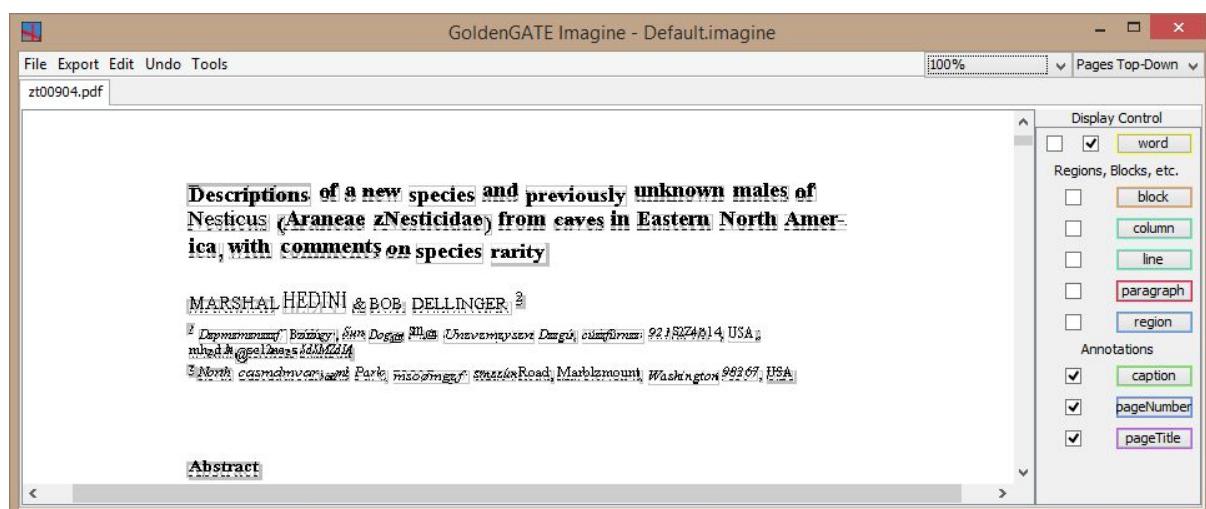


The original text is in the green box, and the box below can be used for editing. Special symbols can be selected through the “Symbols” button in the upper right corner.

Document structure

Though all the tools can be used at any time, detecting document structures

on the right hand appears the new annotations “caption”, “pageNumber” and “pageTitle”



To edit words (tokens), activate the right box left to the word in the upper left Display Control. Whilst highlighting a word, a popup menu will appear that allows the edit the word, choose from a symbol menu, and change the font to bold or italics

If a digital-born document is opened in the scanned mode, the Word display is different and allows editing the words.



, and *N.furtivus*: **tivuk:** Ge-
fema B... I... OK Cancel

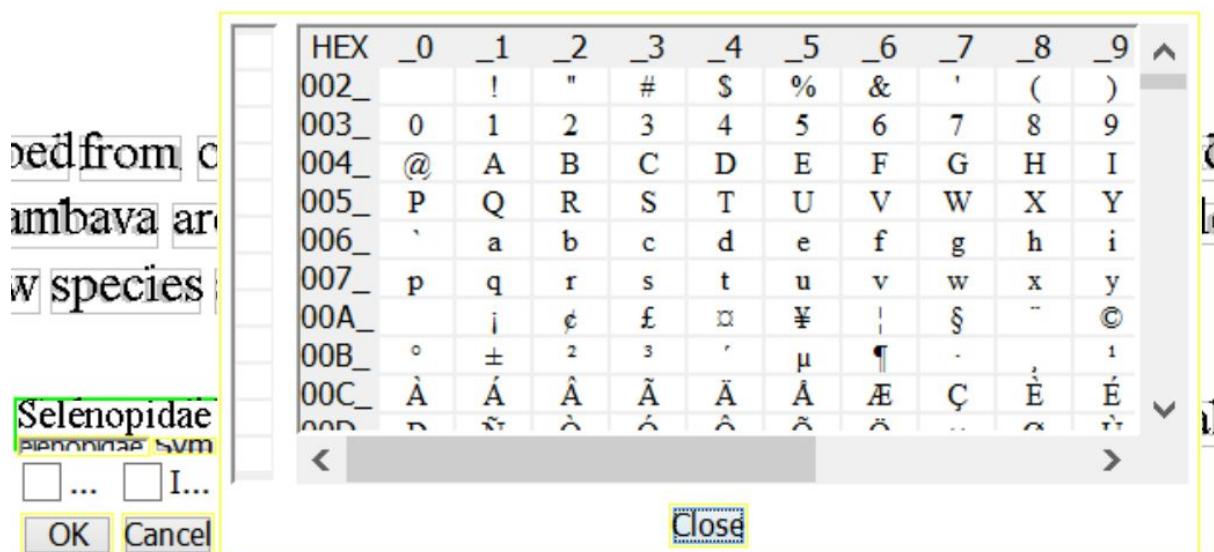


Fig. XX. Select special characters by clicking on the upper right corner of the popup menu. After the selection of a special character, the character will show up in the left column the next time the popup is called.

N. furtivus: G, and *N. furtivus*: G females are all tak



Annotating blocks of texts (names, treatments, etc)

The screenshot shows the GoldenGATE Imagine software interface. The main window displays a document page with the following text:

deposited in the American Museum of Natural History, New York (AMNH). All other specimens referenced with MCH numbers are currently housed in the personal collection of the first author. Upon completion of on-going studies of Appalachian *Nesticus*, these specimens will be donated to the AMNH.

Annotations are present in the text:

- Family Nesticidae Simon 1894**
- Genus *Nesticus* Thorell 1869**
- Nesticus barri* Gertsch 1984**
- Figs. 1–10
- Nesticus barri* Gertsch, 1984:36, figs. 121–123, 161–163.
- Nesticus valentinei* Gertsch, 1984:29, figs. 150–152.
- Material Examined.** □ *Nesticus valentinei* Holotype. □ Female, U.S.A.: Tennessee: Marion County, Heatin Stove Monteagle Cave, coll. 30 May 1937, JM. Valentine (housed at AMNH).

The right side of the interface features a "Display Control" panel with checkboxes for "word", "block", "column", "line", "paragraph", and "region". A vertical scroll bar is visible on the right edge of the main window.

scroll down to end of treatment and click on the last token.

The screenshot shows the GoldenGATE Imagine software interface with a modal dialog box titled "Enter Annotation Type". The dialog contains the following text:

Enter or select type of annotation to create materialsCitation

OK Cancel

The background document page contains the following text:

(904) action of the Tate Springs Add annotation starting from *Nesticus* nomenclature with *N. barri*. The paracymbium is particularly similar, with a distal and ventral paracymbium that is essentially unmodified, and a simple, thin dorsal process (compare Figs. 9 & 10; also see Gertsch 1984, figs. 121–123). There is variation in the shape of the tegular apophysis across populations (Figs. 9 & 10), but the condition seen in the Tate Springs male is very close to that illustrated by Gertsch (figs. 121–123). Again, we interpret this as geographic variation in a single, dispersal-limited, species.

Although *N. valentinei* population with the many Article 24.2 (determination Nomenclature (ICZN, 4th ed.)

DNA sequences — Mitochondrial 16S (Genbank accession numbers AF004596 – 99) and nuclear ITS (AF003769 – 70) DNA sequences for several populations of *N. barri* were reported in Hedin (1997 a).

***Nesticus stygius* Gertsch 1984**

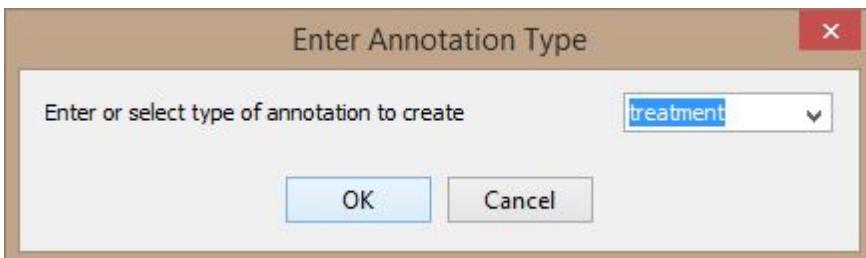
Figs. 11, 11–12

Nesticus stygius Gertsch 1984:36, figs. 170–172.

The right side of the interface features a "Display Control" panel with checkboxes for "word", "block", "column", "line", "paragraph", and "region". A vertical scroll bar is visible on the right edge of the main window.



select treatment



GoldenGATE Imagine - Default.imagine
File Export Edit Undo Tools
z00904.pdf

deposited in the American Museum of Natural History, New York (AMNH). All other specimens referenced with MCH numbers are currently housed in the personal collection of the first author. Upon completion of on-going studies of Appalachian *Nesticus*, these specimens will be donated to the AMNH.]

ZOOTAXA (904)

Family *Nesticidae* Simon 1894
Genus *Nesticus* Thorell 1869
***Nesticus barri* Gertsch 1984**
Figs. 1–10

Nesticus barri Gertsch, 1984: 36, figs. 121–123, 161–163.
Nesticus valentinei Gertsch, 1984: 29, figs. 150–152.

Material Examined. *Nesticus valentinei* Holotype, Female, U.S.A.; Tennessee; Marion County, Heatin Stove Monteagle Cave, coll. 30 May 1937, JM. Valentine (housed at AMNH).
N. barri specimens examined U.S.A.; Tennessee; Franklin County, Crownover Salt peter Cave, 35°00' N, 84°51' W, 21 August 1999, coll. W. Reeves (1 female); Salt River Cave, ~ 2.1 km W of Gonce, Alabama, 34°59' N, 85°58' W, elev. 210 m, 24 March 1995, coll. M. Hedin & J. Hedin (3 females, 2 males); Lost Cove Cave, ~ 4.7 km N/NE of Sherwood, 35°07' N, 85°54' W, elev. 240 m, 23 Sept. 1992, coll. M. Hedin, J. Hedin, and S. O'Kane (2 females, 1 male); Keith Cave, 35°07' N, 86°00' W, elev. 320 m, 24 March 1995, coll. M. Hedin & J. Hedin (8 females, 1 male); Custard Hollow Cave, 24 April 1960, coll. T. Barr (5 females); Fishtrap Spring Cave, 6.4 km NNW S. Pittsburg, 28 August 1968, coll. S. Peck; Ranie Willis Cave, on Tennessee-Alabama Line, 13 January 1957, coll. T. C. Barr, Jr. (1 female); Marion County, Nickajack Cave, July 29 1967, coll. S. Peck & A. Fiske (3 females); Honeycutt Cave, 7 mi. NW S. Pittsburg, 28 August 1968, S.

Display Control
Regions, Blocks, etc.
Annotations
 treatment

Annotate taxonomic names

highlight *Nesticus barri* Gertsch 1984

Pars taxon name



Export treatment of document to SRS

upload not use, but use “export to GoldenGate SRS”



go edit

Edit document metadata

Edit document metadata



Get Meta Data for Document zt00904.pdf

Publication Type: Book / Thesis / Monograph

Authors (use '&' to separate):

Title:

Year: Pagination:

Journal:

Part Designators: volume: issue: numero:

Publisher:

Location:

Editors (use '&' to separate):

Volume Title:

Publication Url:

DOI Identifier:

Handle Identifier:

ISBN Identifier:

ISSN Identifier:

Extract Validate OK Cancel

select journal article



Get Meta Data for Document zt00904.pdf

Publication Type:	Journal Article
Authors (use '&' to separate):	
Title:	
Year:	
Journal:	
Part Designators:	volume: <input type="text"/> issue: <input type="text"/> numero: <input type="text"/>
Publisher:	
Location:	
Editors (use '&' to separate):	
Volume Title:	
Publication Url:	
DOI Identifier:	
Handle Identifier:	
ISBN Identifier:	
ISSN Identifier:	

extract



Document View

Publication Type: Journal Article

More ... Author Title Year Pagination Journal Volume Publisher Location Editor Volume Title Publication Url

Zootaxa 904: 1–19 (2005) www.mapress.com/zootaxa/
Copyright © 2005 Magnolia Press
ISSN 1175-5326 (print edition) ZOOTAXA ISSN 1175-5334 (online edition)
Descriptions of a new species and previously unknown males of *Nesticus* (Araneae: Nesticidae) from caves in Eastern North America, with comments on species rarity
MARSHAL HEDIN1 & BOB DELLINGER2
1 Department of Biology, San Diego State University, San Diego, California 92182-4614, USA;
mhedin@sciences.sdsu.edu
2 North Cascades National Park, 7280 Ranger Station Road, Marblemount, Washington 98267, USA
Abstract
We describe the new species, *Nesticus pecki*, an apparently cave-limited (but not troglomorphic) species from the southeastern edge of the Cumberland Plateau in southern Tennessee. Previously unknown males of *N. dilutus* Gertsch, *N. stygius* Gertsch, and *N. furtivus* Gertsch are illustrated and described. These male specimens, along with additional females, are all taken from their respective type localities. *Nesticus valentinei* Gertsch is considered a junior subjective synonym

Close

highlight all the content of the highlighted boxes, and then click the respective box



Get Meta Data for Document zt00904.pdf

Publication Type:	Journal Article		
Authors (use '&' to separate):	Marshal Hedin1 & Bob Dellinger2		
Title:	ae: Nesticidae) from caves in Eastern North America, with comments on species rarity		
Year:	2005	Pagination:	1-19
Journal:	Zootaxa		
Part Designators:	volume: 904	issue:	numero:
Publisher:			
Location:			
Editors (use '&' to separate):			
Volume Title:			
Publication Url:			
DOI Identifier:			
Handle Identifier:			
ISBN Identifier:			
ISSN Identifier:			
<input type="button" value="Extract"/> <input type="button" value="Validate"/> <input type="button" value="OK"/> <input type="button" value="Cancel"/>			

Detect treatment structure

Run Tools, Detect treatment structure

Link to external resources

Scientific names

[Zoobank](#) and IPNI are the reference name servers for nomenclatorial acts that cover the creation of new taxonomic names and other acts governed by the respective Codes. There is no other system that covers the subsequent acts, and the Catalogue of Life covers only names that currently are considered valid, but without the track that ultimately ends with the original acts.



Since there is currently no agreement on the kind of (persistent) identifiers, Imagine has no generic format to link to external name services.

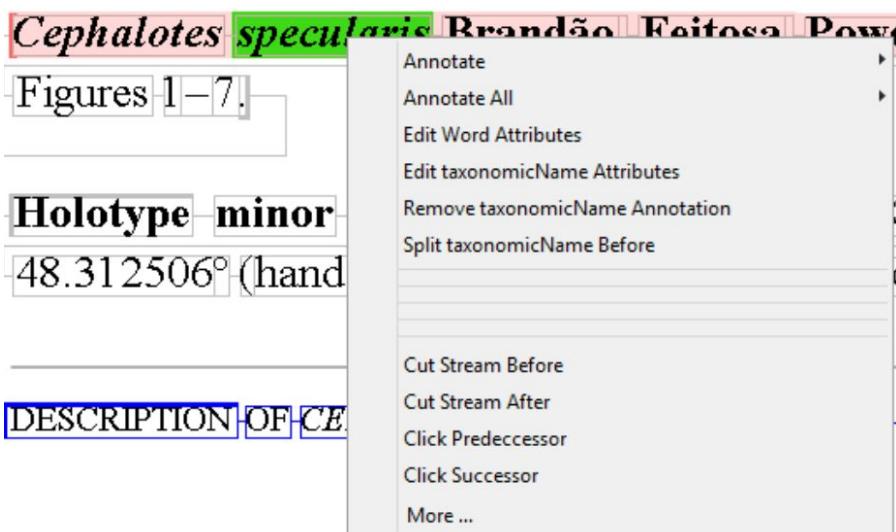
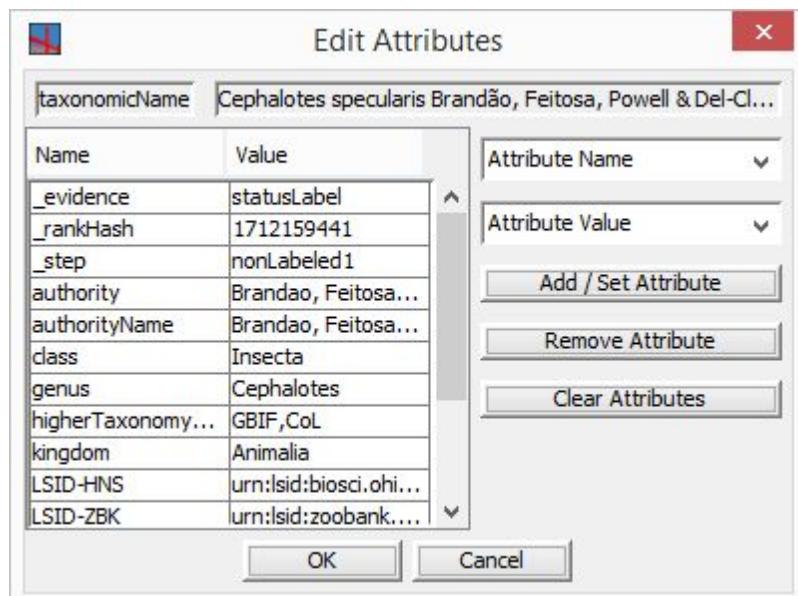


Fig.XX Annotation menu.

Select Edit taxonomicName Attributes



For Zoobank add AttributeName="LSID-ZBK", and as Value the URN, e.g. "urn:lsid:zoobank.org:act:B6C072CF-1CA6-40C7-8396-534E91EF7FBB".



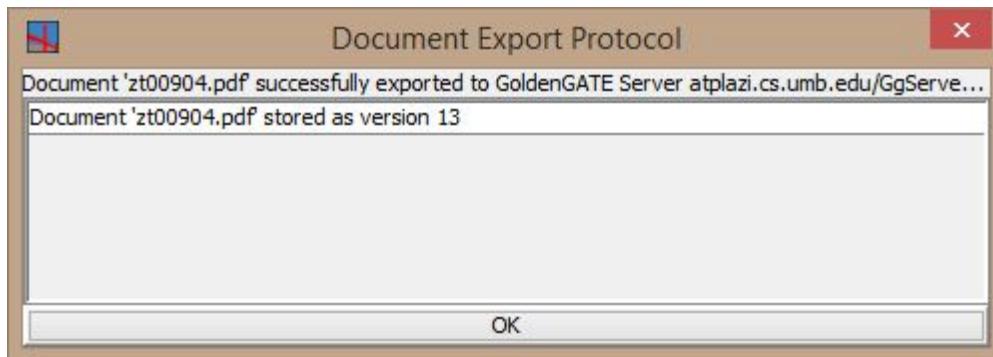
For Hymenoptera Name Server (HNS) add AttributeName="LSID-HNS", and as Value the URN, e.g. urn:lsid:biosci.ohio-state.edu:osuc_concepts:235700

Make sure to select Add/Set Attribute, OK.

Export

Minimal requirement for export to GG SRS

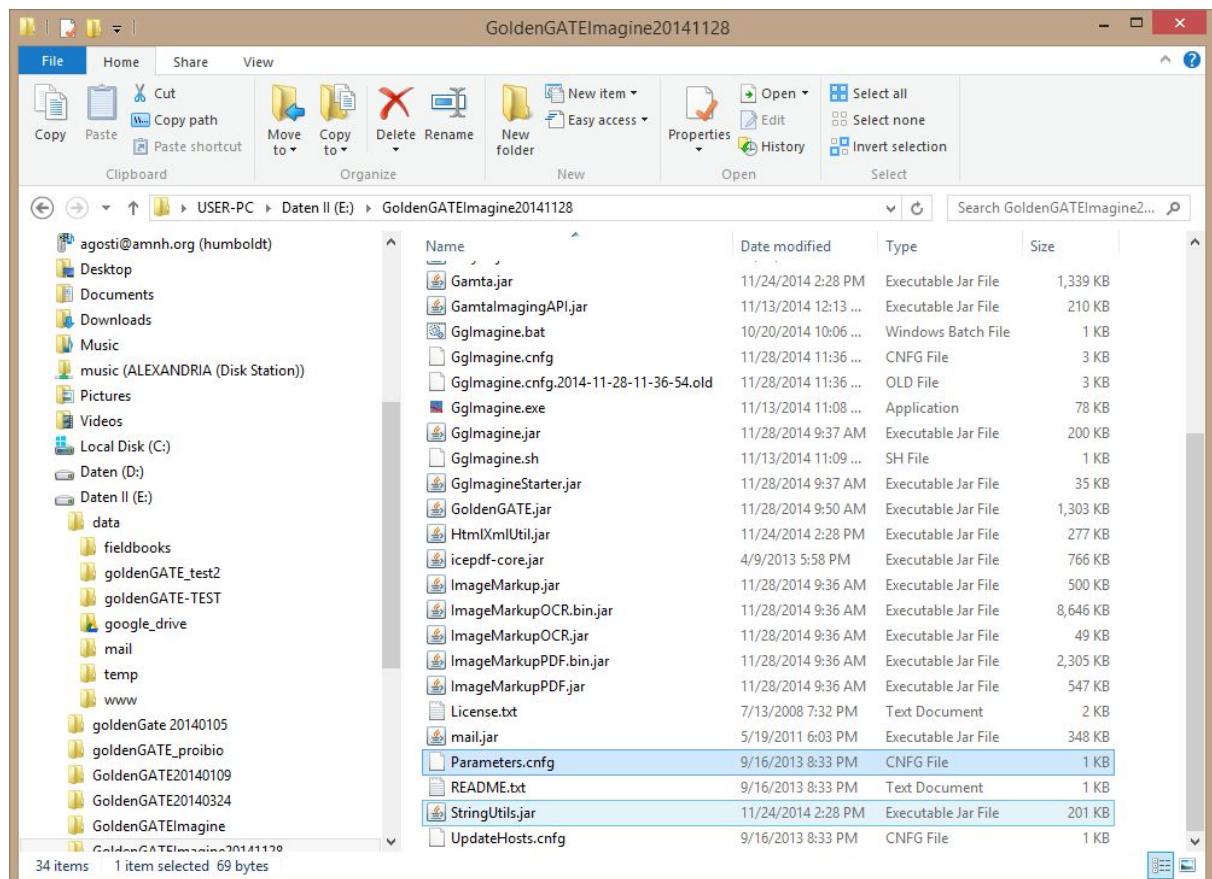
1. Metadata for article added
2. Treatment marked up
3. Nominate taxon name marked up



check

Trouble Shooting

File Format not valid



This message appears if Imagine has been opened from GoldgenGATE, NOT the GoldenGATE-Imagine version. Close GG and restart from the GG-I version

Loading of the file stalls

A reason can be that not enough memory is available for GG. To change memory allocation, open “parameters.cnfg” in a text editor, and in

```
START_MEMORY=128  
MAX_MEMORY=512  
LOG_SYSTEM_OUT=LOG  
LOG_ERROR=LOG
```

change

```
MAX_MEMORY=512
```



to

MAX_MEMORY=1536

Opening a scanned PDF with underlying OCR as born digital

Opening a non born digital pdf with underlying OCR with the option “born digital” results in the black image.





VILLEFRANCHE SUR MER, 1911
EDITION M. BÉDOS ET C. POCHE

LIBRAIRIE MARCET

ÉTUDE SUR LES FAUNES
BLUES D'IVIBOINE
PAR
Camille PISTET
PHOTOGRAPHIES
AVEC 100 planches en 12 et 16.

Introduction

Les Pycnoaires qui localement sont nommés *peperomia* ont été rencontrés pendant un séjour de deux mois et demi que j'ai fait avec le M. Redouté dans l'île d'Amboine, du 26 juin au 7 septembre 1911.

La plupart des espèces sont littorales, deux ou trois seulement ont été pêchées à une profondeur de 60 à 80 mètres, mais bien que nous étions très éloignés pour pêcher dans le profond, il nous fut impossible de trouver faire d'une approche aussi grande. Nous avons également plus rencontré que la pour profondeur paraît être très difficile. Hydrozoaires, tandis que la zone littorale au contraire est abondante en autres invertébrés de toutes classes, surtout en Anthozoaires, n'en rencontrant que très peu.

La plus grande partie des espèces a été rencontrée à marée basse soit de vieux pilotis ou sur des îles, soit dans la zone littorale proprement dite, parfois dans zones marées, soit dans la zone des eaux sombres. Leur nombre varie assez à 30 espèces, dont 7 *Symploblaues*, 25 *Calyptoclastes*, 70 possibles.

Env. Janvier de 1911 à 1912.



Page images generated under the wrong assumption (eg born digital vs not born digital):

In this case go into the GoldenGate directory /PagelImages and delete the respective images created during the initial conversion process.

The screenshot shows a Windows File Explorer window with the following details:

- Toolbar:** Includes Copy, Delete, Rename, New folder, New item, Open, Properties, Select all, Select none, and Invert selection.
- Address Bar:** (E:) > GoldenGATEImagine20150305 > PagelImages
- Search Bar:** Search PagelImages
- Table:** A list of files with columns: Name, Date, Type, Size, and Tags. The table contains the following data:

Name	Date	Type	Size	Tags
BB5E2D42FFDC7F24...	3/5/2015 5:25 PM	PNG image	33 KB	
BB5E2D42FFDC7F24...	3/5/2015 5:25 PM	PNG image	32 KB	
BB5E2D42FFDC7F24...	3/5/2015 5:25 PM	PNG image	42 KB	
BB5E2D42FFDC7F24...	3/5/2015 5:25 PM	PNG image	43 KB	
BB5E2D42FFDC7F24...	3/5/2015 5:25 PM	PNG image	39 KB	
FFCE9A74FFF93E5D...	3/5/2015 4:27 PM	PNG image	107 KB	
FFCE9A74FFF93E5D...	3/5/2015 4:27 PM	PNG image	87 KB	
FFCE9A74FFF93E5D...	3/5/2015 4:27 PM	PNG image	83 KB	



Editing and menu choices

Start window

tiling of pages

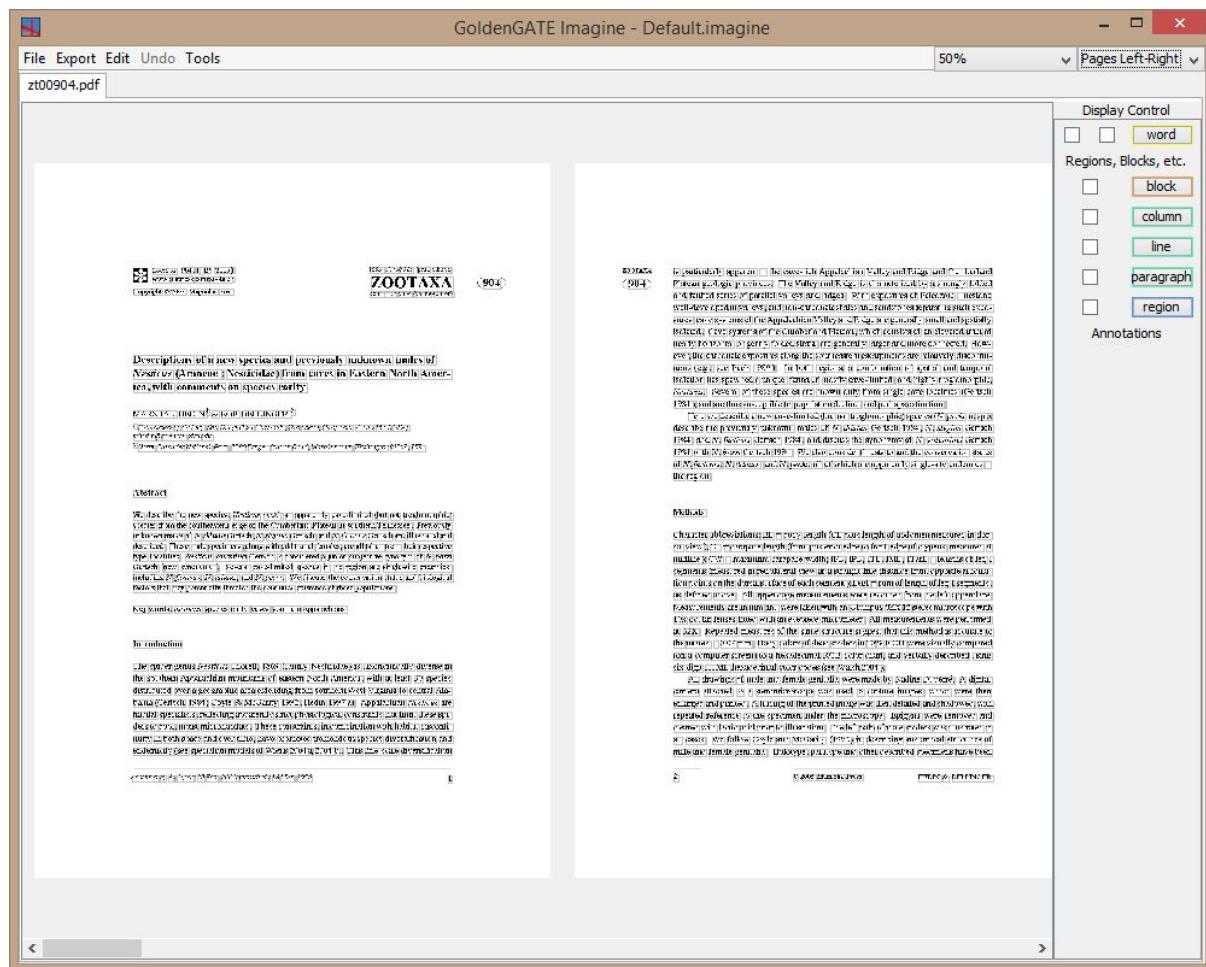


Fig: xx Tiling of pages: Lef-Right

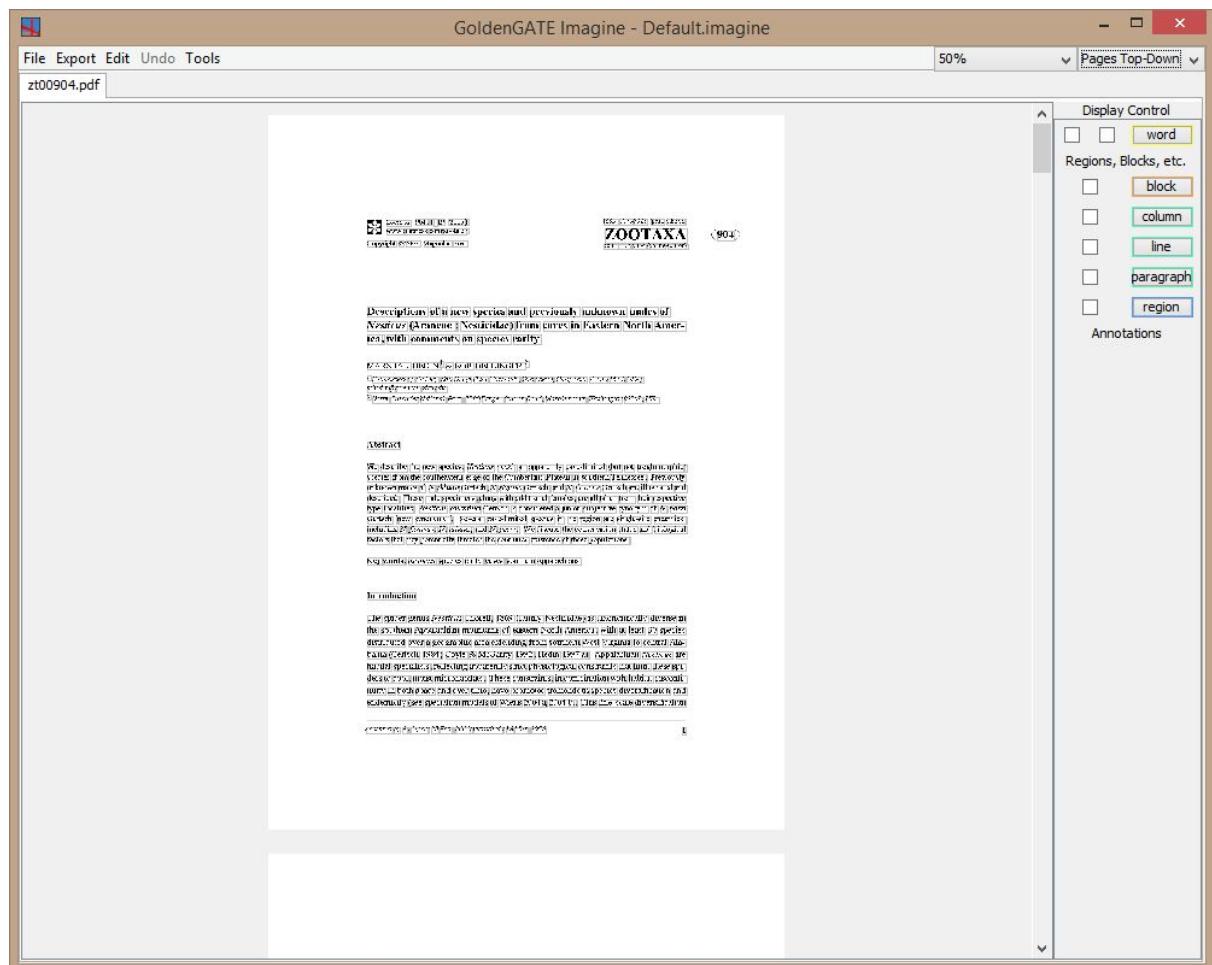
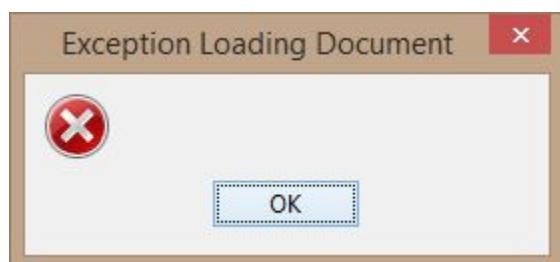
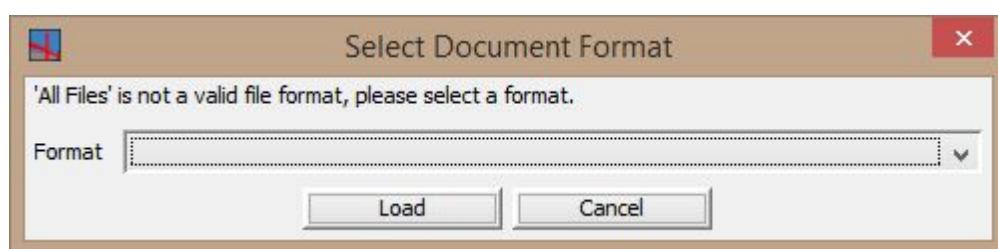
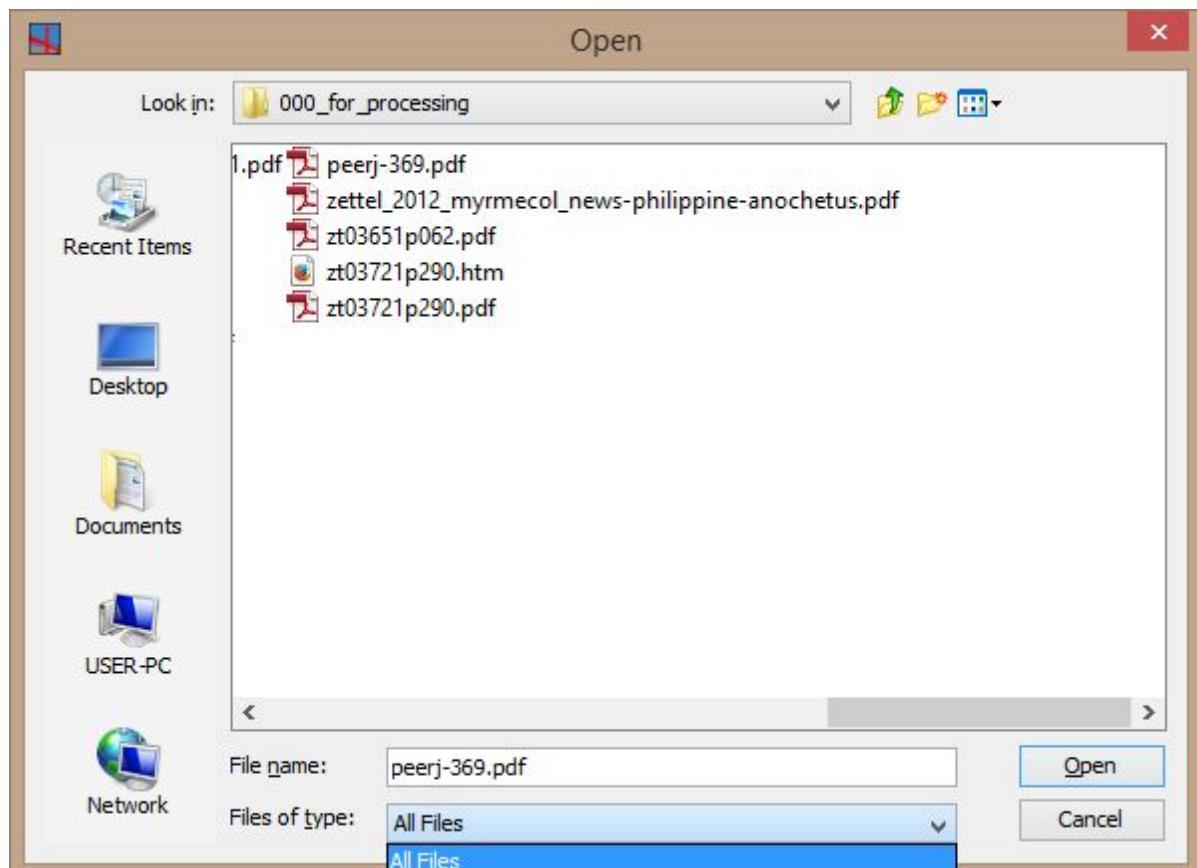


Fig: xx Tiling of pages: Top-Down



When opening the editor,



Glossary

GoldenGate Expressions

Word / Expression	Explanation
Token	A word, number or punctuation mark in the document. Treated as the atomic text unit by the Annotation Editor
Annotation	Representation of an XML tag in the GoldenGATE editor's data model
Type of an Annotation	The element name of the XML tag represented by the Annotation
Attribute of an Annotation	Representation of an attribute of the XML tag represented by the Annotation
Value of an Annotation	The textual content enclosed by the XML tag represented by the Annotation
Document Part	Special type of Annotation which can be edited as if it was a document of its own
Resource	Some object fulfilling some purpose related to handling or editing documents, or making editing more convenient
Annotation Source	Special type of resource for analyzing a document and creating Annotations marking some parts of it
Document Processor	Special type of resource for applying some processing to a document, whatever this might be for some particular document processor
Resource Provider	A sub-modul of the GoldenGATE editor managing and providing some type of resource (Annotation Source Provider and Document Processor Provider a have respective meaning)

TaxonX Expressions

Abstract	a summary of the present publication	A general element in modern scientific publications
Acknowledgements	a listing of acknowledgments to everybody who contributed to the	ACKNOWLEDGMENTS



	research leading to the present publication, often including a reference to the funding agencies	This work was supported by the National Science Foundation under Grant No. DEB-0344731 to B.L. Fisher and P.S. Ward. Fieldwork that provided (...)
Bibref	a bibliographic reference	Brown, 1974 or BROWN, W.L. 1974. A remarkable new island isolate in the ant genus <i>Proceratium</i> (Hymenoptera: Formicidae). <i>Psyche</i> 81:70–83.
Biology_ecology	content relating to the ecology, biology and behavior of the taxon	(...)this species is a subterranean nester and forager (...)
Citation	a reference to an earlier description or listing of the taxon described	<i>Proceratium avium</i> Brown, 1974: 71, figs. 1 and 2 (worker, gyne and male). Mauritius: Le Pouce Mt, 700- 800 m, Native forest, 1 Apr. 1969 (coll. W.L. Brown) [examined] AntWeb MCZTYPE32216 (MCZC)
Description	the morphological (and possibly molecular) description of the taxon	(...)Cranium (excluding median part of clypeus) entirely covered with decumbent / appressed hairs among which standing hairs are scattered [major]; (...)
Diagnosis	the characters which make his taxon unique and separate it from others, and often those allowing to recognize the taxon immediately	The following character combination differentiates <i>berlita</i> from all its congeners: scrobe absent, (...)
Discussion	anything which relates to the description, the nomenclatorial	<i>Discothyrea</i> of Madagascar belong to the first group...



	history, the behavior or comments relating to the taxon	
Distribution	a summary statement of the distribution of the species. The individual records are listed in "materials examined"	Distribution: North Borneo
Document head	the title, author and there addresses of the publication	
Etymology	the origin of the name of the taxon	The specific name is an arbitrary combination, to be treated as a noun in apposition.
Introduction	the introduction to the present paper; this is an element often marked with a specific title in recent publications or then the first general section in older publications proceeding the description of the taxa. Often general issues of the taxa are summarized, which is in modern paper more often to be found in the discussion.	A general element in modern scientific publications
Key	an identification tool to taxa. In most cases, these are dichotomous, that is a couplet of alternatives referring to the next and finally to a taxon name.	1. Red curved hair on occiput....2 Yellow straight hair on occiput3
Materials & methods	a section describing the techniques, measurements and methods used to derive the results in the respective publication	A general element in modern scientific publications
Materials examined	a listing of all the individual collecting events used in the description of this taxon, that is, not necessarily conclusively, a combination of locality, date, collector, sample number, habitat, etc.	Sample No. 4186; type locality: Poring Hot Spring, East ridge, 820 - 860 m a.s.l., Sabah, Malaysia (leg. Annette K. F. Malsch, 16.V.1998)
Multiple	Generally, anything that can not be assigned to any of the annotation above.	
Nomenclature	any elements pertaining to the naming of taxon according the International Code of Zoological Nomenclature	



Reference group	the section containing the bibliographic references <bibref> in the publication. This is an element hardly known in the old legacy literature	This refers to the “Literature cited section”, A general element in modern scientific publications
Synopsis	A list of taxa (e.g. all the taxa treated in a revision)	<i>Metapone</i> species: <i>M. leae</i>

Element Descriptions and Examples

The following table gives an overview about the TaxonX elements. A detailed description is available here <http://taxonx.org/documentation/taxonx1.xsd.html#h519182448>:

author	The author of the original description of a taxon in a nomenclature or synonymy section of a taxonomic treatment.
bibref	A bibliographical reference <tax:citation> <tax:xid identifier="doi:10.1046/j.1523-1739.1998.96177.x"/> SAFFORD, R.J., AND C.G. JONES. 1998. Strategies for Land-Bird Conservation on Mauritius. Conservation Biology 12:169–176. </tax:citation>
character	A morphological character.
citation	A bibliographic reference.
collection_event	Contains information regarding the collection of a specimen.
div	A block level textual division of a text. Attributes: n, number or name of division; type: type of division. If div occurs inside a treatment, suggested values are: abstract, acknowledgments, biology_ecology, description, diagnosis, discussion, distribution, etymology, introduction, materials_examined, materials_methods, multiple, synopsis.
figure	A figure or graphic.
figures	The statement identifying the figures related to a given treatment. <tax:nomenclature> <tax:name>Proceratium avium</tax:name> <tax:author>Brown</tax:author>, <tax:year>1974</tax:year>



	<tax:figures>Figs. 5-13.</tax:figures>...
head	A heading, such as the title of a section, etc.
locality	A geographical location.
name	A scientific name of a taxon as it appears in the source text.
nomenclature	The heading of a taxonomic treatment containing the scientific name of the taxon described.
note	A note, such as a footnote or endnote, in the source text. Use the place attribute to indicate the placement of the note in the source document (e.g., "foot", "end"). Use the n attribute to contain the number or symbol used to label the note in the source text.
p	A paragraph or other textual block. <tax:p> Venter very glossy. Ostiolar peritreme ligulate, gently curved, quite long, its apex nearly reaching lateral margin of plate. Rostrum reaching onto seventh abdominal sternite. Legs pale yellow, irregularly spotted and blotched with castaneous spots, terminal tarsal segment tending to become rosy. </tax:p>
pb	Page break. Indicates the point in the source text where a new page begins. Use the n attribute to record number of the new page; use the url attribute to link to an electronic graphical representation of the page.
ref_group	A group of bibliographic references.
seg	Segment. A phrase-level segment of text.
state	A character state.
TaxonX	Contains a single TaxonX document, including a TaxonXHeader and TaxonXBody (see also the documentation for taxonx: http://taxonx.org/documentation/taxonx1.xsd.html#h519182448
TaxonXBody	Contains a single text including at least one taxonomic treatment.
TaxonXHeader	Contains identification and description of the TaxonX document and its source, expressed in the Metadata Object Description Standard (MODS).
treatment	A taxonomic treatment



xid	External identifier. A pointer to an identifier assigned to the parent object in an external system. Contains optional attributes "identifier", the identifier, "source", the system in which the identifier can be found, "uri" a uniform resource identifier.
xmldata	A wrapper element used to include data from an external schema.
year	A year in a citation of a document, a scientific name, or a specimen.

A complete TaxonX documentation can be found at
<http://taxonx.org/documentation/taxonx1.xsd.html>.

GoldenGate Internal Markup Elements

taxonomicName	The scientific name of a taxon	Formica rufa
taxonomicNameAuthority	The author of a scientific name of a taxon	Formica rufa Linnaeus Formica rufa Linnaeus, 1758
taxonomicNameLabel	A qualifier indication the status of the taxon	Nov.sp., comb.nov.

Acknowledgments

This program is part of a bilateral digital library grant awarded by the Deutsche Forschungsgemeinschaft (DFG BIB47) and US National Science Foundation (IIS-0241229) to Klemens Boehm, Universität Karlsruhe (TH), and Christie Stephenson, American Museum of Natural History, New York. The preparation of the manual and its field testing has been supported by GBIF and Pro-iBiosphere (EU-FP7 program).

Notes

I've added an editor for the document meta data, as well as an export facility to push documents onto our servers. Don't worry exporting the



same document several times; the data model underneath GG Imagine has a well-defined way of generating UUIDs, which has effectively fixed the old duplication problems.

So, have a good time playing with it ;) You'll find what used to be the custom functions at the top of the 'Tools' menu. There is no paginator yet, so document cleanup is mostly handywork, but automation is in the works, and you don't have to do the cleanup before exporting a treatment anyway.

Exporting a treatment is as simple as this:

- open a document, say, a Zootaxa PDF
- mark a treatment
- mark its taxon
- parse the taxon
- enter the bibliographic meta data via "Edit -> Edit Document Meta Data"
- export the treatment via "Export -> Export to GoldenGATE Server"

If there are any problems, please save the logs and give them to me together with the problem files. This will help me fix problems and speed up making it a stable tool we can hand out to a wider range of people.

Notes and Questions (ex Jeremy)

Open PDF

Detect document structure (this will detect page numbers, so can omit this step) - not automatic

Mark captions, images, check headers and page numbers, text stream

Clean up text, sex symbols, paragraphs, tables, blocks

Parse bibliography

Step through bibliography references: Edit/List annotations, enter bibRef



Problem with text stream in paragraph:

Check taxon names

Mark taxon names

Mark treatments by hand

Check paragraph blocks (how to fix?)

Treatment structure

Mark materials citations

Parse materials citations

Edit document metadata

how to clean up noise associated with figures?

Tables?

male female OCR

how do you know if you are working within a selected treatments or all treatments? (got it)

Hard to click on the end of a section because usually ends with small box period

Hard to read text sometimes, and find the exact point to click on to get the text you need.

Hard to see the status of a selected object - what are its attributes? (e.g., name of treatment, materials



citations, location; or description)

The coloring system is not so easy to intuit and match up with the right menu - old gg was better in this respect; fill color makes sense - what does frame color mean?

Need to be able to step through materials citations records (next, previous; copy previous value would be nice)

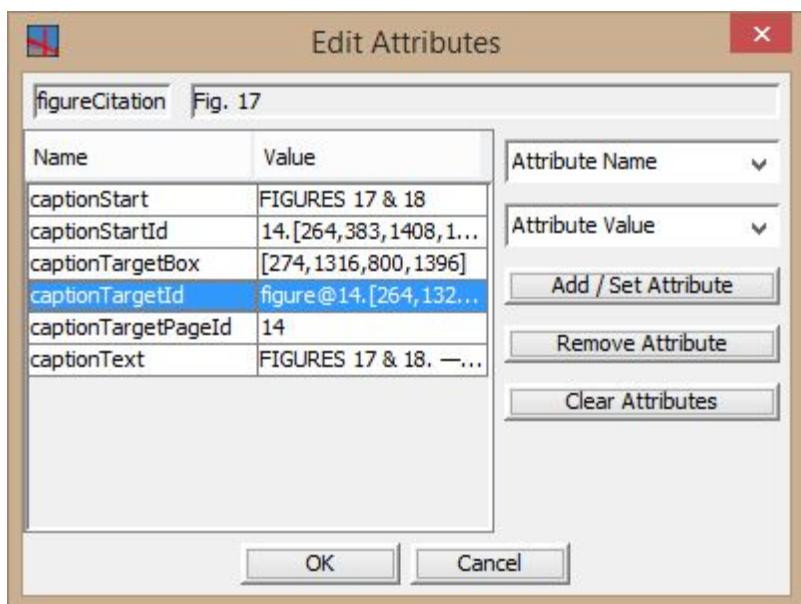
How to deal with noise in figures?

How to deal with inline tables? Example
8012_Peng_&_Xie_1995_Bul_10_57-64

Removing regions. When there are several of the same kind of region, you can't tell which you are removing before you remove it. Better to remove all of the same class at one time I think



Editing steps to be added



Editing attributes of a figure caption



GoldenGate List of Functions

PI: page image editor (edit page image & words, scanned documents only)

EM: edit menu (in main window)

TM: tools menu (in main window)

otherwise: context menu (showing after click or selection in document)

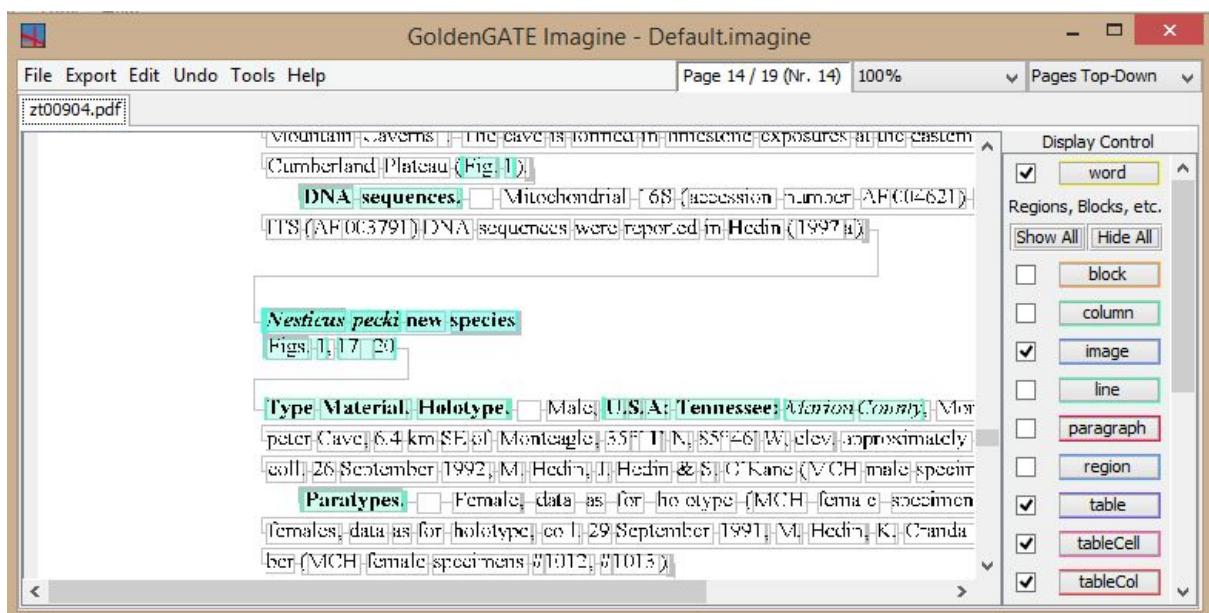


Fig. Main window (see above)

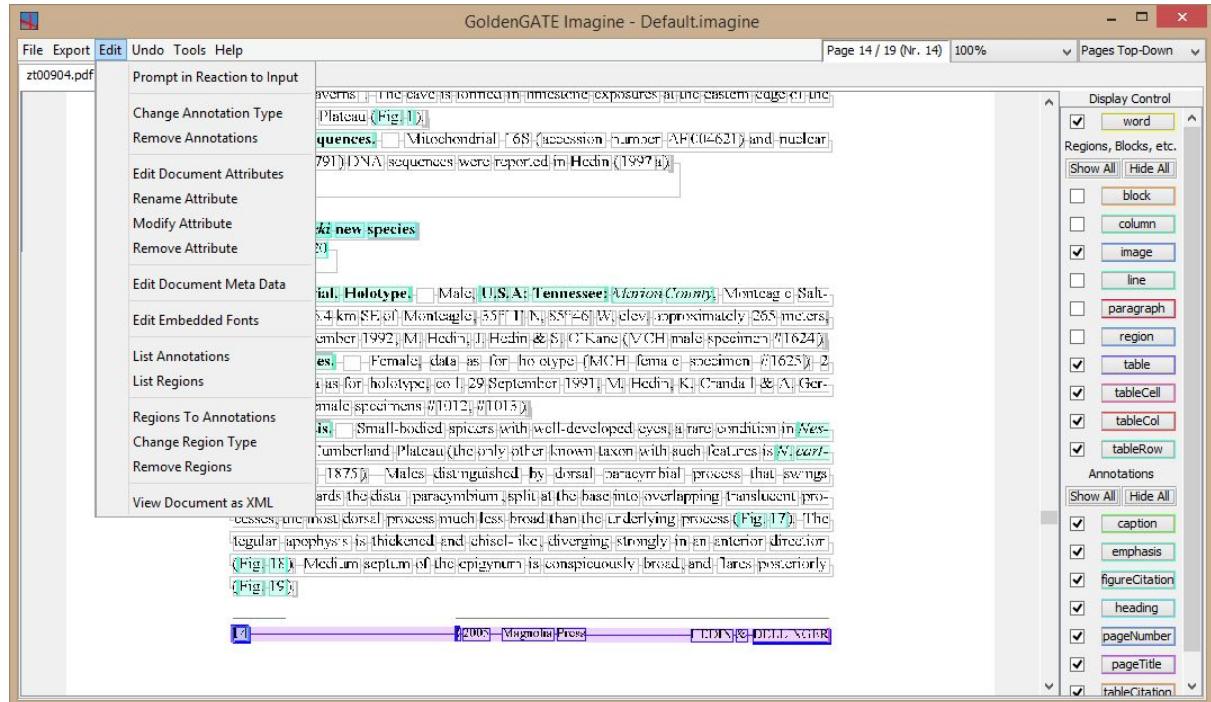


Fig Edit menu

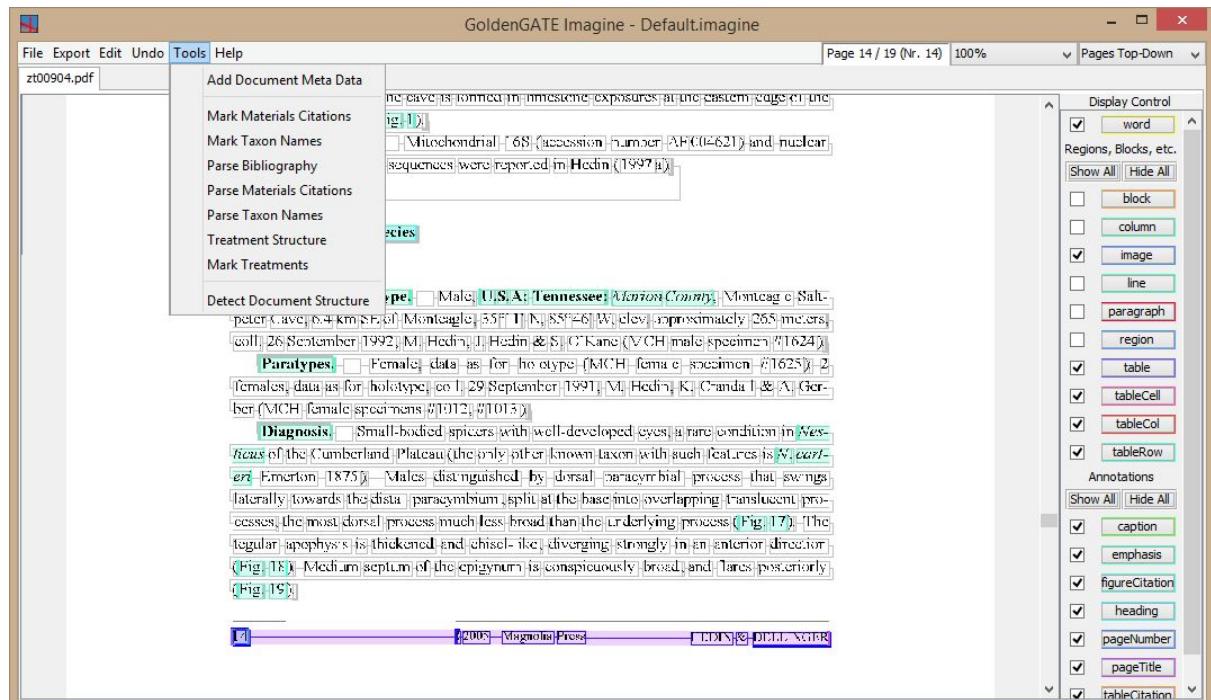


Fig Tools menu

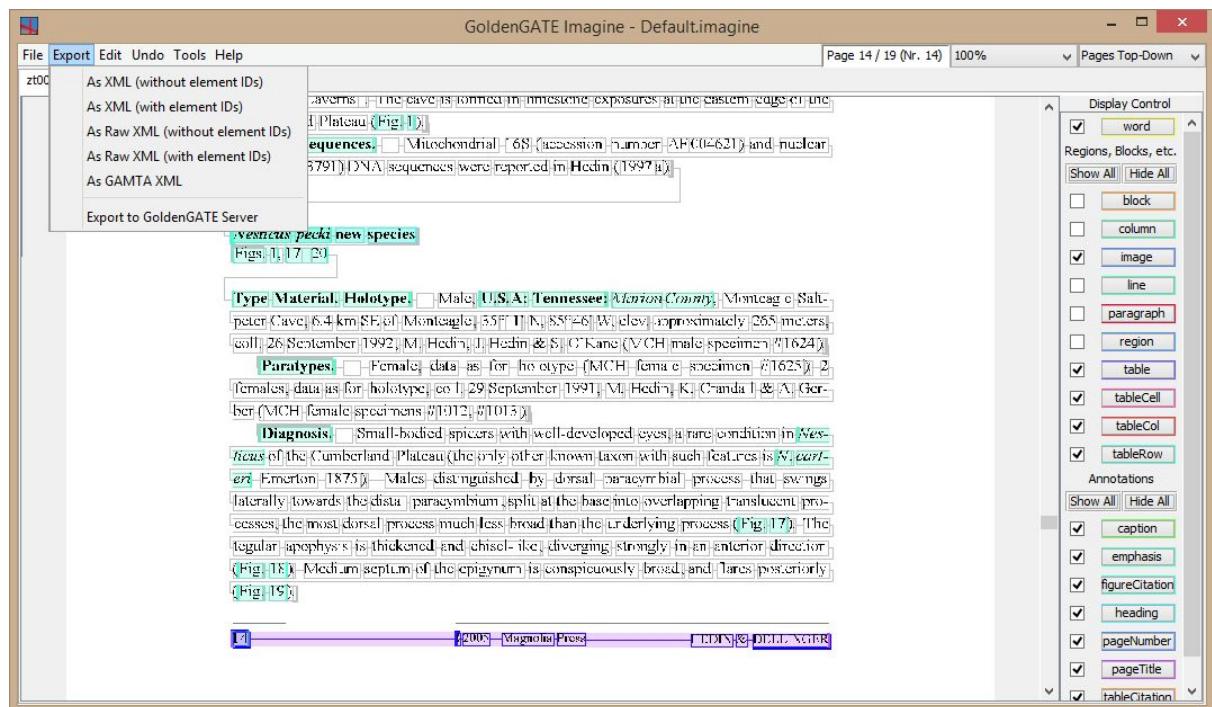


Fig. Export menu

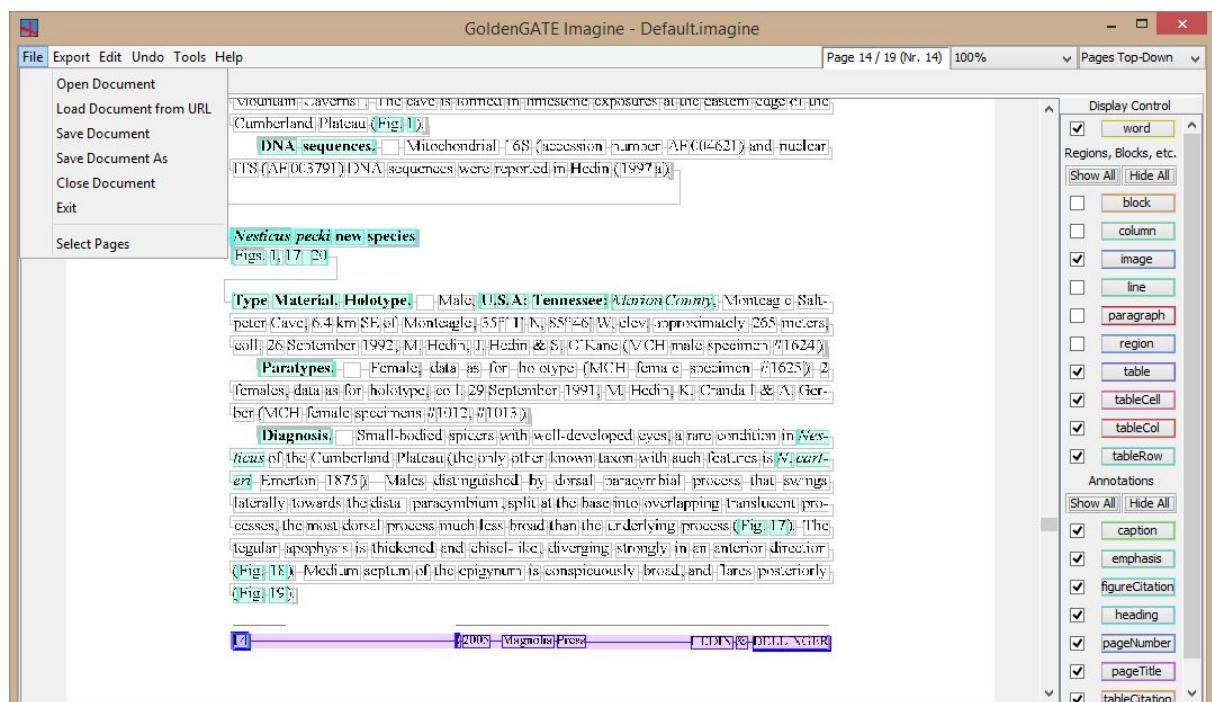


Fig. File menu

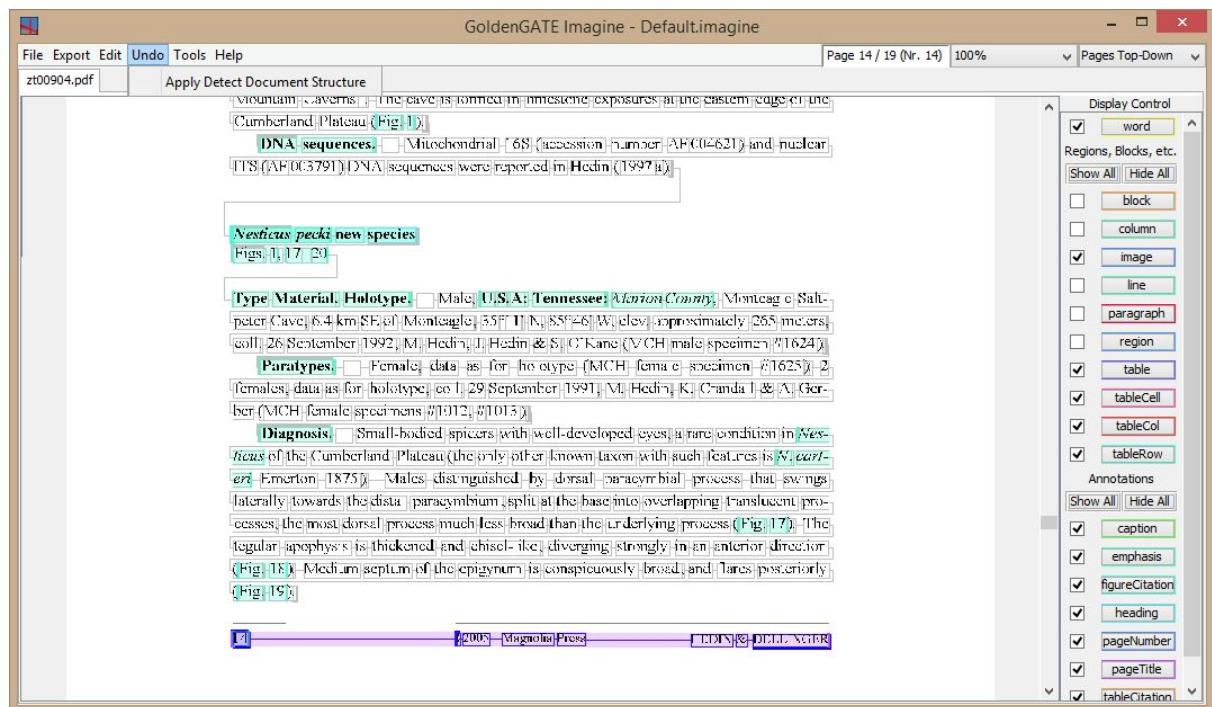


Fig Undo menu

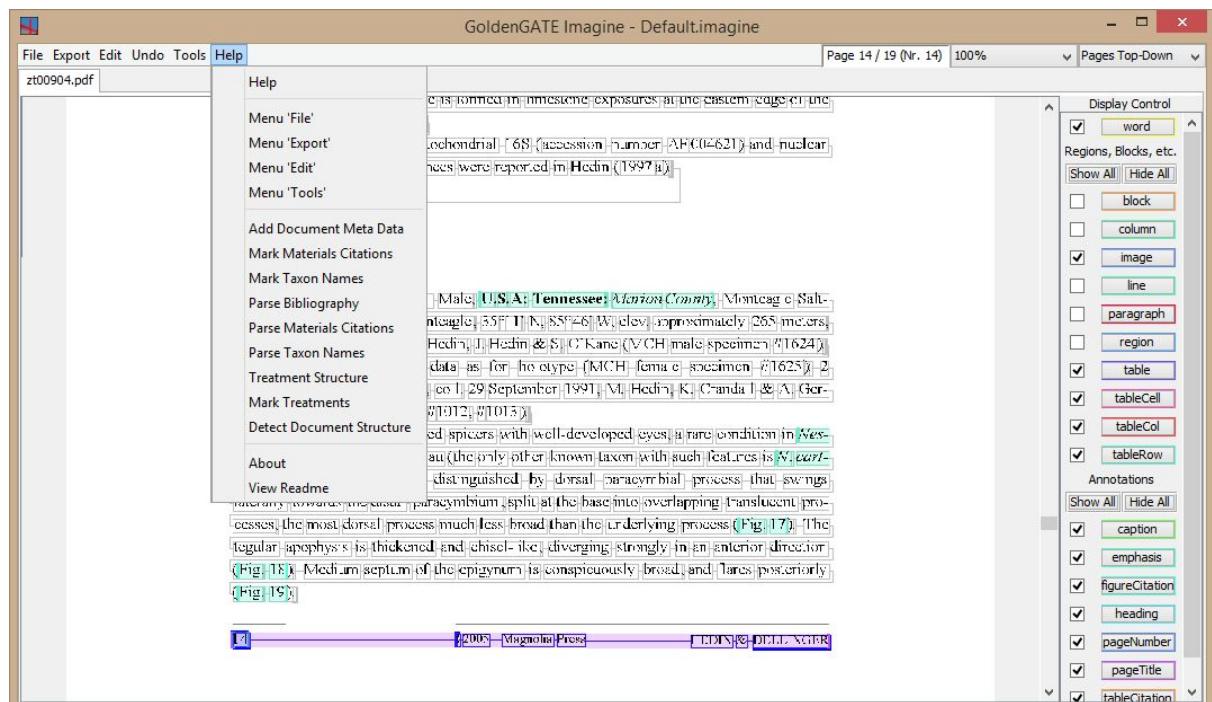


Fig. Help menu



Editors

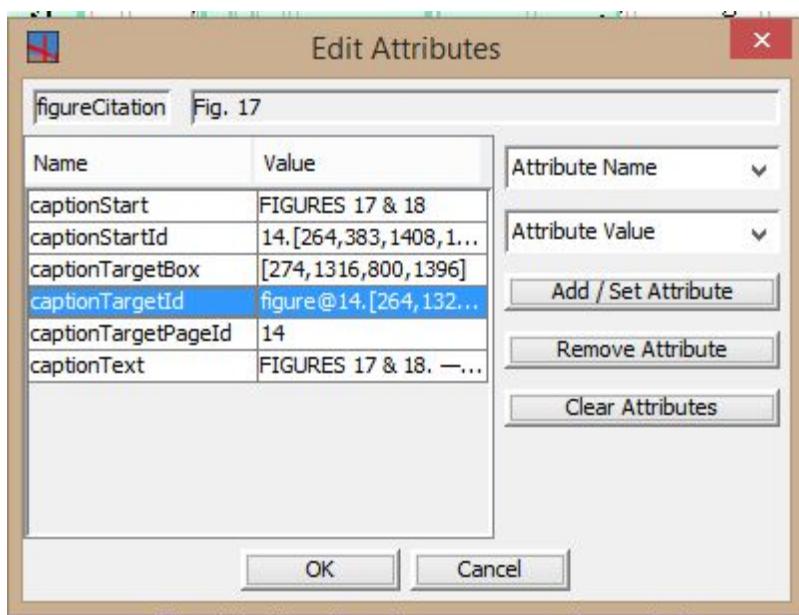


Fig. Attribute editor window

==== Annotation Actions

- annotate (annotate one or more words of a single text stream; annotation types visible in editor can be pre-selected from menu, as can some pre-configured ones)
- annotate all (annotates all occurrences of the string value of selected words; annotation types visible in editor can be pre-selected from menu, as can some pre-configured ones)
- edit word attributes
- edit annotation attributes
- remove annotation
- remove all (removes the selected annotation, as well as all annotations whose type and string value match the selected one)
- change annotation type
- start annotation (starts an annotation from a single selected word, to be completed at the next clicked word; good for annotations that span multiple pages)
- split annotation before (splits a selected annotation before a single selected word)
- split annotation after (splits a selected annotation after a single selected word)
- merge annotations (merges multiple selected annotations of same type into a single one)



- extend annotation (extends an annotation covering part of selection to include whole selection)
- copy text (copies the string value of a selected annotation to the system clipboard)
- copy XML (copies the a selected annotation to the system clipboard, possibly along with nested annotations)
- EM globally change type of annotations
- EM globally remove annotations

==== Attribute Tools

- EM edit document attributes
- EM globally rename object attribute
- EM globally modify object attribute
- EM globally remove object attribute

==== Basic Image Editing

- PI pens of different size (make black markings on page image, e.g. for repairing letters)
- PI erasers of different shapes (erase all non-white from page image, e.g. for cleaning up artifacts)
- PI flex area eraser (erases a custom-sized rectangle)
- PI block rotator (rotates a block of a page image by a custom angle)
- PI remove words
- PI mark words (automatically runs OCR on newly marked word)

==== Region Actions

- mark region (mark region of some type; region types visible in editor can be pre-selected from menu, as can some pre-configured ones)
- edit page attributes
- edit page image & words (opens current page in page image and word editor)
- mark image (removes all embedded regions, marks image region and shrinks it to remove white margins, attempts to link to caption marked within 1 inch below image)
- edit region attributes
- remove region
- change region type
- convert region to annotation
- split paragraph (splits a single partially selected paragraph, distributing lines above, inside and below selection; adjusts text stream word connections after split)



- merge paragraphs (merges two or more selected paragraph regions; adjusts text stream word connections after merge)
- revise block paragraphs (specify how to split a single selected block into paragraphs; adjusts text stream word connections after modification)
- merge blocks (merge several blocks into one, then re-detect lines and group them into paragraphs; adjusts text stream word connections after modification)
- EM globally change type of regions
- EM globally remove regions
- EM globally convert regions to annotations

==== Text Actions

- edit word string & layout attributes (scanned documents only)

==== Text Block Actions

- mark caption (makes selection into standalone 'caption' text stream, annotates caption, attempts to link to image block within 1 inch above caption)
- mark footnote (makes selection into standalone 'footnote' text stream, annotates footnote)
- mark page header (makes selection into standalone 'pageTitle' text stream, annotates pageTitle)
- mark parenthesis (makes selection into standalone 'mainText' text stream, annotates parenthesis)
- mark artifact (dissolves selection into individual words of type 'artifact')
- mark image (removes all selected words, as well as embedded regions, marks image region and shrinks it to remove white margins, attempts to link to caption marked within 1 inch below image)

==== Text Stream Actions

- make stream (cuts out selected words from parent stream, joins latter at the cuts)
- order stream ((re)arranges selected words to text stream in left-right and top-down order, links ends of selection to predecessor of first and successor of last selected word)
- set word relation (modifies relation between two adjacent layout words of same text stream)
- merge words (sets relation as multiple words to 'continued')
- merge paragraphs (sets word relations inside selection from 'paragraph break' to 'separate')
- copy text (put the string values of the selected words into the system clipboard)



- cut stream before (cut text stream before a single selected word)
- cut stream after (cut text stream after a single selected word)
- cut stream (cuts text stream between two selected words)
- make successor (makes second selected word successor of first selected word)
- click predecessor (set up next clicked word to become predecessor of currently selected word)
- click successor (set up next clicked word to become successor of currently selected word)
- set text stream type (modify type of text stream selected words belong to)

==== Block OCR

- PI run OCR on selected block (runs OCR on selected block, heuristically increasing line margin beforehand if necessary, and replaces existing block words with ones from OCR result)

==== Document Meta Data Editor

- EM edit document meta data

==== Document Structure Detector

- TM detect document structure (mark page titles, page numbers, captions, footnotes, tables, images, emphasized words in general, and headings in particular)

==== Font Editor

- edit font of word (modify how individual glyphs transcribe to Unicode, as well as font attributes; (partially) re-renders page images after font modification)
- EM edit fonts (modify how individual glyphs transcribe to Unicode, as well as font attributes; (partially) re-renders page images after font modification)

==== Drag&Drop Based Linking

- drop a URL to link to it (stores the dropped URL in specific attributes of certain annotations to establish links)

==== Table Actions

- mark table (mark selected block as a table, and identify columns, rows, and cells; makes table words into separate 'table' text stream, and orders stream based on table trid)
- split table row (splits a partially selected table row into two or three rows; adjusts table cells and text stream after split)



- merge table rows (merges two or more selected table rows into one; adjusts table cells and text stream after merge)
- split table column (splits a partially selected table column into two or three columns; adjusts table cells and text stream after split)
- merge table columns (merges two or more selected table columns into one; adjusts table cells and text stream after merge)
- update table (recomputes table structure after generic modifications)
- copy table data (copy data from table, using comma, semicolon, or tab as separator)
- connect table rows (logically connect rows in selected table to ones in table selected next, former on the left, latter on the right; tables need to have same number of rows for this to work)
- disconnect table rows (reverse of connecting rows)
- connect table columns (logically connect columns in selected table to ones in table selected next, former on the top, latter on the bottom; tables need to have same number of columns for this to work)
- disconnect table columns (reverse of connecting columns)
- copy table grid data (copy data from table and connected tables, using comma, semicolon, or tab as separator)

==== Object Listing

- EM list layout objects (lists layout objects in a dedicated dialog, supports XPath filters)
- EM list annotations (lists annotations in a dedicated dialog, supports XPath filters)
- TM list XYZ (lists layout objects or annotations with a pre-configured filters)

==== Custom Markup Tools

- TM various markup tools (apply specific markup tools; labels should be good hint, and they bring their own help texts)