

NOSIBLE Rich Search - Data Dictionary V2.1.0

As of: 2026-07-20

This document describes every input and output field of the NOSIBLE Rich Search endpoint so that a technical reader can understand the API at a glance. Rich Search ([POST https://www.nosible.ai/search/v2/rich-search](https://www.nosible.ai/search/v2/rich-search)) synchronously returns up to 100 maximally-enriched search results that come with full embeddings, quant signals, content classifications, and website insights. It accepts every Fast Search input plus five enrichment toggles. Requests are authenticated with the [api-key](#) header, and results come back as a JSON object of the shape `{"response": [...]}`, where each entry in the list is one enriched search result. Please be aware: latencies are high when all enrichments are enabled.

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.....	9
Example Request and Response.....	15

Request Headers

Every request to this endpoint is an HTTP POST with a JSON body. Authentication is handled by a single header, documented below. You should also send **Content-Type: application/json**, and **Accept-Encoding: gzip** is recommended to cut transfer sizes.

Header	Details
api-key	Definition: Your NOSIBLE Search API key. To get an API key please visit app.nosible.com , where you can manage organizations, users, subscriptions, and API keys. Type: string Length: 42–50 characters Required: Yes - requests without a valid key fail with HTTP 422. Example: "nos_sk_XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"

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. The five `enrich_*` toggles at the end of the table control which enrichment blocks are attached to each result.

Input Field	Details
question	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: "Why did Argus Filch confiscate the Marauder's Map?" Example: "What conservation efforts exist for the Queen Alexandra's Birdwing butterfly?"
expansions	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: [] 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. Type: integer Range: 10–100 Default: <code>10</code> Example: <code>100</code>
n_probes	Definition: The number of shards (collections) to probe. More probes give deeper coverage at higher latency. Type: integer Range: 5–50 Default: <code>30</code> Example: <code>50</code>

Input Field	Details
n_contextify	Definition: The amount of context you would like per search result, in words. Controls snippet lengths. Type: integer Range: 64–1024 Default: 256 Example: 512
enrich_profile	Definition: Whether to enrich the search results with the profile of the source website. Adds the profile block to each search result. Type: Boolean Default: true Example: false
enrich_targeting	Definition: Whether to enrich the search results with ad targeting information for the source website. Adds the targeting block to each search result. Type: Boolean Default: true Example: false
enrich_history	Definition: Whether to enrich the search results with the publishing track record of the source website. Adds the history block to each search result. Type: Boolean Default: true Example: false
enrich_signals	Definition: Whether to enrich the search results with quantitative signals (sentiment and forward-looking classifications). Adds the signals block to each search result. Type: Boolean Default: true Example: false
enrich_vectors	Definition: Whether to enrich the search results with 1-bit quantized, base64 encoded vectors a.k.a. embeddings. Adds the vectors block to each search result. Type: Boolean Default: true Example: false

Input Field	Details
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 nested search result object made up of the blocks below. The first four blocks (`page`, `snippet`, `tokens`, `semantics`) are always present; the last five appear only when their `enrich_*` toggle is on.

Output Field	Details
response	Definition: The top-level wrapper of the response body. An array of enriched search result objects, with at most <code>n_results</code> entries, sorted by descending similarity. Type: array of objects Example: <pre>{ "response": [{...}, {...}, ...] }</pre>
page	Definition: Details about the source page. Type: object Fields: <code>url</code> (str) - the full URL of the source page <code>title</code> (str) - the title/headline of the page <code>author</code> (str) - the author of the page, when detected <code>published</code> (str) - the publish date, YYYY-MM-DD <code>visited</code> (str) - the date NOSIBLE crawled the page, YYYY-MM-DD <code>description</code> (str) - the description of the page <code>language</code> (str) - ISO 639-1 two-letter code, e.g. en

Output Field	Details
	netloc (str) - the domain of the source verified (bool) - whether the site is verified; currently always false url_hash (str) - 24-character hash that uniquely identifies the page Example: <pre>{ "url": "https://www.rainforest-times.com/birdwing", "title": "New sanctuary offers hope for the world's largest butterfly", "author": "Jane Doe", "published": "2025-08-12", "visited": "2026-06-30", "description": "Conservationists in Papua New Guinea have secured a 10,000-hectare corridor.", "language": "en", "netloc": "www.rainforest-times.com", "verified": false, "url_hash": "aB3dE_fG7hIj-KlMnOpQrStU" }</pre>
snippet	Definition: The matched text of the result. Type: object Fields: best_chunk (str) - the snippet with the highest similarity to your question content (str) - the contextified content: the best snippet expanded with surrounding words, up to n_contextify words Example: <pre>{ "best_chunk": "The Queen Alexandra's Birdwing ...", "content": "... The corridor links two forest reserves ..." }</pre>
tokens	Definition: Tokenized equivalents of the snippet text, useful for highlighting and downstream NLP. Type: object Fields: best_chunk (str) - the tokenized form of snippet.best_chunk content (str) - the tokenized form of snippet.content Example: <pre>{ "best_chunk": "the queen alexandra s birdwing ...", </pre>

Output Field	Details
	<pre>"content": "... the corridor links two forest reserves ..."</pre> <pre>}</pre>
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, rounded to 6 decimal places Example: <pre>{ "origin_shard": 143, "chunks_total": 42, "chunks_matched": 18, "chunks_kept": 6, "similarity": 0.874213 }</pre>
profile	Definition: The profile of the source website. Built from NOSIBLE netloc data plus schema.org structured data found on the site. Only present when <code>enrich_profile</code> is <code>true</code>. Type: object Fields: <code>netloc</code> (str) - the domain this profile describes <code>verified</code> (bool) - whether the site is verified; currently always false <code>legal_name</code> (str/null) - the legal entity behind the website <code>title</code> (str/null) - the website's title <code>description</code> (str/null) - the website's description <code>contact_email</code> (str) - the first contact email found on the homepage, or "Unknown" <code>telephone_num</code> (str) - the first phone number found, or "Unknown" <code>street_address</code> (str) - the street address found, or "Unknown" <code>legal_jurisdiction</code> (str/null) - the site's legal jurisdiction <code>privacy_policy</code> (str/null) - path to the site's privacy policy <code>terms_of_service</code> (str/null) - path to the site's terms of service Example:

Output Field	Details
	<pre>{ "netloc": "www.rainforest-times.com", "verified": false, "legal_name": "Rainforest Times Ltd", "title": "Rainforest Times", "description": "Independent environmental news from the Pacific.", "contact_email": "news@rainforest-times.com", "telephone_num": "Unknown", "street_address": "Unknown", "legal_jurisdiction": null, "privacy_policy": "/legal/privacy", "terms_of_service": "/legal/terms" }</pre>
targeting	Definition: Ad-targeting classifications of the source website. Only present when <code>enrich_targeting</code> is <code>true</code>. Type: object Fields: <code>netloc</code> (str) - the domain these classifications describe <code>verified</code> (bool) - whether the site is verified; currently always false <code>brand_safety</code> (str/null) - Safe, Sensitive, or Unsafe <code>continent</code> (str/null) - the site's continent <code>region</code> (str/null) - the site's geographic region <code>country</code> (str/null) - the site's country <code>language</code> (str/null) - the site's primary language <code>sector</code> (str/null) - the site's GICS sector <code>industry_group</code> (str/null) - the site's GICS industry group <code>industry</code> (str/null) - the site's GICS industry <code>sub_industry</code> (str/null) - the site's GICS sub-industry <code>iab_tier_1</code> (str/null) - the site's tier 1 IAB category <code>iab_tier_2</code> (str/null) - the site's tier 2 IAB category <code>iab_tier_3</code> (str/null) - the site's tier 3 IAB category <code>iab_tier_4</code> (str/null) - the site's tier 4 IAB category <code>ads_txt_url</code> (str/null) - path to the site's ads.txt file Example: <pre>{ "netloc": "www.rainforest-times.com", "verified": false, "brand_safety": "Safe",</pre>

Output Field	Details
	<pre> "continent": "Oceania", "region": "Southeast Asia", "country": "Papua New Guinea", "language": "en", "sector": "Other", "industry_group": "Other", "industry": "Other", "sub_industry": "Other", "iab_tier_1": "Science", "iab_tier_2": "Environment", "iab_tier_3": "Environment", "iab_tier_4": "Nature", "ads_txt_url": null } </pre>
history	Definition: The publishing track record of the source website. Only present when <code>enrich_history</code> is <code>true</code>. Type: object Fields: <code>netloc</code> (str) - the domain this history describes <code>verified</code> (bool) - whether the site is verified; currently always false <code>first_published</code> (str/null) - the earliest publish date we observed for the site <code>first_archived</code> (str/null) - the first time the site was seen by the Wayback Machine <code>first_crawled</code> (str/null) - the first time NOSIBLE crawled the site <code>2000_published_pct</code> (float) - share of the site's pages published as of 2000, rounded to 6 decimals <code>2005_published_pct</code> (float) - as above, for 2005 <code>2010_published_pct</code> (float) - as above, for 2010 <code>2015_published_pct</code> (float) - as above, for 2015 <code>2020_published_pct</code> (float) - as above, for 2020 <code>2025_published_pct</code> (float) - as above, for 2025 Example: <pre> { "netloc": "www.rainforest-times.com", "verified": false, "first_published": "2008-03-14", "first_archived": "2008-06-02", "first_crawled": "2021-11-19", "2000_published_pct": 0.0, </pre>

Output Field	Details
	<pre> "2005_published_pct": 0.0, "2010_published_pct": 0.031882, "2015_published_pct": 0.144209, "2020_published_pct": 0.412355, "2025_published_pct": 0.183012 } </pre>
signals	Definition: Quantitative signals computed over <code>snippet.content</code> by NOSIBLE's classifiers. Only present when <code>enrich_signals</code> is <code>true</code>. Keys are returned in the fixed order below. Type: object Fields: <code>signal_sentiment</code> (str/null) - positive, neutral, or negative; null on classifier failure <code>signal_forward</code> (str/null) - forward or not-forward; null on classifier failure <code>prob_positive</code> (float) - probability the text is positive; defaults to 0.33 on failure <code>prob_neutral</code> (float) - probability the text is neutral; defaults to 0.33 on failure <code>prob_negative</code> (float) - probability the text is negative; defaults to 0.33 on failure <code>prob_forward</code> (float) - probability the text is forward looking; defaults to 0.50 on failure <code>prob_not_forward</code> (float) - probability the text is not forward looking; defaults to 0.50 on failure Example: <pre> { "signal_sentiment": "positive", "signal_forward": "not-forward", "prob_positive": 0.81334, "prob_neutral": 0.12401, "prob_negative": 0.06265, "prob_forward": 0.201152, "prob_not_forward": 0.798848 } </pre>
vectors	Definition: 1-bit quantized embeddings of the result, base64 encoded. Only present when <code>enrich_vectors</code> is <code>true</code>. Type: object Fields:

Output Field	Details
	nosible_bitstring (str/null) - the page's NOSIBLE embedding: a 1536-bit, 1-bit quantized vector packed into 192 bytes and base64 encoded qwen3_8b_bitstring (str/null) - the Qwen3-Embedding-8B embedding of snippet.best_chunk: a 4096-bit, 1-bit quantized vector packed into 512 bytes and base64 encoded; null if the embedding service failed Example: <pre>{ "nosible_bitstring": "kK9mQ2xZ...vB4w==", "qwen3_8b_bitstring": "7Hf2pLnR...Xq1A==" }</pre>

Example Request and Response

The example below runs a fully-enriched search end to end. The request is shown as Python, and the response is shown truncated to a single result. All field values in the response are illustrative.

Example Request:

```
import requests

response = requests.post(
    url="https://www.nosible.ai/search/v2/rich-search",
    json={
        "question": "Conservation efforts for the Queen Alexandra's Birdwing?",
        "enrich_profile": True,
        "enrich_targeting": True,
        "enrich_history": True,
        "enrich_signals": True,
        "enrich_vectors": True,
        "algorithm": "hybrid-3",
        "n_results": 100,
        "n_probes": 30,
        "n_contextify": 256
    },
    headers={
        "Accept-Encoding": "gzip",
        "Content-Type": "application/json",
        "api-key": "YOUR_NOSIBLE_API_KEY"
    }
)

print(response.json())
```

Example Response:

```
{
  "response": [
    {
      "page": {
        "url": "https://www.rainforest-times.com/2025/08/12/birdwing-
sanctuary",
        "title": "New sanctuary offers hope for the world's largest
butterfly",
        "author": "Jane Doe",
        "published": "2025-08-12",
        "visited": "2026-06-30",
        "description": "Conservationists in Papua New Guinea have secured a
10,000-hectare corridor.",
        "language": "en",
        "netloc": "www.rainforest-times.com",
        "verified": false,
        "url_hash": "aB3dE_fG7hIj-KlMnOpQrStU"
      },
      "snippet": {
        "best_chunk": "The Queen Alexandra's Birdwing, the world's largest
Province.",
        "content": "... The corridor links two forest reserves ..."
      },
      "tokens": {
        "best_chunk": "the queen alexandra s birdwing the world s largest
province",
        "content": "... the corridor links two forest reserves ..."
      },
      "semantics": {
        "origin_shard": 143,
        "chunks_total": 42,
        "chunks_matched": 18,
        "chunks_kept": 6,
        "similarity": 0.874213
      },
      "profile": {
        "netloc": "www.rainforest-times.com",
        "verified": false,
        "legal_name": "Rainforest Times Ltd",
        "title": "Rainforest Times",
        "description": "Independent environmental news from the Pacific.",
        "contact_email": "news@rainforest-times.com",
        "telephone_num": "Unknown",
        "street_address": "Unknown",
        "legal_jurisdiction": null,
        "privacy_policy": "/legal/privacy",
        "terms_of_service": "/legal/terms"
      }
    }
  ]
}
```



```

"targeting": {
  "netloc": "www.rainforest-times.com",
  "verified": false,
  "brand_safety": "Safe",
  "continent": "Oceania",
  "region": "Southeast Asia",
  "country": "Papua New Guinea",
  "language": "en",
  "sector": "Other",
  "industry_group": "Other",
  "industry": "Other",
  "sub_industry": "Other",
  "iab_tier_1": "Science",
  "iab_tier_2": "Environment",
  "iab_tier_3": "Environment",
  "iab_tier_4": "Nature",
  "ads_txt_url": null
},
"history": {
  "netloc": "www.rainforest-times.com",
  "verified": false,
  "first_published": "2008-03-14",
  "first_archived": "2008-06-02",
  "first_crawled": "2021-11-19",
  "2000_published_pct": 0.0,
  "2005_published_pct": 0.0,
  "2010_published_pct": 0.031882,
  "2015_published_pct": 0.144209,
  "2020_published_pct": 0.412355,
  "2025_published_pct": 0.183012
},
"signals": {
  "signal_sentiment": "positive",
  "signal_forward": "not-forward",
  "prob_positive": 0.81334,
  "prob_neutral": 0.12401,
  "prob_negative": 0.06265,
  "prob_forward": 0.201152,
  "prob_not_forward": 0.798848
},
"vectors": {
  "nosible_bitstring": "kK9mQ2xZ...vB4w==",
  "qwen3_8b_bitstring": "7Hf2pLnR...Xq1A=="
}
},
... more results, up to n_results in total ...
]
}

```