

---

記事

[Megumi Kakechi](#) · 2022年7月26日 5m read

## Embedded Python : lxmlを使用しWebスクレイピングした表の結果をグローバルに格納する

[Embedded Python で Excel のデータを IRIS グローバルに格納する方法](#) では Excel データを pandas.DataFrame に取り込んで、それを InterSystems IRIS グローバルに保存する方法をご紹介しました。

今回は「lxml を使用し Web スクレイピングした表(テーブル)のデータを InterSystems IRIS グローバルに格納する」方法をご紹介します。

以下は Windows 上の IRIS でのインストール方法になります。

UNIX

ベースのシステムでは、pip3 コマンドを使用してインストールします。詳細は [ドキュメント](#) をご覧ください。

最初に、irisip コマンドで必要なライブラリをインストールします。

今回は、pandas、lxml の2つのライブラリをインストールします。

```
>cd C:\InterSystems\IRIS\bin
C:\InterSystems\IRIS\bin>irisip install --target C:\InterSystems\IRIS\mgr\python pandas
C:\InterSystems\IRIS\bin>irisip install --target C:\InterSystems\IRIS\mgr\python lxml
```

今回サンプルで使用するのは、Yahoo!ファイナンスでみる 日本郵政(株) の株価時系列データです。

<https://finance.yahoo.co.jp/quote/6178.T/history>

VSCode または IRISスタジオで次のようなクラスを作成します。

メソッドの宣言部に [ Language = Python ] と指定することで、クラスのメソッドに直接 Pythonコードを書くことができます。

```
Class User.PythonTest2 Extends %Persistent
{

ClassMethod fromPythonString(in As %String)
{
    set ^ISJ2($increment(^ISJ2))=$LISTFROMSTRING(in,"")
}

ClassMethod PythonWebScraping(webadd as %String) [ Language = python ]
{
    import iris
    import pandas as pd
    import lxml
    dfs = pd.read_html(webadd)
    for key,row in dfs[0].iterrows():
        moji=",".join(list(map(str, row)))
```

```
        iris.cls(__name__).fromPythonString(moji)}  
}
```

IRISターミナルから正しく実行できるかを確認します。

```
USER>kill ^ISJ2
```

```
USER>do ##class(User.PythonTest2).PythonWebScraping("https://finance.yahoo.co.jp/quote/6178.T/history")
```

```
USER>zwrite ^ISJ2
```

```
^ISJ2=20  
^ISJ2(1)=$lb("2022?7?25?", "963.0", "966.3", "962.1", "965.2", "5200900", "965.2")  
^ISJ2(2)=$lb("2022?7?22?", "965.1", "967.1", "959.0", "965.0", "8665700", "965.0")  
^ISJ2(3)=$lb("2022?7?21?", "973.4", "977.5", "971.5", "972.1", "11539000", "972.1")  
^ISJ2(4)=$lb("2022?7?20?", "971.2", "974.8", "968.2", "974.8", "8864700", "974.8")  
^ISJ2(5)=$lb("2022?7?19?", "966.8", "970.9", "962.4", "966.6", "5714100", "966.6")  
^ISJ2(6)=$lb("2022?7?15?", "970.0", "970.6", "956.6", "962.7", "9064300", "962.7")  
^ISJ2(7)=$lb("2022?7?14?", "975.0", "976.9", "966.5", "968.7", "10921300", "968.7")  
^ISJ2(8)=$lb("2022?7?13?", "984.0", "986.5", "981.8", "982.3", "6361500", "982.3")  
^ISJ2(9)=$lb("2022?7?12?", "981.0", "985.8", "978.8", "981.7", "8801600", "981.7")  
^ISJ2(10)=$lb("2022?7?11?", "979.0", "987.0", "975.0", "981.5", "10416500", "981.5")  
^ISJ2(11)=$lb("2022?7?8?", "971.7", "977.6", "968.3", "973.6", "10220600", "973.6")  
^ISJ2(12)=$lb("2022?7?7?", "979.0", "984.0", "965.1", "969.8", "10874300", "969.8")  
^ISJ2(13)=$lb("2022?7?6?", "975.0", "976.1", "964.1", "970.8", "12283000", "970.8")  
^ISJ2(14)=$lb("2022?7?5?", "985.0", "989.6", "980.2", "988.4", "9458900", "988.4")  
^ISJ2(15)=$lb("2022?7?4?", "975.0", "983.1", "972.3", "983.1", "9618800", "983.1")  
^ISJ2(16)=$lb("2022?7?1?", "974.0", "986.7", "969.2", "971.0", "15510400", "971.0")  
^ISJ2(17)=$lb("2022?6?30?", "957.9", "971.0", "952.1", "969.1", "11908700", "969.1")  
^ISJ2(18)=$lb("2022?6?29?", "965.5", "972.8", "961.3", "962.9", "28123200", "962.9")  
^ISJ2(19)=$lb("2022?6?28?", "951.5", "967.6", "948.3", "967.6", "12559200", "967.6")  
^ISJ2(20)=$lb("2022?6?27?", "960.0", "960.6", "946.8", "952.3", "15059600", "952.3")
```

```
USER>
```

Pythonプログラムのデバッグを行いたい場合は、Python シェルに切り替えて確認することができます。

```
USER>kill ^ISJ2
```

```
USER>do ##class(%SYS.Python).Shell()
```

```
Python 3.9.5 (default, May 31 2022, 12:35:47) [MSC v.1927 64 bit (AMD64)] on win32  
Type quit() or Ctrl-D to exit this shell.
```

```
>>> import pandas as pd  
>>> import lxml  
>>> import iris  
>>> dfs = pd.read_html(f'https://finance.yahoo.co.jp/quote/6178.T/history')  
>>> for key,row in dfs[0].iterrows():  
...     moji=",".join(list(map(str, row)))  
...     iris.cls('User.PythonTest2').fromPythonString(moji)  
...  
>>> exit()
```

```
USER>zw ^ISJ2
```

```
^ISJ2=20  
^ISJ2(1)=$lb("2022?7?25?", "963.0", "966.3", "962.1", "965.2", "5200900", "965.2")  
^ISJ2(2)=$lb("2022?7?22?", "965.1", "967.1", "959.0", "965.0", "8665700", "965.0")
```

```
^ISJ2(3)=$lb("2022?7?21?", "973.4", "977.5", "971.5", "972.1", "11539000", "972.1")
^ISJ2(4)=$lb("2022?7?20?", "971.2", "974.8", "968.2", "974.8", "8864700", "974.8")
^ISJ2(5)=$lb("2022?7?19?", "966.8", "970.9", "962.4", "966.6", "5714100", "966.6")
^ISJ2(6)=$lb("2022?7?15?", "970.0", "970.6", "956.6", "962.7", "9064300", "962.7")
^ISJ2(7)=$lb("2022?7?14?", "975.0", "976.9", "966.5", "968.7", "10921300", "968.7")
^ISJ2(8)=$lb("2022?7?13?", "984.0", "986.5", "981.8", "982.3", "6361500", "982.3")
^ISJ2(9)=$lb("2022?7?12?", "981.0", "985.8", "978.8", "981.7", "8801600", "981.7")
^ISJ2(10)=$lb("2022?7?11?", "979.0", "987.0", "975.0", "981.5", "10416500", "981.5" )
^ISJ2(11)=$lb("2022?7?8?", "971.7", "977.6", "968.3", "973.6", "10220600", "973.6")
^ISJ2(12)=$lb("2022?7?7?", "979.0", "984.0", "965.1", "969.8", "10874300", "969.8")
^ISJ2(13)=$lb("2022?7?6?", "975.0", "976.1", "964.1", "970.8", "12283000", "970.8")
^ISJ2(14)=$lb("2022?7?5?", "985.0", "989.6", "980.2", "988.4", "9458900", "988.4")
^ISJ2(15)=$lb("2022?7?4?", "975.0", "983.1", "972.3", "983.1", "9618800", "983.1")
^ISJ2(16)=$lb("2022?7?1?", "974.0", "986.7", "969.2", "971.0", "15510400", "971.0")
^ISJ2(17)=$lb("2022?6?30?", "957.9", "971.0", "952.1", "969.1", "11908700", "969.1" )
^ISJ2(18)=$lb("2022?6?29?", "965.5", "972.8", "961.3", "962.9", "28123200", "962.9" )
^ISJ2(19)=$lb("2022?6?28?", "951.5", "967.6", "948.3", "967.6", "12559200", "967.6" )
^ISJ2(20)=$lb("2022?6?27?", "960.0", "960.6", "946.8", "952.3", "15059600", "952.3" )
```

USER>

いかがでしょうか？

みなさんもWebスクレイピングでいろいろなデータをIRISに取り込んでみましょう！

今回使用したサンプルはこちらにあります。

<https://github.com/Intersystems-jp/WebScrapingToGlobal>

[#Embedded Python](#) [#ヒントとコツ](#) [#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)

---

ソースURL:<https://jp.community.intersystems.com/post/embedded-python%E3%82%92%E4%BD%BF%E7%94%A8%E3%81%97web%E3%82%B9%E3%82%AF%E3%83%AC%E3%82%A4%E3%83%94%E3%83%B3%E3%82%B0%E3%81%97%E3%81%9F%E8%A1%A8%E3%81%AE%E7%B5%90%E6%9E%9C%E3%82%92%E3%82%B0%E3%83%AD%E3%83%BC%E3%83%90%E3%83%AB%E3%81%AB%E6%A0%BC%E7%B4%8D%E3%81%99%E3%82%8B>