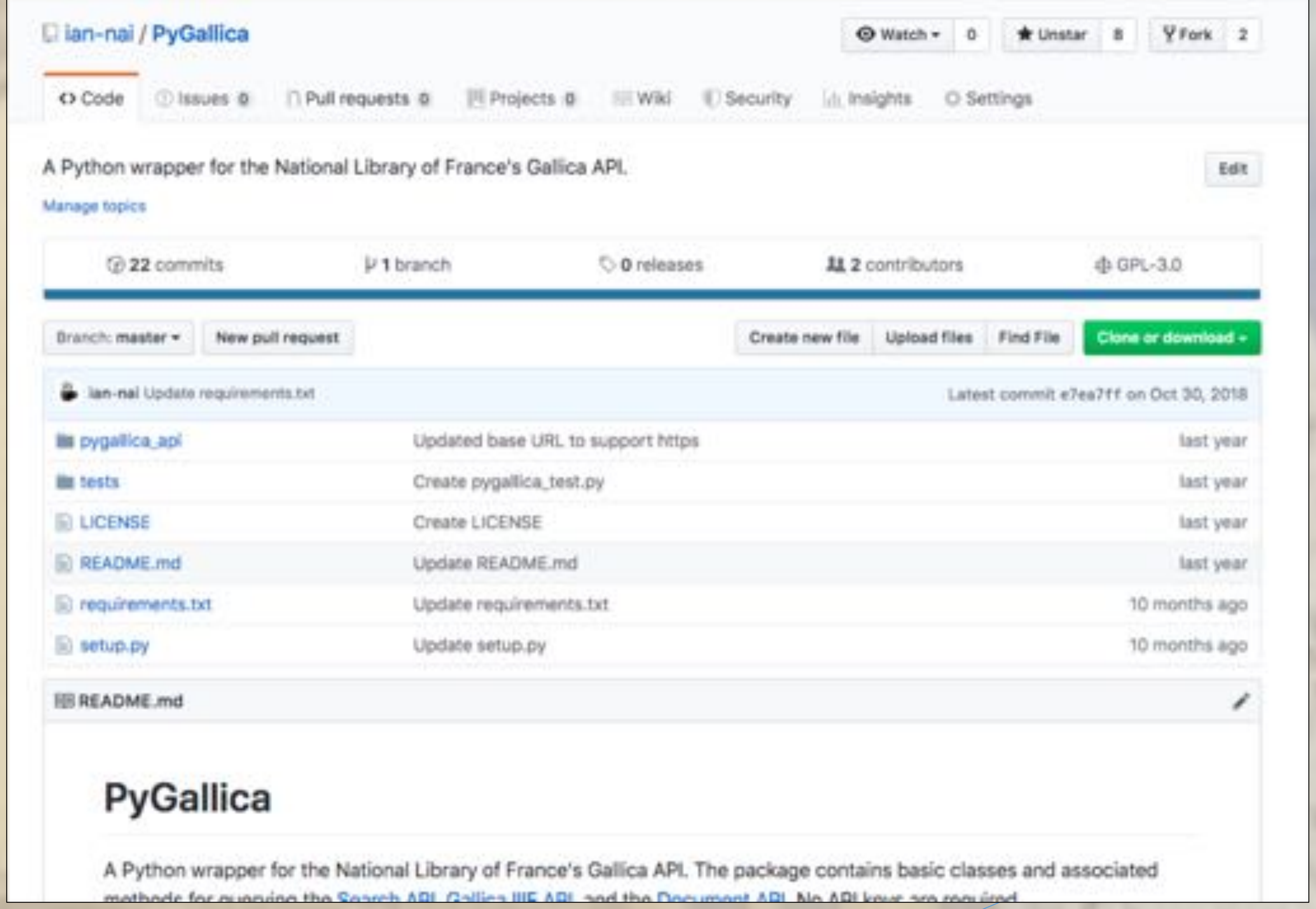
PyGallica: Creating a Python Wrapper for the Gallica Digital Library

By Ian Goodale, UT-Austin Libraries

1. About PyGallica

PyGallica is a Python wrapper for the National Library of France's Gallica API. The package contains basic classes and associated methods for querying the Search API, Gallica IIIF API, and the Document API. No API keys are required.

Gallica is the digital library of the National Library of France and its partners. It has been online since 1997, and now offers access to several million documents, with thousands of new documents added each week. Its API allows extensive searching and downloading of its materials using a number of parameters, which is made easier with this API wrapper.



A Python wrapper for the National Library of France's Gallica API.

ian-nai Update requirements.txt — Latest commit e7ea7ff on Oct 30, 2018

pygallica_api	Updated base URL to support https	last year
tests	Create pygallica_test.py	last year
LICENSE	Create LICENSE	last year
README.md	Update README.md	last year
requirements.txt	Update requirements.txt	10 months ago
setup.py	Update setup.py	10 months ago

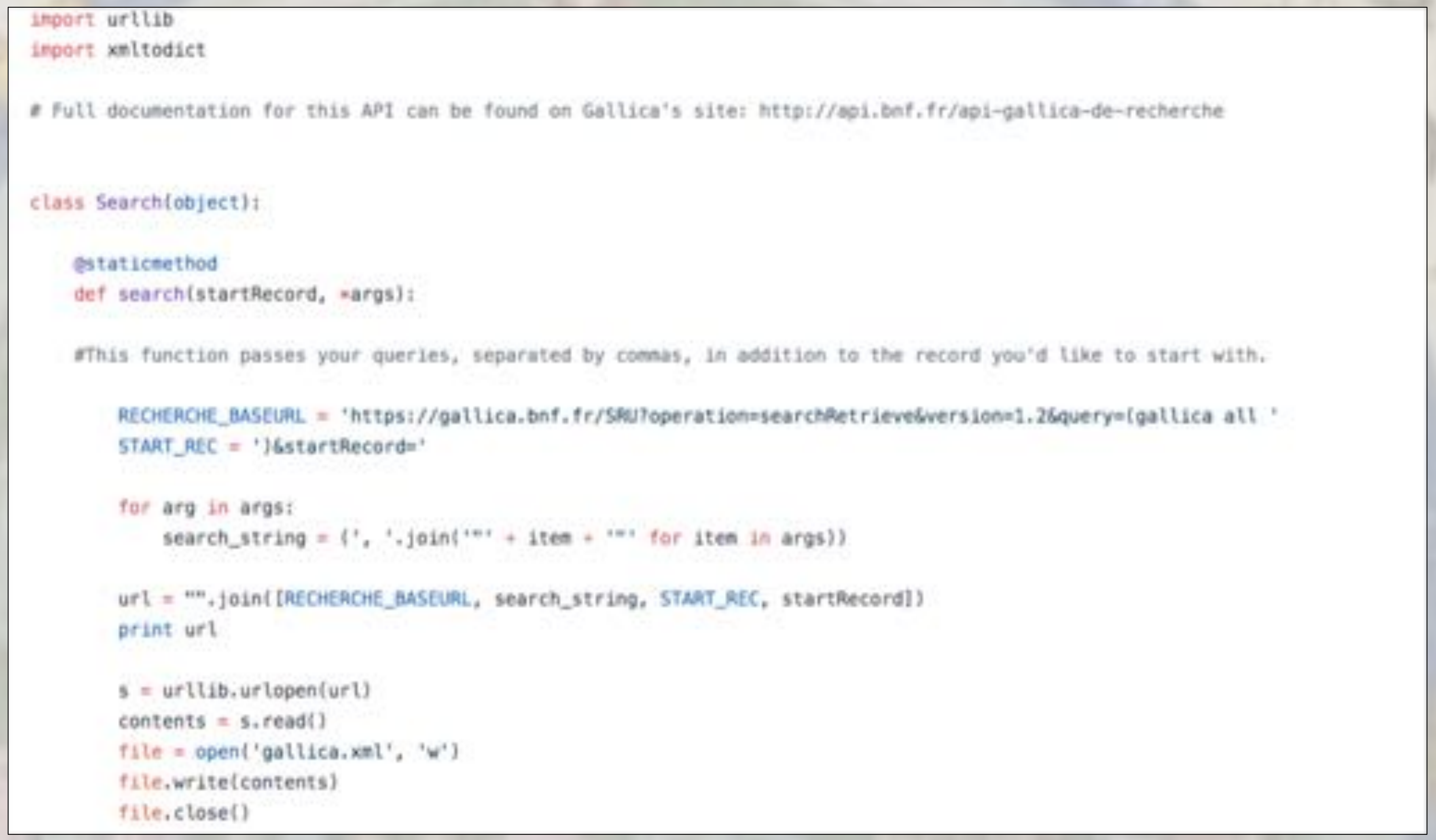
README.md

PyGallica

A Python wrapper for the National Library of France's Gallica API. The package contains basic classes and associated methods for querying the Search API, Gallica IIIF API, and the Document API. No API keys are required.

2. Why Create a Python Wrapper?

Creating a coding project such as a Python wrapper can be a rewarding experience for librarians of varying technical abilities. Learning to write the code allows librarians to deeply engage with others' digital scholarship work, and to broaden one's skillset to more strongly engage with digital humanists and humanities coders across various disciplines and positions. Creating such a project allows one to broaden one's own skillset while creating a tangible product that will be valuable both to the librarian and to others, perhaps on an international scale.

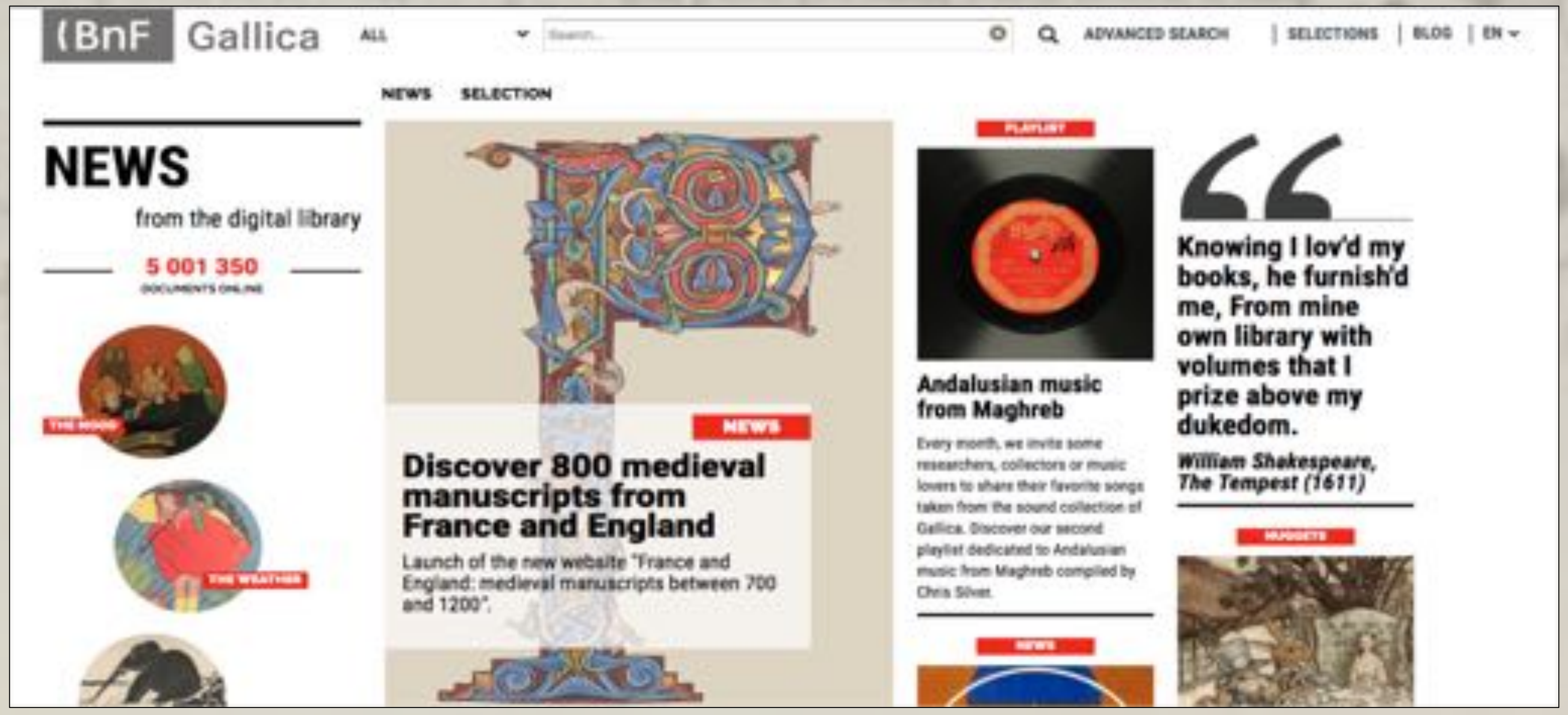


```
import urllib
import xmltodict

# Full documentation for this API can be found on Gallica's site: http://api.bnf.fr/api-gallica-de-recherche


class Search(object):

    @staticmethod
    def search(startRecord, *args):

    #This function passes your queries, separated by commas, in addition to the record you'd like to start with.

        RECHERCHE_BASEURL = 'https://gallica.bnf.fr/SRU?operation=searchRetrieve&version=1.2&query=(gallica all '
        START_REC = ')&startRecord='

        for arg in args:
            search_string = (', '.join('"' + item + '"' for item in args))

        url = "".join([RECHERCHE_BASEURL, search_string, START_REC, startRecord])
        print url

        s = urllib.urlopen(url)
        contents = s.read()
        file = open('gallica.xml', 'w')
        file.write(contents)
        file.close()
```

3. What is a Python Wrapper?

API wrappers can simplify the process of interacting with APIs. An API wrapper provides a way to access an API through a particular programming language or interface, which can help streamline the process of making API calls. Python is commonly used as a programming language for digital humanists constructing their own wrappers.

Librarians can use API wrappers in a number of ways:
- By coding them, creating their own products of digital scholarship that enhance one's own and others' research
- To teach, implementing API instruction into workshops, classes, or consults
- To further their own research, finding resources or conducting analyses of collections



NEWS

from the digital library

5 001 350

DOCUMENTS ONLINE

Discover 800 medieval manuscripts from France and England

Launch of the new website "France and England: medieval manuscripts between 700 and 1200".

Andalusian music from Maghreb

Every month, we invite some researchers, collectors or music lovers to share their favorite songs taken from the sound collection of Gallica. Discover our second playlist dedicated to Andalusian music from Maghreb compiled by Chris Silver.

Knowing I lov'd my books, he furnish'd me, From mine own library with volumes that I prize above my dukedom.

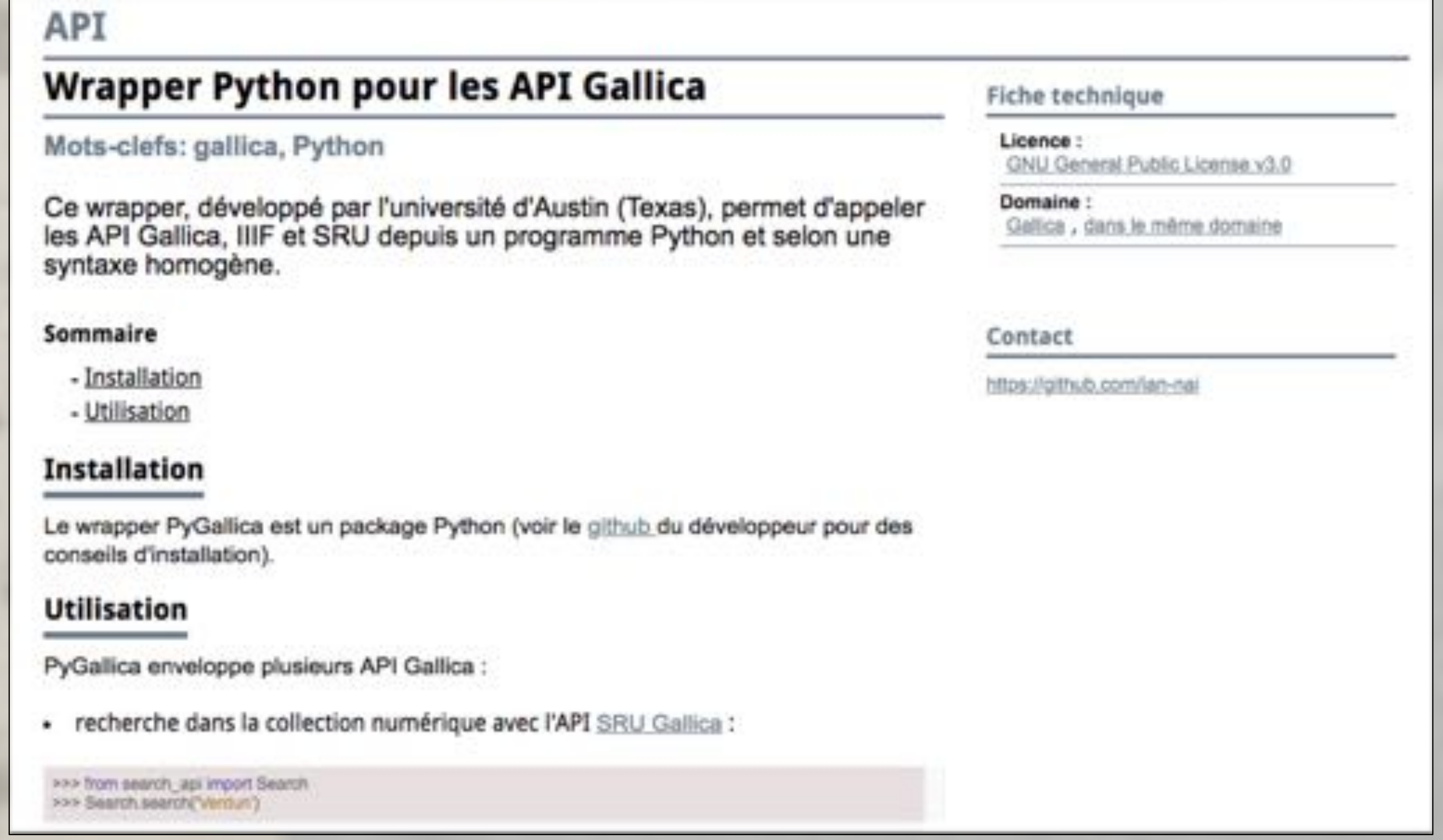
William Shakespeare, The Tempest (1611)

4. Similar Projects and Publishing Venues

Similar projects have been created, such as:
- The Hathi Trust Python wrapper: (https://github.com/rlmv/hathitrust-api)
- The DPLA Python wrapper: (https://github.com/lbjay/dplapy)

These projects, as well as PyGallica, have been made available on GitHub, the leading website for hosting open access code. Other publishing options include GitLab and Bitbucket.

Licensing options vary. I used the GNU General Public License, chosen for its openness to reuse and modification, but other good options exist, such as Creative Commons licenses and the MIT License.



API

Wrapper Python pour les API Gallica

Mots-clefs: gallica, Python

Ce wrapper, développé par l'université d'Austin (Texas), permet d'appeler les API Gallica, IIIF et SRU depuis un programme Python et selon une syntaxe homogène.

Sommaire
- Installation
- Utilisation

Installation

Le wrapper PyGallica est un package Python (voir le github du développeur pour des conseils d'installation).

Utilisation

PyGallica enveloppe plusieurs API Gallica :

- recherche dans la collection numérique avec l'API SRU Gallica :

```
>>> from search_api import Search
>>> Search.search("Verdun")
```

Fiche technique

Licence : GNU General Public License v3.0

Domaine : Gallica , dans le même domaine

Contact

https://github.com/ian-nai

5. The Tool's Future

I initially published the code on my GitHub account, then contacted Gallica to have the tool featured on their website. Working with Gallica staff allowed for the creation of a permanent page on the Gallica website highlighting the tool. Being featured on the Gallica website has greatly enhanced the wrapper's profile, and has led to others working with the tool via GitHub, fostering an ongoing collaboration with other developers.