

The
Automated Journalist

by
Frederik Nurnus



*“A journalist is an individual that
collects/gathers information in form of text, audio, or pictures,
processes them into a news-worthy form, and
disseminates it to the public.”*

Wikipedia

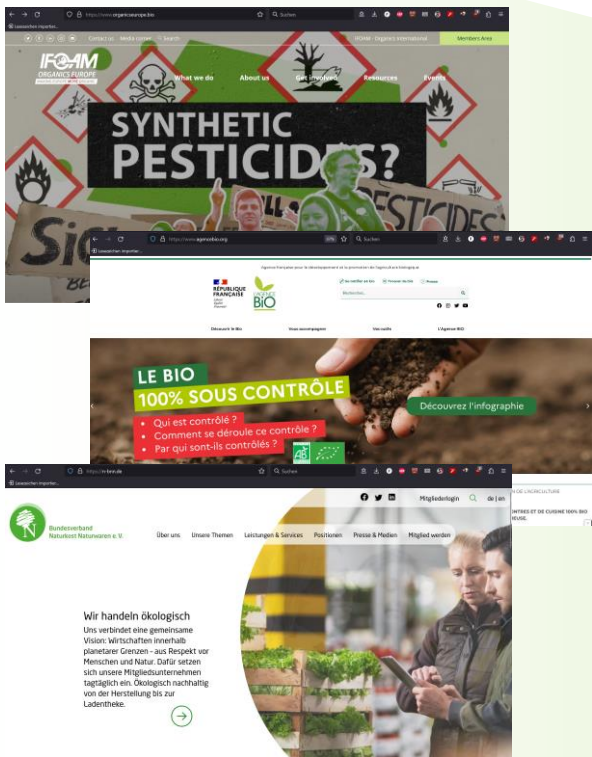


collect information

process information

disseminate information

challenges & learnings



```
Scraping.ipynb (5)

[38]: url15="https://www.organicseurope.bio/library/?qterms=7&qyears=&qtopics="

[39]: response5 = requests.get(url15)
      soup5 = BeautifulSoup(response5.content, "html.parser")
      response5.status_code

[39]: 200

[40]: links5 = []

[41]: #pagewrap > article > div > div.library-list > div:nth-child(1) > div > h2 > a

[42]: for link in soup5.select("#pagewrap > article > div > div.library-list")[0].find_all('a', href=True):
      links5.append(link.get('href'))
      links5 = [ links5 for links in links5 if "qtopics" not in links5 ]
      links5 = list(dict.fromkeys(links5))
      del links5[10:]

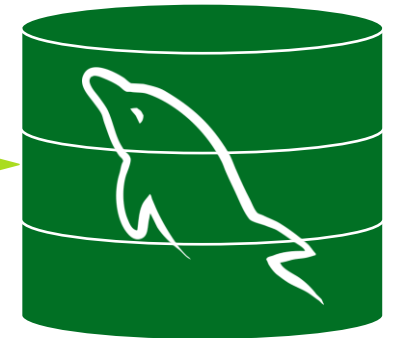
[43]: links5

[43]: ['https://www.organicseurope.bio/library/green-claims-the-fight-against-greenwashing-must-be-based-on-methodologies-promoting-the-transition-towards-more-sustainable-food-systems/',
      'https://www.organicseurope.bio/library/ifoam-organics-europe-diversified-communications-partner-to-support-organic-market-growth/',
      'https://www.organicseurope.bio/library/scientific-study-shows-synthetic-pesticides-significantly-more-dangerous-than-natural-ones/',
      'https://www.organicseurope.bio/library/eu-organic-movement-warns-against-greenwashing/',
      'https://www.organicseurope.bio/library/la-federation-europeenne-de-lagriculture-bio-agit-en-justice-contre-un-affichage-environnemental-trompeur/',
      'https://www.organicseurope.bio/library/eu-organic-movement-takes-legal-action-to-defend-reliable-environmental-labelling-of-food-products/',
      'https://www.organicseurope.bio/library/boosting-organic-farming-aquaculture-to-reach-the-farm-to-fork-targets/',
      'https://www.organicseurope.bio/library/bold-pesticide-reduction-plan-faces-attack-on-saturday/',
      'https://www.organicseurope.bio/library/commission-proposal-on-the-certification-of-carbon-removals-leaves-crucial-questions-unanswered/',
      'https://www.organicseurope.bio/library/green-claims-better-methodologies-to-assess-environmental-performance-of-food-textiles-products/']

[44]: country5 = []
      while len(country5) < len(links5):
          country5.append("EU")

[45]: source5 = []
      while len(source5) < len(links5):
          source5.append("IFOAM EU")

[46]: article5 = []
      for link in links5:
          response = requests.get(link)
          reader = PyPDF2.PdfReader(io.BytesIO(response.content))
          text=reader.pages[0].extract_text()
          article5.append(' '.join(text.split()))
```





collect information

process information

disseminate information

challenges & learnings



Author

Article



SEO Expert

Title



```
File Edit View Run Kernel Tabs Settings Help
Textconverter.ipynb x Scraping.ipynb x WPAutor.ipynb x + Python 3 (pykernel)

Remodel through API

[8]: import os
import openai
openai.organization = "org-7L"
openai.api_key = open("oai.txt", "r").read()
from datetime import date

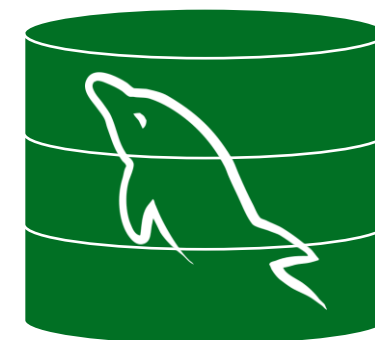
[9]: def convert_article():
    completion = openai.ChatCompletion.create(
        model="gpt-3.5-turbo",
        messages=[{"role": "system",
                    "content": "You are a journalist in the organic foods sector, working for an organic trade journal in europe."},
                  {"role": "user", "content": "Please transform the following Article: \"%orig_article\"}],
        temperature=0.5)
    tran_article = completion.choices[0].message.content
    return tran_article_list.append(tran_article)

Transfer to Database

[10]: def retrieve_idorig_article():
    connection = connect_server()
    cursor = connection.cursor()
    use_database("automated_journalist", connection)
    query = ('SELECT idorig_article FROM orig_article WHERE orig_article_text = %s')
    val = ((orig_article))
    cursor.execute(query, val)
    result = cursor.fetchall()
    cursor.close()
    disconnect_server(connection)
    return result[0][0]

[11]: def insert_tran_article(tran_article):
    connection = connect_server()
    cursor = connection.cursor()
    use_database("automated_journalist", connection)
    add_entry = ("INSERT INTO tran_article (tran_article_text, idorig_article, creation_date) VALUES (%s, %s, %s)")
    data_entry = (tran_article, retrieve_idorig_article(), date.today())
    cursor.execute(add_entry, data_entry)
    connection.commit()
    cursor.close()
    disconnect_server(connection)

[26]: while count_tran_article_entries() < count_orig_article_entries():
    tran_article_list=[]
    orig_article = retrieve_article(count_tran_article_entries()+1)
    convert_article()
    insert_tran_article(tran_article_list[0])
```



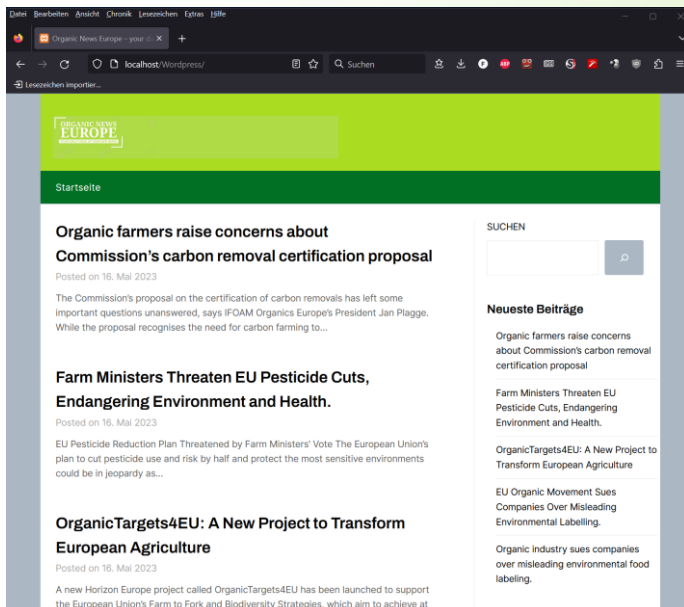


collect information

process information

disseminate information

challenges & learnings



```

File Edit View Run Kernel Tabs Settings Help
Textconverter.ipynb Scraping.ipynb WPAutor.ipynb Python 3 (ipykernel)

[15]: 'Chef Rafa Jiménez Garrido from the School of Healthy and Sustainable Cooking Karuna will be offering a showcooking with organic food on Friday, May 12th at 12:00 pm at the Enogastronomic Laboratory of the University of Malaga. The event is part of the XXX Technical Days of the Spanish Society of Organic Agriculture and Agroecology (SEAE) and is aimed at students of Tourism, Gastronomy, Nutrition and similar degrees. The initiative, called "Communicating Food: Communication Initiatives to Promote the 2030 Agenda," invites future chefs and nutritionists to "care for the climate from the kitchen" and to show the benefits of organic food in achieving the Sustainable Development Goals, particularly goals 12 and 13. Chef Rafa Jiménez Garrido, known for his energy and passion, will showcase the benefits of using organic food for the health of people, the planet and the taste of our dishes. The event is free and can be followed on Instagram: @comunicaquealimenta.'

[16]: def retrieve_tran_title(idtran_article):
      connection = connect_server()
      cursor = connection.cursor()
      use_database("automated_journalist",connection)

      query = ('SELECT title FROM tran_article WHERE idtran_article = %s')
      val = (idtran_article)
      cursor.execute(query,val)
      result = cursor.fetchall()
      cursor.close()
      disconnect_server(connection)
      return result[0][0]

[17]: retrieve_tran_title(1)

[17]: 'Organic Chef Showcases Benefits of Sustainable Cooking at University Event'

[48]: count_tran_article_entries ()

[49]: 42

[50]: def write_wp_article(idtran_article):
      wp_url = "https://localhost/wordpress/wp-json/wp/v2/posts"
      auth = HTTPBasicAuth("f", "1")

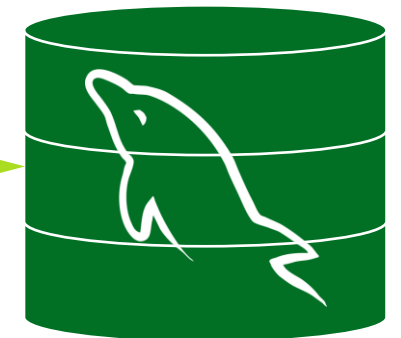
      headers = {
          "Accept": "application/json",
          "Content-Type": "application/json"
      }

      payload = json.dumps({
          "status": "publish",
          "title": retrieve_tran_title(idtran_article),
          "content": retrieve_tran_article(idtran_article)
      })

      response = requests.request(
          "POST",
          wp_url,
          data=payload,
          headers=headers,
          auth=auth,
          verify=False
      )

      print(response)

[51]: for i in range(1,count_tran_article_entries ()):
      write_wp_article(i)
  
```





collecting

- scraping
different formats
- connecting
Python to db

processing

- be precise in your
AI prompts
- crap in, crap out

posting

- take your time
setting up your
environment
- connecting Python
to WordPress