

Python requests and LanguageTool encoding error

Asked 1 year, 8 months ago Modified 1 year, 8 months ago Viewed 351 times



I am trying to post text data to a [languagetool](#) server. My text includes trademark symbols and copyright symbols etc.

-1



On my first attempt to just post the text like so:

```
response = requests.post(
    LANGUAGE TOOL_URL,
    data=f"language=en-US&text={text}"
)
```



I received an error from requests:

```
UnicodeEncodeError: 'latin-1' codec can't encode character '\u2122' in position
317: Body ('™') is not valid Latin-1. Use body.encode('utf-8') if you want to send
it encoded in UTF-8.
```

Following [this post](#) I updated my request as follows:

```
response = requests.post(
    LANGUAGE TOOL_URL,
    data=f"language=en-US&text={text}".encode('utf-8')
)
```

Now requests does not error but the languagetool server complains that it cannot decode the query:

```
2022-01-23 13:09:47.366 +0000 INFO [lt-server-thread-6] [logError] rID:-
org.languagetool.server.LanguageToolHttpHandler An error has occurred: 'Could not
decode query. Query length: 3085 Request method: POST', sending HTTP code 400.
Access from 172.17.0.1, HTTP user agent: python-requests/2.27.1, User agent param:
null, Referrer: null, language: null, h: 1, r: 29, time: 0m: ALL, l: DEFAULT,
Stacktrace follows:org.languagetool.server.BadRequestException: Could not decode
query. Query length: 3085 Request method: POST
    at
    org.languagetool.server.LanguageToolHttpHandler.getParameterMap(LanguageToolHttpHand
    at
    org.languagetool.server.LanguageToolHttpHandler.parseQuery(LanguageToolHttpHandler.j
    at
    org.languagetool.server.LanguageToolHttpHandler.getRequestQuery(LanguageToolHttpHand
    at
    org.languagetool.server.LanguageToolHttpHandler.handle(LanguageToolHttpHandler.java:
    at jdk.httpserver/com.sun.net.httpserver.Filter$Chain.doFilter(Filter.java:77)
    at jdk.httpserver/sun.net.httpserver.AuthFilter.doFilter(AuthFilter.java:82)
    at jdk.httpserver/com.sun.net.httpserver.Filter$Chain.doFilter(Filter.java:80)
    at
    jdk.httpserver/sun.net.httpserver.ServerImpl$Exchange$LinkHandler.handle(ServerImpl.
    at jdk.httpserver/com.sun.net.httpserver.Filter$Chain.doFilter(Filter.java:77)
    at
    jdk.httpserver/sun.net.httpserver.ServerImpl$Exchange.run(ServerImpl.java:694)
    at
```

```
java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:
    at
java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java
    at java.base/java.lang.Thread.run(Thread.java:829)
```

I've checked all the languagetool docs and cannot find anything about encodings. I don't know at this stage whether the problem is requests, languagetool, or something else I'm doing wrong. Is it possible to post characters like a trademark symbol to languagetool and if so how?

[python](#) [utf-8](#) [python-requests](#) [character-encoding](#) [languagetool](#)

Share Improve this question Follow

asked Jan 23, 2022 at 13:12



Neil

3,040

5

25

48

- 1 I'm not sure I correctly understand the [API docs](#), but it seems that you should send the text as a query parameter, not payload data. In that case, you need to URL-encode it (and provide it to `requests`'s `params` kwarg, not `data`). – [lenz](#) Jan 23, 2022 at 15:59

1 Answer

Sorted by:

Highest score (default)



1

```
import requests
import json
```



```
response = requests.post(
    'https://api.languagetoolplus.com/v2/check',
    data={'text': 'check for misspelling™ © 2022', 'language': 'en-US'})
```



```
print(json.dumps(response.json(), ensure_ascii=False, indent=2))
```



Output:

```
{
  "software": {
    "name": "LanguageTool",
    "version": "5.7-SNAPSHOT",
    "buildDate": "2022-01-18 13:50:09 +0000",
    "apiVersion": 1,
    "premium": true,
    "premiumHint": "You might be missing errors only the Premium version can find. Contact us at support<at>languagetoolplus.com.",
    "status": ""
  },
  "warnings": {
    "incompleteResults": false
  }
}
```

```

},
"language": {
  "name": "English (US)",
  "code": "en-US",
  "detectedLanguage": {
    "name": "English (US)",
    "code": "en-US",
    "confidence": 0.924
  }
},
"matches": [
  {
    "message": "This sentence does not start with an uppercase letter.",
    "shortMessage": "",
    "replacements": [
      {
        "value": "Check"
      }
    ],
    "offset": 0,
    "length": 5,
    "context": {
      "text": "check for misspelling™ © 2022",
      "offset": 0,
      "length": 5
    },
    "sentence": "check for misspelling™ © 2022",
    "type": {
      "typeName": "Other"
    },
    "rule": {
      "id": "UPPERCASE_SENTENCE_START",
      "description": "Checks that a sentence starts with an uppercase letter",
      "issueType": "typographical",
      "category": {
        "id": "CASING",
        "name": "Capitalization"
      },
      "isPremium": false
    },
    "ignoreForIncompleteSentence": true,
    "contextForSureMatch": -1
  },
  {
    "message": "Possible spelling mistake found.",
    "shortMessage": "Spelling mistake",
    "replacements": [
      {
        "value": "misspelling"
      },
      {
        "value": "dispelling"
      },
      {
        "value": "mi spelling"
      }
    ],
    "offset": 10,
    "length": 10,
    "context": {
      "text": "check for misspelling™ © 2022",
      "offset": 10,
      "length": 10
    },
    "sentence": "check for misspelling™ © 2022",
    "type": {

```

```
    "typeName": "Other"
  },
  "rule": {
    "id": "MORFOLOGIK_RULE_EN_US",
    "description": "Possible spelling mistake",
    "issueType": "misspelling",
    "category": {
      "id": "TYP0S",
      "name": "Possible Typo"
    },
    "isPremium": false
  },
  "ignoreForIncompleteSentence": false,
  "contextForSureMatch": 0
}
]
```

Share Improve this answer

edited Jan 23, 2022 at 17:42

answered Jan 23, 2022 at 17:33

Follow



[Mark Tolonen](#)

167k 26 169 251
