Digital Herbarium Documentation

Release

Dmitry E. Kislov

Jul 30, 2017

Оглавление

Printable version

Глава

ONE

DIGITAL HERBARIUM HTTP-API DESCRIPTION

1.1 Intro

This document describes HTTP-API (Application Programming Interface over HTTP protocol) for working with the Herbarium Database.

HTTP-API is working in readonly mode. There is no way to make changes in the database using this API.

1.2 Description of request parameters

Only GET-requests are allowed when talking with the HTTP-API. To establish connection with the service, one can use either HTTP or HTTPS protocols.

Requests with multiple parameters, e.g. colstart=2016-01-01 and collectedby=Bak, are treated as AND-type queries: in this example, all records collected after 2016-01-01 and including bak (case insensetive inclusion is checked) as a substring in Collectors field of the main table of the database will be returned.

OR-type query behaviour can be emulated by series of consequent queries to the database and isn't natively implemented in the current version of the API.

List of allowed GET-parameters:

- family — family name (case insensetive, records should have the same family name as provided);
- genus — genus name (case insensetive, records should have the same gunus name as provided), note: if its value contradicts with the family name provided in the same request, an error will be returned as a part of json-response;

- species_epithet — species epithet (case insensetive, records should have the provided value as a substring in the correspongin field of the database);
- place — place of collection (case insensetive, records should have the provided value as a substring in the correspongin field of the database), if the parameter is given, searching is performed over the following fields: Place, region, district, note;
- collectedby — collectors (case insensetive, records should have the provided value as a substring in the correspongin field of the database); if the field is filled in cyrillic characters, searching will be performed on its transliterated value;
- identifiedby — identifiers; (case insensetive, records should have the provided value as a substring in the correspongin field of the database);if the field is filled in cyrillic characters, searching will be performed on its transliterated value;
- country — country name (case insensetive, records should have the provided value as a substring in the correspongin field of the database); the system knows country names according to standards ISO3166-1-ru and ISO3166-1-en; Russian Federation is replaced with Russia for short;
- colstart — date when collection was started (yyyy-mm-dd);
- colend — date when collection was finished (yyyy-mm-dd);
- acronym — name of the herbarium acronym (case insensetive, records should have exactly the same acronym as provided);
- subdivision — name of the herbarium subdivision/branch (case insensetive, records should have the provided value as a substring in the correspongin field of the database);
- latl — lower bound of latitude, should be in (-90, 90);
- latu — upper bound of latitude, should be in (-90, 90);
- lonu — upper bound of longitude, should be in (-180, 180);
- lonl — lower bound od longitude, should be in (-180, 180);
- synonyms — boolean parameter, allowed values are false or true; absence of the parameter in GET-request is treated as its false value; true value (e.g. synonyms=true) tells the system to search records taking into account the table of species synonyms; note: when performing search including known (known by the system) species synonyms one should provide both genus and species_epithet values, if only one of these is provided or both are leaved empty, a warning will be shown and this search condition will be ignored;
- additionals — boolean parameter, allowed values are false or true; absence of the parameter in GET-request is treated as its false value; true value (e.g. additionals=true) tells the system to search within additionals species (if such exist); some herbarium records include more than one species, e.g. bryophyte records;
- id — record's ID ; if this parameter is provided in GET-request, all the other search parameters are ignored and the only one record (if it exists and is published) with the requested ID is returned;
- fieldid — field number;
- itemcode — storage number (used in the herbarium storage);
- authorship — authorship of the main species (case insensetive, records should have the provided value as a substring in the correspongin field of the database);

Note The system performs only one-way transliteration of fields collectedby and identifiedby to Englsh language. So, if you will try to search, e.g. collectedby=Боб (that coincide to Bob in English), the system will find records include either Боб or Bob substrings. On the contrary, If you wil try to send collectedby=Bob search query, only records include Bob will be found (still in case insensetive way). Be careful: transliteration

is fully automatic and could be quite silly, e.g. Джон will be transliterated into something like Dzhon, instead of John, as it would expected.

1.3 Description of server response

The server response is a JSON-formatted text transferred via HTTP-protocol and having the following attributes:

- errors — array of errors (strings) occurred during search request evaluation;
- warnings — array of warnings (strings) occurred during search request evaluation; note: warnings are informative messages used to tell the user whats happened in an unexpected way: e.g. which search parameters contradict each other, which parameters were ignored, which parameters weren't recognized by the system etc.
- data — array of structured data, i.e. result of the search query.

1.3.1 Structured data format

data attribute is a json-formatted array. Each item of the array describes a herbarium record and have the following attributes:

- family — family name (latin uppercase letters);
- family_authorship — family authorship;
- genus — genus name;
- genus_authorship — genus authorship;
- species_epithet — species epithet;
- species_id — ID of species instance (unique integer value); don't mix with ID of the herbarium record. ID of the herbarium record is unique for all herbarium records, ID of the related to the herbarium record species instance is unique among all species instances;
- species_authorship — species authorship;
- species_status — current species status; the term species status is related to species instance not herbarium record; it describes a degree of acceptance the species by scientific community (current state); Possible values of species_status are 'Recently added' — the species was recently included to the database and wasn't checked by an expert, 'Approved' — the species was approved by an expert (a user having some prevelegies), 'Deleted' — the species name is probably obsolete and should be avoided, 'From plantlist' — the species was imported from the http://theplantlist.org;
- species_fullname — full species name, e.g. Genus + species epithet + species authorship;
- id — integer identifier of a herbarium record, it is unique;
- gpsbased — boolean parameter, its true value means that a herbarium record position is obtained via the GNSS (GPS/GLONASS); note (for VBGI Herbarium): unfortunately, its false value doesn't meant anything: there are lots of records with geographic coordinates obtained via GNSS, but having unchecked gpsbased flag;
- latitude — latitude, degrees (WGS84);
- longitude — longitude, degrees (WGS84);
- fieldid — field number; an arbitrary string assigned by a collector;

- itemcode — storage number, a string assigned by curator of the herbarium; it is used to identify the position of a record in herbarium storage house;
- acronym — herbarium acronym (e.g. VBGI);
- branch — herbarium branch (e.g. “Herbarium of Fungi”, “Bryophite Herbarium” etc.);
- collectors — collectors;
- identifiers — identifiers;
- devstage — development stage; available values: Development stage partly, Life form of empty string;
- updated — the date the record was saved/updated;
- created — the date the record was created;
- identification_started — the date a species identification was started;
- identification_finished — the date a species identification was finished;
- country — country;
- country_id — unique id of the country;
- altitude — altitude (sea surface is zero-level), this parameter is a string, therefore its form of altitude’s representation might be quite fuzzy: ‘100-300’, ‘100-300 m’, ‘100’, ‘100 m’ etc.
- region — region of collection;
- district — district of collection;
- details — environmental conditions of collection, additional info;
- note — everything that wasn’t yet included in the previous fields (this field could include information on place of collection, environmental conditions etc.);
- dethistory — an array; history of species identifications for this herbarium record;
- additionals — some herbarium records could include more than one species, this array describes all of these;
- images — list of images related to the herbarium record; the list is formatted as follows: [] – an empty list, means that no images are provided; [(‘http://someresource.com/path/to/image1’, ‘image1 type’, ‘meta information1’), (‘http://someresource.com/path/to/image2’, ‘image2 type’, ‘meta information2’), . . .]
  - http://. . . – first field of image record; it is a path (link), where the image coulde be downloaded from;
  - image type – allowed values are eiter ‘p’ or ‘s’; ‘p’ = ‘place’ – the image is related to the place of collection (e.g. s
    ‘s’ = ‘sheet’ – snapshot of the herbarium sheet;
  - meta information – json-formatted string included auxiliary information about the image; e.g. snapshot authorship, snapshot date, etc. In case of snapshot authorship, sample meta-string would be “{‘photographer’: ‘Pavel Krestov’, ‘organization’: ‘Vladivostok Botanical Garden Institute’}” There is no restriction about names of meta-fields, such as ‘photographer’ or ‘organization’; meta-fields could be arbitrary, but ones having intuitive values are preffered.

Note: Attributes region, district, details, note, altitude could be filled in bilingual mode, that means it could include special symbol “|”. For instance, let’s consider region and its value “Russian Far East|Дальний Восток России”. The region stringconsist of two parts English and Russian. In current implementation the API-system doesn’t care about what part of the string is really needed to the user and returns the entire string. Handling such cases, e.g. removing unnecessary substrings from left or right side of the “|” symbol, should be performed by the end user.

Structure of dethistory and additionals arrays are described below.

History of species identifications and additional species

History of species identifications

Each item of the array "History of species identifications" (dethistory) describes an attempt of speciment reidentification in the current herbarium record/sheet and have the following fields:

- valid_from — beginning date of validity of identification;
- valid_to — ending date of validity of identification; empty field means that species assignment to the herbarium record is actual since valid_from date;
- family — family name;
- family_authorship — family authorship;
- genus — genus name;
- genus_authorship — genus authorship;
- species_epithet — species epithet;
- species_id — ID of species instance;
- species_authorship — species authorship;
- species_status — species instance status;
- species_fullname — full species name;

Dates of validity valid_from and valid_to allow to descirbe species reidentificationsin the future, storing in the database species identification history.

Note If herbarium record/sheet include more than one species, than "history of species identifications" is related to main species of the record only.

Additional species

Each item of the array "Additional species" (additionals) describes all species attached to the current herbarium record/sheed and have the following fields (fields have almost the same meaning as for dethistory array):

- valid_from — beginning date of validity of identification;
- valid_to — ending date of validity of identification; empty field means that species assignment to the herbarium record is actual since valid_from date;
- family — family name;
- family_authorship — family authorship;
- genus — genus name;
- genus_authorship — genus authorship;
- species_epithet — species epithet;
- species_id — ID of species instance;
- species_authorship — species authorship;
- species_status — species instance status;
- species_fullname — full species name;

Example

Let us consider an example of additionals array of the following form (not all fields are shown for short):

```
[
{'genus': 'Quercus', 'species_epithet': 'mongolica', ... ,'valid_from': '2015-05-05', 'valid_to': '2016-01-01
↪'},
{'genus': 'Quercus', 'species_epithet': 'dentata', ... ,'valid_from': '2016-01-01', 'valid_to': ''},
{'genus': 'Betula', 'species_epithet': 'manshurica', ... ,'valid_from': '2015-05-05', 'valid_to': ''},
{'genus': 'Betula', 'species_epithet': 'davurica', ... ,'valid_from': '2015-05-05', 'valid_to': ''},
]
```

Inetpretation:

So, if today is 2015, 1 Sept, than the array includes Quercus mongolica, Betula manshurica and Betula davurica, but Quercus dentata should be treated as out-of-date for this date.

If today is 2017, e.g. 1 Jan 2017, than out-of-date status should be assigned to Quercus mongolica, and, therefore, actual set of species includes Quercus dentata, Betula manshurica и Betula davurica.

Note: The array "Additional species" should include complimentary (additional) species to the main species of the herbarium sheet/record only; the main species should never be duplicated in the additional species array.

1.4 Service restrictions

Due to each HTTP-request to the service could lead to transferring big amount of data, there are some restrictions on creating such long running keep-alive HTTP-connections.

The number of allowed simultaneous connections to the service is determined by JSON_API_SIMULTANEOUS_CONN value.

When the number of simultaneous connections is exceeded, the server don't evaluate search requests, but an error message is returned.

This behaviour isn't related with the search-by-id queries. This type of query is evaluated quickly and have no special restrictions.

Unpublished records are ignored when do searching.

Attempt to get data for unpublished record by its ID leads to an error message.

1.5 Examples

To get tested with the service, just build an search request using your web-browser (follow the links below):

http://botsad.ru/hitem/json/?genus=riccardia&collectedby=bakalin

Follow through the link will lead to json-response that includes all known (and published) herbarium records with genus Riccardia and collected by bakalin.

Searching by ID (colstart will be ignored):

http://botsad.ru/hitem/json?id=500&colstart=2016-01-01

http://botsad.ru/hitem/json?id=44

http://botsad.ru/hitem/json?id=5

1.6 See also

Accessing Digital Herbarium using Python

Accessing Digital Herbarium using R