

Component Report

Project Acronym: OpenUp!
Grant Agreement No: 270890
Project Title: Opening up the Natural History Heritage for Europeana

C4.1.2 Data quality service for zoological names prototype

Revision: V4a (Final)

Authors:

Tommi Koskinen (UH)
Hanna Koivula (UH)
Jörgen Palmgren (UH)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	



REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision	Date	Author	Organisation	Description
1	2011-07-15	Tommi Koskinen	University of Helsinki	The software is adjusted to use an SQL Server version of the Catalogue of Life database
2	2011-08-04	Tommi Koskinen	University of Helsinki	Implementation of combined result sets from two source portals (Catalogue of Life and eu.nomen.eu)
3	2011-08-15	Tommi Koskinen	University of Helsinki	Implementation of a concept check against the Catalogue of Life
4	2011-08-25	Tommi Koskinen	University of Helsinki	A new field added for monitoring the response time
4a	2011-09-06	Coordination Team	BGBM	Minor editing.

Statement of Originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Distribution

Recipient	Date	Version	Accepted YES/NO
Anton Güntsch, TMG leader	2011-09-06	4	YES
Project Coordinator	2011-09-06	4a	YES



Zoological quality service prototype

Purpose

The purpose of the zoological data quality service is to provide a machine readable back end data service. This service is to be used by a Data Quality Toolkit that is being separately developed by the OpenUp project.

Functions available in the prototype

The prototype provides an application programming interface (API) for searching and matching zoological scientific names. A locally installed version of the Catalogue of Life database running on an SQL Server is used in the background as the data source. The service provides query parsing with error handling and offers two search modes: single search mode and multiple search mode. A brief description of the service is available; this is called the Metadata service.

The concept search includes both synonyms and accepted names in the result set and the relevant status information. EU-nomen.eu via SOAP interface is available as an additional data source.

Features to be developed later

- support for fuzzy matching using a variant of Levenshtein distance
- search optimization

The URL structure will follow this pattern:

- names/col (using CoL only)
- names/nomen (using eu.nomen only)
- names/all (using all available sources)
- concepts/col (concept check using CoL)

Additional data sources will be cached to the local service database to optimize and fasten the fuzzy queries from all available sources. Additional sources such as the GBIF Sweden DYN TAXA, eu-nomen, and other local resources can be added when the agreement draft is accepted. Further efforts to harmonize the result set fields are needed, since all data sources offer a different set of fields.



Examples

A list of links is provided below, demonstrating the current functionality of the zoological name services. Please note that the result pages of these links are best viewed by installing the **JSONView plugin**. This plug-in is available both for the Firefox and Chrome browsers.

Name search

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/name/process?query={%22query%22:%22Poecile%20cincta%22,%22limit%22:3,%22type%22:%22name%22}>

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/name/process?queries={%22q1%22:{%22query%22:%22tribolium%20castaneum%22},%22q2%22:{%22query%22:%22branta%20leucopsis%22}}>

Concept search

The concept search returns both synonyms and accepted names of a search term. Status information is included.

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/concept?queries={%22q1%22:{%22query%22:%22pusa%20hispida%22},%22q2%22:{%22query%22:%22lyrurus%20tetrax%22}}>

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/concept?query={%22query%22:%22Poecile+cincta%22,%22limit%22:3,%22type%22:%22concept%22}>

Error handling with malformed search parameters

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/name/process?queries=q1%22:{%22query%22:%22tribolium%20castaneum%22},%22q2%22:{%22query%22:%22branta%20leucopsis%22}}>

Metadata service with no search parameters

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/name/process>

SOAP search

This is an example of a SOAP response from eu.nomen.eu.

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/nomen?query={%22query%22:%22tribolium%20castaneum%22,%22limit%22:3,%22type%22:%22concept%22}>



Combined result set

Here result sets from both source portals are combined.

<http://kirjotikka.fmnh.helsinki.fi/ci/index.php/names/all?queries={%22q1%22:%22query%22:%22tribolium%20castaneum%22},{%22q2%22:%22query%22:%22branta%20leucopsis%22}}>