

Tài liệu API Hán Nôm

Đây là tài liệu mô tả các API mà trang web Kim Hán Nôm hiện có
(<https://tools.clc.hcmus.edu.vn/>)

1. Mô tả cấu trúc chung

1.1 Domain hiện tại của API:

<https://tools.clc.hcmus.edu.vn>

Ví dụ: <https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/image-upload>

1.2 Cấu trúc json chuẩn của response trả về:

```
{  
  "is_success": true,  
  "code": "000000",  
  "message": null,  
  "data": null  
}
```

trong đó:

- **code:** là mã lỗi của response. với mã 000000 là trạng thái thực thi thành công. các mã khác 000000 là mã lỗi.
- **message:** là object thông tin cụ thể về mã. khi mã 000000 thì field này null.
 - **title:** tiêu đề mã lỗi.
 - **text:** nội dung mã lỗi
 - **debug_msg:** chi tiết lỗi (throw error) nếu có. có thể on/off khi chạy ở môi trường thực tế.

data: dữ liệu trả về nếu có. với mã **khác** 000000 thì field này null.

Lưu ý với thư viện requests của python:

Khi xài thư viện requests của python thì cần thêm phần User-Agent. Và phần header “User-Agent” **không** để giá trị có tiền tố là “python-requests”

Ví dụ:

Giá trị đúng: headers = {"User-Agent": "test 123"}

Giá trị sai: headers = {"User-Agent": "python-requests test 123"}

2. Mô tả các API

POST /api/web/clc-sinonom/image-upload

Là API upload hình ảnh lên server

The screenshot displays a REST client interface for a POST request to the endpoint `https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/image-upload`. The request is configured with the method `POST` and the body type set to `form-data`. A single form field is defined with the key `image_file` and the value `hanh_chinh_0001.JPG`. The response section shows a `200 OK` status with a response time of `360 ms` and a body size of `288 B`. The response body is formatted as JSON, indicating a successful upload with the file name `tempd561fba0e59e48dbaaeac2189a9deef8.jpg`.

Key	Value	Description
<input checked="" type="checkbox"/> image_file	File ▾ hanh_chinh_0001.JPG	
Key	Text ▾ Value	Description

```
1 {
2   "is_success": true,
3   "code": "000000",
4   "message": null,
5   "data": {
6     "file_name": "tempd561fba0e59e48dbaaeac2189a9deef8.jpg"
7   }
8 }
```

Thông tin request:

Method: POST, Body: form-data
image_file : [file data]

Thông tin response:

```
{
  "is_success": true,
  "code": "000000",
  "message": null,
  "data": {
    "file_name": " tempd561fba0e59e48dbaaeac2189a9deef8.jpg"
  }
}
```

POST /api/web/clc-sinonom/image-download

Là API tải hình ảnh đã đưa lên server


GET ▼ https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/image-download?file_name=tempd561fba0e59e48dbaaecf... Send ▼

Params ● Authorization Headers (7) Body Scripts Settings Cookies

Query Params

<input checked="" type="checkbox"/>	Key	Value	Description	...	Bulk Edit
<input checked="" type="checkbox"/>	file_name	tempd561fba0e59e48dbaaec2189a9deef8....			
	Key	Value	Description		

ody Cookies Headers (4) Test Results 200 OK • 449 ms • 824.57 KB • Save Response ...



Thông tin request:

Method: GET, Param: file_name

Thông tin response:

[image_data]

POST /api/web/clc-sinonom/image-classification

Là API phân loại hình ảnh

The screenshot shows a REST client interface with a POST request to `https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/image-classification`. The request body is a JSON object with a single key `file_name` pointing to a temporary image file. The response is a 200 OK status with a JSON body indicating success and providing classification details.

Request:

```
1 {
2   "file_name": "tempd561fba0e59e48dbaaeac2189a9deef8.jpg"
3 }
```

Response:

```
1 {
2   "is_success": true,
3   "code": "000000",
4   "message": null,
5   "data": {
6     "ocr_id": 2,
7     "ocr_name": "Hành chính"
8   }
9 }
```

Thông tin request:

Method: POST, Body: Raw-Json

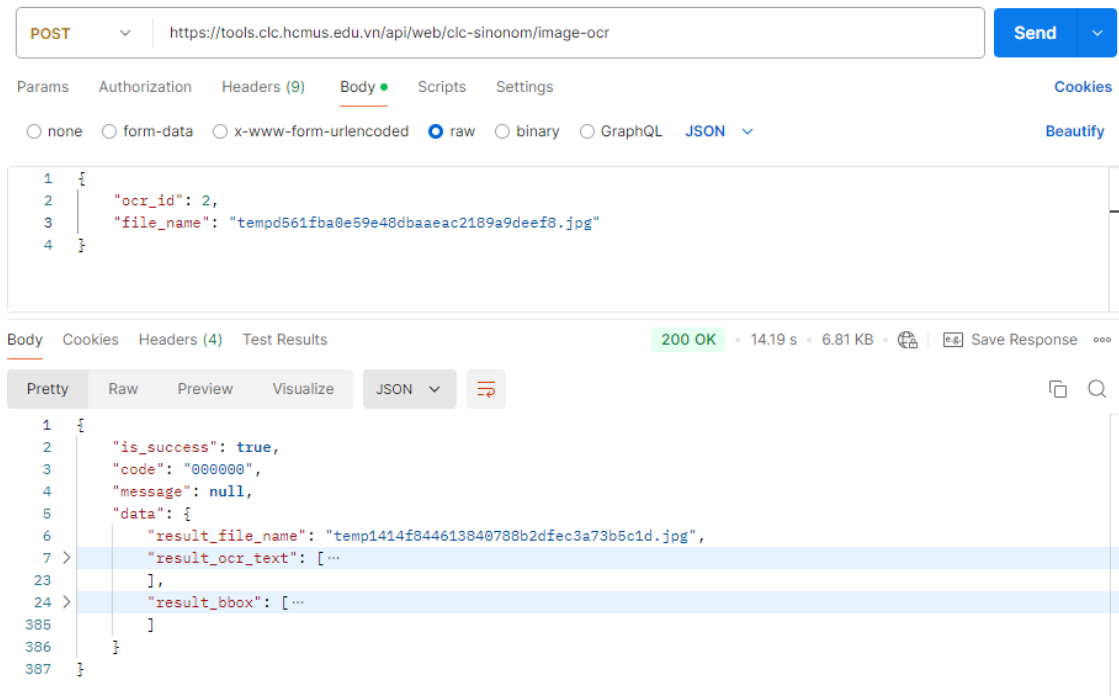
```
{
  "file_name": "tempd561fba0e59e48dbaaeac2189a9deef8.jpg"
}
```

Thông tin response:

```
{
  "is_success": true,
  "code": "000000",
  "message": null,
  "data": {
    "ocr_id": 2,
    "ocr_name": "Hành chính"
  }
}
```

POST /api/web/clc-sinonom/image-ocr

Là API OCR hình ảnh



Thông tin request:

Method: POST, Body: Raw-Json

```
{
  "ocr_id": 2,
  "file_name": " tempd561fba0e59e48dbaaeac2189a9deef8.jpg"
}
```

Thông tin response:

```
{
  "is_success": true,
  "code": "000000",
  "message": null,
  "data": {
    "result_file_name": "
temp1414f844613840788b2dfec3a73b5c1d.jpg",
    "result_ocr_text": [
      "承"
    ],
    "result_bbox": []
  }
}
```

Note: ocr_id là một trong các giá trị:

1: thông thường dọc, 2: hành chính, 3: ngoại cảnh, 4: thông thường ngang

POST /api/web/clc-sinonom/sinonom-transliteration

Là API dịch âm (chuyển tự) Hán-Nôm

The screenshot shows a REST client interface with a POST request to `https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/sinonom-transliteration`. The request body is a JSON object with a `text` field containing the Vietnamese text: `"text": "皇帝制曰朕惟人臣事上百勞靡此于壯懷"`. The response is a 200 OK status with a JSON body containing success information and transliteration results.

```
1 {
2   "text": "皇帝制曰朕惟人臣事上百勞靡此于壯懷"
3 }
```

body Cookies Headers (4) Test Results 200 OK • 422 ms • 475 B | Save Response

```
1 {
2   "is_success": true,
3   "code": "000000",
4   "message": null,
5   "data": {
6     "result_hannom_text": [
7       "皇帝制曰朕惟人臣事上百勞靡此于壯懷"
8     ],
9     "result_text_transcription": [
10      "hoàng đế chế viết trăm duy nhân thần sự thượng bách lao mỹ thử vu tráng hoài "
11    ]
12  }
13 }
```

Thông tin request:

Method: POST, Body: Raw-Json

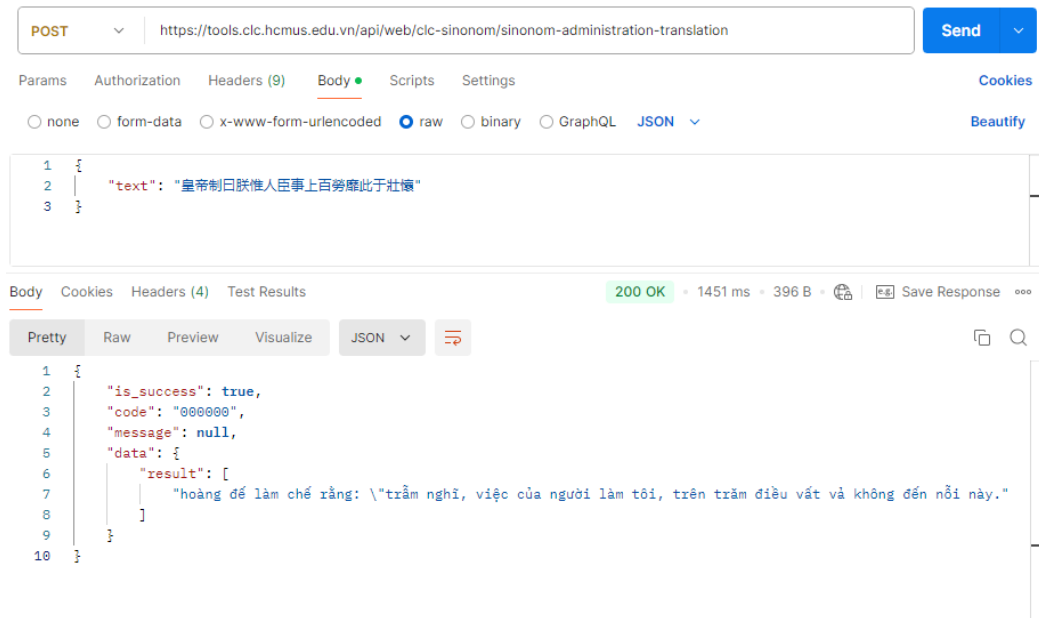
```
{
  "text": "皇帝制曰朕惟人臣事上百勞靡此于壯懷"
}
```

Thông tin response:

```
{
  "is_success": true,
  "code": "000000",
  "message": null,
  "data": {
    "result_hannom_text": [
      "皇帝制曰朕惟人臣事上百勞靡此于壯懷"
    ],
    "result_text_transcription": [
      "hoàng đế chế viết trăm duy nhân thần sự thượng bách
lao mỹ thử vu tráng hoài "
    ]
  }
}
```

POST /api/web/clc-sinonom/sinonom-administration-translation

Là API dịch nghĩa (interpretation) của văn bản hành chính chữ Hán



Thông tin request:

Method: POST, Body: Raw-Json

```
{
  "text": "皇帝制曰朕惟人臣事上百勞靡此于壯懷"
}
```

Thông tin response:

```
{
  "is_success": true,
  "code": "000000",
  "message": null,
  "data": {
    "result": [
      "hoàng đế làm chế rằng: \"trẫm nghĩ, việc của người
làm tôi, trên trăm điều vất vả không đến nỗi này.\"
    ]
  }
}
```

POST /api/web/clc-sinonom/sinonom-verse-translation

Là API dịch nghĩa của văn bản chữ Hán

The screenshot shows a REST client interface with the following details:

- Method:** POST
- URL:** https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/sinonom-verse-translation
- Body (Raw):**

```
1 {
2   "text": "須志在四方張天以火冊而揚"
3 }
```
- Response (JSON):**

```
1 {
2   "is_success": true,
3   "code": "000000",
4   "message": null,
5   "data": {
6     "result": [
7       "tu chí ở bốn phương giương lên trời để nêu cao sách vở"
8     ]
9   }
10 }
```
- Status:** 200 OK, 717 ms, 332 B
- Buttons:** Send, Beautify, Save Response

Thông tin request:

Method: POST, Body: Raw-Json

```
{
  "text": "須志在四方張天以火冊而揚"
}
```

Thông tin response:

```
{
  "is_success": true,
  "code": "000000",
  "message": null,
  "data": {
    "result": [
      "tu chí ở bốn phương giương lên trời để nêu cao sách
vở"
    ]
  }
}
```


POST /api/web/clc-sinonom/sinonom-prose-translation

Là API dịch nghĩa của văn xuôi chữ Hán

The screenshot shows a REST client interface. At the top, the method is set to POST and the URL is `https://tools.clc.hcmus.edu.vn/api/web/clc-sinonom/sinonom-prose-translation`. The 'Body' tab is selected, showing a raw JSON request:

```
{  "text": "須志在四方張天以火冊而揚"}
```

. Below the request, the 'Test Results' section shows a successful response with status '200 OK', a response time of '1189 ms', and a size of '334 B'. The response is displayed in the 'Pretty' JSON format:

```
{  "is_success": true,  "code": "000000",  "message": null,  "data": {    "result": [      "cần chí ở bốn phương, giảng cò trời để đốt lửa sách."    ]  }}
```

Thông tin request:

Method: POST, Body: Raw-Json

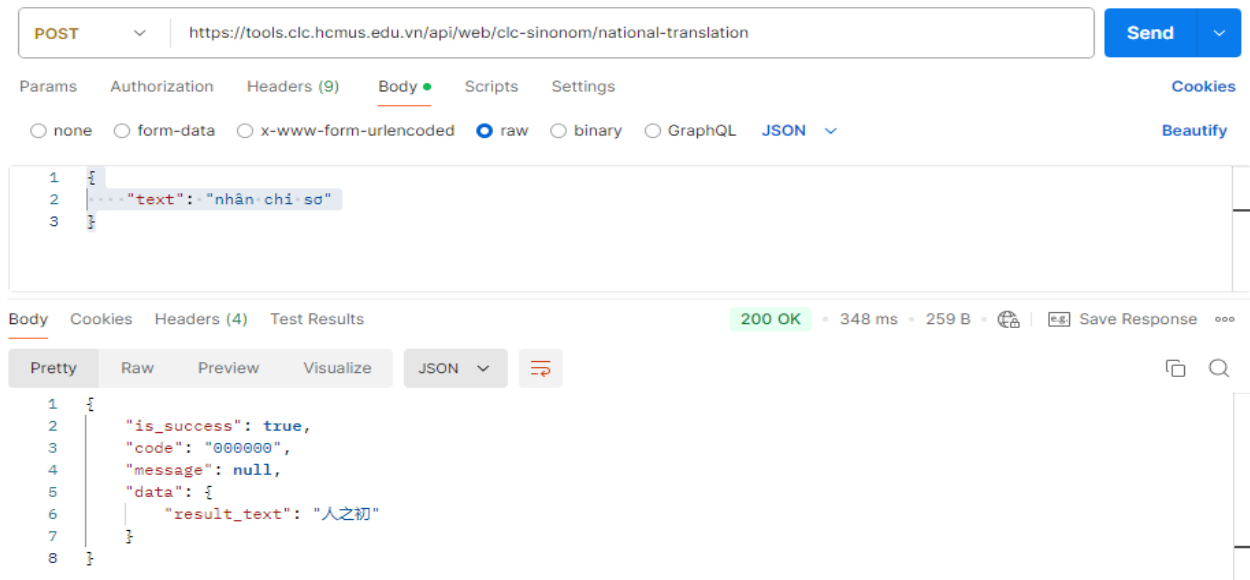
```
{  "text": "須志在四方張天以火冊而揚"}
```

Thông tin response:

```
{  "is_success": true,  "code": "000000",  "message": null,  "data": {    "result": [      "cần chí ở bốn phương, giảng cò trời để đốt lửa sách."    ]  }}
```

POST /api/web/clc-sinonom/national-transliteration

Là API chuyển tự (dịch âm) Quốc Ngữ thành Hán-Nôm



Thông tin request:

Method: POST, Body: Raw-Json

```
{  "text": "nhân chi sơ"}
```

Thông tin response:

```
{  "is_success": true,  "code": "000000",  "message": null,  "data": {    "result_text": "人之初"  }}
```

Tóm lại:

- Mỗi loại ảnh văn bản (thông thường dọc, thông thường ngang, hành chánh và ngoại cảnh) sẽ gọi các API-OCR khác nhau tương ứng với ID của nó.
- Sau khi OCR, phần dịch âm Hán-Nôm sang Quốc Ngữ sẽ gọi chung 1 API chuyển tự SinoNom-Transliteration; phần dịch âm từ QN sang HánNôm sẽ gọi API chuyển tự National-Transliteration.
- Phần dịch nghĩa sẽ được gọi cho các văn bản toàn chữ Hán mà thôi. Với văn bản thông thường có thể có loại Hán Nôm (vd: Truyện Kiều) và toàn Hán (vd: Đại Việt Sử ký toàn thư), còn các loại còn lại (hành chánh, ngoại cảnh) thì toàn Hán.