

NOSIBLE Free Search - Data Dictionary V2.1.0

As of: 2026-07-20

This document describes every input and output field of the NOSIBLE Free Search endpoint so that a technical reader can understand the API at a glance. Free Search ([POST https://www.nosible.ai/search/v2/free-search](https://www.nosible.ai/search/v2/free-search)) is 100% free, permissionless web search: no API key is required. It synchronously returns up to 10 search results, and accepts the same inputs as Fast Search with tighter caps ([n_results](#) and [n_probes](#) are capped at 10, [n_contextify](#) at 128 - the server clamps anything higher). Results come back as a JSON object of the shape `{"response": [...]}`.

Interactive API documentation: this endpoint is documented live in the NOSIBLE Swagger UI at <https://www.nosible.ai/search/v2/docs/> (OpenAPI spec: <https://www.nosible.ai/search/v2/openapi.json>). The version of this data dictionary tracks the API version published there: this document describes NOSIBLE V2.1.0.

Contents

Request Headers.....	2
Input Fields.....	2
Output Fields.....	8
Example Request and Response.....	10

Request Headers

Free Search replaces API-key authentication with a single identification header. You should also send `Content-Type: application/json`; `Accept-Encoding: gzip` is recommended.

Header	Details
<code>ISolemnlySwear ImUpToNoGood</code>	Definition: Free Search needs no API key. To prevent abuse, you must instead identify your bot with this header in the format <code>BotName abuse@email.com description</code> where the description is roughly 30 words (10–60 words are accepted). Type: string Required: Yes - a missing or malformed header fails with HTTP 400. Example: <code>TestBot hello@example.com A simple python script testing out the free search API endpoint to fetch data from the web.</code>

Input Fields

All input fields are passed in the JSON body of the POST request. Every field is optional: any field you omit falls back to the default shown in its row. In practice, `question` is the only field you must think about; everything else tunes recall, filtering, and snippet length.

Input Field	Details
<code>question</code>	Definition: The question you would like your search results to answer. This is the primary query that drives retrieval. Type: string Length: 1–500 characters Default: <code>"Why did Argus Filch confiscate the Marauder's Map?"</code> Example: <code>"What conservation efforts exist for the Queen Alexandra's Birdwing butterfly?"</code>
<code>expansions</code>	Definition: Semantically equivalent but lexically diverse rephrasings of the question. Optional and recommended: including ten expansions materially improves recall. Type: array of strings Items: 0–10 Default: <code>[]</code> Example:

Input Field	Details
	<pre>["How is the Queen Alexandra's Birdwing protected in PNG?", "What threats face the largest butterfly species?"]</pre>
instruction	Definition: The instruction passed along to the embedding model. USE WITH CAUTION. Type: string Length: 1–500 characters Default: "Retrieve semantically similar text." Example: "Retrieve semantically similar text."
sql_filter	Definition: A SQL filter to narrow down your search results. Type: string Default: "SELECT loc FROM engine" Columns: netloc - include or exclude specific websites from your search results published - a YYYY-MM-DD string for filtering on publish dates modified - a YYYY-MM-DD string for filtering on modified dates visited - a YYYY-MM-DD string for filtering on visited/crawled dates language - e.g. en-en (english document, english snippet) stot - include or exclude documents based on total number of snippets snum - include or exclude snippets based on their index on the page words - include or exclude snippets based on their word count chars - include or exclude snippets based on their character count Example: "SELECT loc FROM engine WHERE words>=30"
algorithm	Definition: The algorithm to use when scoring search results. Type: string Default: "hybrid-3" - the recommended choice Options: string, lexical, baseline, hamming, hybrid-1, hybrid-2, hybrid-3, company Note: company is remapped to hybrid-3 by the engine. Example: "hybrid-3"
min_similarity	Definition: The minimum similarity that a result must have to be returned. Type: float Range: 0.0–1.0 Default: 0.0

Input Field	Details
	Example: 0.65
must_include	Definition: Any of the following strings must be included in each result. Be careful when using this: aggressive lists can filter out everything. Type: array of strings Items: 0–100 Default: [] Example: <pre>["Ornithoptera alexandrae"]</pre>
must_exclude	Definition: All of the following strings must be excluded from each result. Be careful when using this: aggressive lists can filter out everything. Type: array of strings Items: 0–100 Default: [] Example: <pre>["casino", "betting"]</pre>
brand_safety	Definition: The brand safety level to apply to the source websites. Type: string or null Options: Safe, Sensitive, Unsafe Default: null Example: "Safe"
language	Definition: The language you would like your search results to be refined to. Type: string or null Options: one of 94 ISO 639-1 two-letter codes, e.g. en, fr, de, es, pt, zh, ja, ar Default: null Example: "en"
continent	Definition: The continent you would like your search results to be refined to. Type: string or null

Input Field	Details
	Options: Africa, Asia, Europe, North America, Oceania, South America, Worldwide Default: null Example: "Oceania"
region	Definition: The geographic region you would like your search results to be refined to. Type: string or null Options: one of 31 named world regions, e.g. Western Europe, Middle East, Southeast Asia, Sub-Saharan Africa, The Balkans, The Caucasus Default: null Example: "Southeast Asia"
country	Definition: The country you would like your search results to be refined to. Type: string or null Options: one of 203 full English country names (including Worldwide), e.g. United States, Germany, South Africa Default: null Example: "Papua New Guinea"
sector	Definition: The GICS sector you would like your search results to be refined to. Type: string or null Options: Other, Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Real Estate, Utilities Default: null Example: "Materials"
industry_group	Definition: The GICS industry group you would like your search results to be refined to. Type: string or null Options: one of 26 GICS industry groups, e.g. Banks, Capital Goods, Media & Entertainment, Software & Services Default: null Example: "Food, Beverage & Tobacco"

Input Field	Details
industry	Definition: The GICS industry you would like your search results to be refined to. Type: string or null Options: one of 75 GICS industries, e.g. Biotechnology, Aerospace & Defense, Automobiles Default: null Example: "Biotechnology"
sub_industry	Definition: The GICS sub-industry you would like your search results to be refined to. Type: string or null Options: one of 164 GICS sub-industries, e.g. Advertising, Apparel Retail, Aluminum Default: null Example: "Environmental & Facilities Services"
iab_tier_1	Definition: The tier 1 IAB content category you would like your search results to be refined to. Type: string or null Options: one of 39 tier 1 IAB categories, e.g. Automotive, Business and Finance, Technology & Computing Default: null Example: "Science"
iab_tier_2	Definition: The tier 2 IAB content category you would like your search results to be refined to. Type: string or null Options: one of 388 tier 2 IAB categories Default: null Example: "Environment"
iab_tier_3	Definition: The tier 3 IAB content category you would like your search results to be refined to. Type: string or null Options: one of 644 tier 3 IAB categories Default: null Example: "Environment"
iab_tier_4	Definition: The tier 4 IAB content category you would like your search results to be refined to.

Input Field	Details
	Type: string or null Options: one of 704 tier 4 IAB categories Default: <code>null</code> Example: <code>"Nature"</code>
companies	Definition: A list of company names. This will be used to refine the search results. Type: array of strings Items: 0–3 Default: <code>[]</code> Example: <code>[</code> <code> "Tesla"</code> <code>]</code>
collection	Definition: The collection to search over. Use <code>this-week</code> to search only recently added data. Type: string Options: <code>everything</code> , <code>this-week</code> Default: <code>"everything"</code> Example: <code>"this-week"</code>
deduplicate	Definition: Whether to deduplicate search results based on duplicate headlines. Type: Boolean Default: <code>false</code> Example: <code>true</code>
n_results	Definition: The number of search results you would like back. The free tier is capped at 10. Type: integer Range: 10 (fixed) Default: <code>10</code> Example: <code>10</code>
n_probes	Definition: The number of shards (collections) to probe. The free tier is capped at 10. Type: integer Range: 5–10 Default: <code>10</code> Example: <code>10</code>

Input Field	Details
n_contextify	Definition: The amount of context you would like per search result, in words. The free tier is capped at 128. Type: integer Range: 64–128 Default: 128 Example: 128
internal_use	Definition: Internal-only. Used to enable hidden features and new algorithms from time to time. Hidden from the public API docs. Type: object Default: {} Example: {}

Output Fields

A successful search returns HTTP 200 with a JSON body of the shape `{"response": [...]}`. Each entry in the `response` array is one flat search result object - the same format as Fast Search.

Output Field	Details
response	Definition: The top-level wrapper of the response body. An array of search result objects, with at most 10 entries, sorted by descending similarity. All fields below describe one search result. Type: array of objects Example: <pre>{ "response": [{...}, {...}, ...] }</pre>
url_hash	Definition: A hash that uniquely identifies the page: an 18-byte SHAKE-128 digest of the normalised URL, URL-safe base64 encoded (24 characters). Used as the document key across NOSIBLE systems. Type: string Example: "aB3dE_fG7hIj-KlMnOpQrStU"

Output Field	Details
url	Definition: The full URL of the source page. Type: string Example: "https://www.rainforest-times.com/2025/08/12/birdwing-sanctuary"
netloc	Definition: The network location (domain) of the source. Type: string Example: "www.rainforest-times.com"
published	Definition: The date the page was published, as a YYYY-MM-DD string. Type: string Example: "2025-08-12"
visited	Definition: The date NOSIBLE visited/crawled the page, as a YYYY-MM-DD string. Type: string Example: "2026-06-30"
language	Definition: The language of the document as an ISO 639-1 two-letter code. Type: string Example: "en"
author	Definition: The author of the page, when one could be detected. Type: string Example: "Jane Doe"
title	Definition: The title/headline of the page. Type: string Example: "New sanctuary offers hope for the world's largest butterfly"
description	Definition: The description of the page. Type: string Example: "Conservationists in Papua New Guinea have secured a 10,000-hectare corridor for the Queen Alexandra's Birdwing."
best_chunk	Definition: The single snippet of the page with the highest similarity to your question.

Output Field	Details
	Type: string Example: "The Queen Alexandra's Birdwing, the world's largest butterfly, now has a dedicated sanctuary in Oro Province."
content	Definition: The contextified content: the best snippet expanded with surrounding words from the page, up to <code>n_contextify</code> words. Type: string Example: "... The Queen Alexandra's Birdwing, the world's largest butterfly, now has a dedicated sanctuary in Oro Province. The corridor links two forest reserves ..."
semantics	Definition: Retrieval diagnostics for the result. Type: object Fields: <code>origin_shard</code> (int) - the shard the document was retrieved from <code>chunks_total</code> (int) - total chunks in the document <code>chunks_matched</code> (int) - chunks that matched the query <code>chunks_kept</code> (int) - chunks kept after filtering <code>similarity</code> (float) - the similarity score, between 0 and 1 Example: <pre>{ "origin_shard": 143, "chunks_total": 42, "chunks_matched": 18, "chunks_kept": 6, "similarity": 0.874213 }</pre>

Example Request and Response

The example below runs a free search end to end. The response format is identical to Fast Search, so see that document for a full response example. All field values are illustrative.

Example Request:

```
import requests

response = requests.post(
    url="https://www.nosible.ai/search/v2/free-search",
```

```

json={
    "question": "Conservation efforts for the Queen Alexandra's Birdwing?",
    "algorithm": "hybrid-3",
    "n_results": 10,
    "n_probes": 10,
    "n_contextify": 128,
    "sql_filter": "SELECT loc FROM engine WHERE words>=30"
},
headers={
    "Accept-Encoding": "gzip",
    "Content-Type": "application/json",
    "ISolemnlySwearImUpToNoGood": (
        "TestBot | hello@example.com | A simple python script testing "
        "out the free search API endpoint to fetch data from the web."
    )
}
)

print(response.json())

```

Example Response:

```

{
    "response": [
        {
            "url_hash": "aB3dE_fG7hIj-KlMnOpQrStU",
            "url": "https://www.rainforest-times.com/2025/08/12/birdwing-sanctuary",
            "netloc": "www.rainforest-times.com",
            "published": "2025-08-12",
            "visited": "2026-06-30",
            "language": "en",
            "author": "Jane Doe",
            "title": "New sanctuary offers hope for the world's largest butterfly",
            "description": "Conservationists in Papua New Guinea have secured a
                10,000-hectare corridor for the Queen Alexandra's
Birdwing.",
            "best_chunk": "The Queen Alexandra's Birdwing, the world's largest
                butterfly, now has a dedicated sanctuary in Oro Province.",
            "content": "... The Queen Alexandra's Birdwing, the world's largest
                butterfly, now has a dedicated sanctuary in Oro Province.
                The corridor links two forest reserves ...",
            "semantics": {
                "origin_shard": 143,
                "chunks_total": 42,
                "chunks_matched": 18,
                "chunks_kept": 6,
                "similarity": 0.874213
            }
        },
        ... more results, up to n_results in total ...
    ]
}

```