MoroccoAI Challenge 2022

MoroccoAI Annual Conference 2022

December 20-21, 2022

Data Scraping of Africa Calls for Tenders

1. Description

Fintech (Financial Technology) is the term used to augment, streamline, digitize, and disrupt traditional financial services. Some of Fintech use cases include risk assessment, fraud detection, credit scoring, loan default prediction, chatbot assistance, and contracts management. Data is driving Fintech and the weightiest piece can be found in an unstructured format, such as documents, websites, and forums. Financial professionals spend a substantial amount of time reading analysts' reports and financial press. Natural Language Processing (NLP) coupled with Machine Learning (ML) can significantly help decrease the amount of manual routine work.

The business model of many Fintech companies and startups relies on retrieving and aggregating open textual data from companies and governments. This is changing as laws and regulations are updated to support open data in order to enable national companies and startups to join the competition worldwide. Indeed, compilation of this data from disparate sources and making sense of it can play a vital role and impact to any country's economy.

Morocco has been engaged for several years in a dematerialized process to enable freedom of access to public data, an equal treatment of competitors, and guarantees competitors' rights and transparency. Consequently, Morocco emerges as the third largest Fintech hub in the Arab World and among the top key markets in Africa [1, 2].

In this spirit, MoroccoAI is organizing the 2022 Challenge on scraping Africa Calls for Tenders. Tenders are procedures applied to generate offers from companies competing for work and service contracts in the framework of Public Procurement Contracts (PPCs). PPCs play a significant role in contract management that is affecting many sectors in the economy of every country such as construction, public works, energy, telecommunications, etc.

NLP recent advances have completely transformed contract management, allowing businesses to interpret, register, and adjust PPCs in a much shorter period of time with significantly higher accuracy. Thereupon, many countries currently provide dedicated portals to publish and share significant amounts of data about tenders and PPCs. For instance, the portal [3] represents Morocco's counterpart of foreign portals that publish PPCs online, such as the Kenyan Portal [4] or Tunisian Portal [5]. These open portals have allowed a worldwide competition in numerous public offerings, including Moroccan calls for tenders launched by the Moroccan administrations.

2. Main Objectives

The main objectives of this challenge are:

- Encourage data scientists to explore realistic scenarios in the everyday's life of data scientists. Indeed, data gathering, preparation and preprocessing are usually underestimated tasks.
- Understand the regulations behind scraping and gathering data.
- Leverage the outstanding advances in NLP and ML.
- Raise awareness and usefulness of open data.
- Meet the opportunities offered by contract management and understand the opportunities the Fintech industry is offering.

By the organization of this challenge, MoroccoAI aims to facilitate for Moroccan companies the participation in calls for tenders launched worldwide by gathering all public information related to government's PPCs of several countries. We believe this will allow Moroccan companies to increase their competitiveness in key industries. On the other hand, understanding such data will ensure better allocation of economic resources and more rational use of public funds and giving preference to the best-performing national companies. Furthermore, this will reinforce respect for the principles of transparency, equal treatment and efficiency and reduce the risk of fraud and corruption.

3. The Challenge

Candidates are asked to collect and structure textual data of public procurement contracts of the current year (Start_date is in 2022). These could be gathered from public sources from official portals of African Countries like Morocco [3], Kenya [4] and Tunisia [5]. Then the scraped data should be structured according to the following schema (Tender_id, Language, Country, City, Portal_url, Reference_id, Category, Subject, Detail, Keywords, Start_date, End_date, End_time, Award_date, Public_buyer, Supplier, Value, Currency).

For example, from the Morocco portal [3], one can find the tenders in the table below. The scraped data will then be the following: (Tender_id (**1**), Language (**French**), Country (**Maroc**), City (**NOUACEUR**), Portal_url (**https://www.marchespublics.gov.ma/**), Reference_id (**10/2022**), Category (**Fournitures**), Subject (**Achat de Materiels et Mobilier ...**), Detail (**Achat de Materiels et Mobilier de Bureau destiné au centre de santé BARAKA...**), Keywords (**Fourniture, Materiels, Mobilier, Bureau, Santé**), Start_date (**11/11/2022**), End_date (**05/01/2023**), End_time (**11:00**), Award_date (**null**), Public_buyer (**DELEGUE DU MINISTERE DE LA SANTE ...**), Supplier (**null**), Value (**null**), Currency (**MAD**)). The same can be done for the 2nd and 3rd row in the table.

Similarly, for the Kenyan procurements bellow: (Tender_id (**100**), Language (**English**), Country (**Kenya**), City (**KISUMU**), Portal_url (**https://tenders.go.ke/**), Reference_id (**TKNP/2022/23/24**), Category (**Public Colleges**), Subject (**Provision of Catering Services**), Detail (**Provision of Catering Services**), Keywords (**Provision, Catering**), Start_date (**Oct 3 2022**), End_date (**Oct 2 2024**), End_time (**null**), Award_date (**Sept 5 2022**), Public_buyer (**The Kisumu National Polytechnic**), Supplier (**the Alps Hotel...**), Value (**150,000.00**), Currency (**KSH**)).

Annonces de consultation

Nombre de résultats : 1256   Afficher [10 ▼] résultats / page   [1] / 126 ◀◀ ◀ ▶ ▶▶

Procédure Catégorie ▾ Publié le ▾	Référence ▾ Objet ▾ Acheteur public ▾	Lots Lieu d'exécution	Date limite de remise des plis ▾	Actions
AOO ... Fournitures 11/11/2022	**10/2022** **Objet :** Achat de Matériel et Mobilier de Bureau destiné au centre de santé BARAKA relevant de la province NOUACEUR en lot unique. **Acheteur public :** DELEGUE DU MINISTERE DE LA SANTE A LA PROVINCE NOUACEUR	- NOUACEUR	05/01/2023 11:00	
AOO ... Services 02/11/2022	**08/2022/CHPKH** **Objet :** Externalisation des activités d'accueil à l'hôpital provincial de KHENIFRA. **Acheteur public :** CENTRE HOSPITALIER PROVINCIAL DE KHENIFRA	- MAROC KHENIFRA ...	20/12/2022 14:00	
AOO ... Services 02/11/2022	**07/2022/CHPKH** **Objet :** Le brancardage des malades à l'intérieur du Centre Hospitalier Provincial de Khénifra. **Acheteur public :** CENTRE HOSPITALIER PROVINCIAL DE KHENIFRA	- MAROC KHENIFRA ...	20/12/2022 12:00	

https://www.marchespublics.gov.ma/

Supplier	PE	Procurement ...	Contract N...	Value	Description	Award Date	V...
The Alps Hotel Nakuru LTD	THE KISUMU NATIO...	Open Tender	TKNP/2022/...	150,000.00	PROVISION OF CATERING S...	05/09/2022	View
JKUAT ENTERPRISES LIMIT...	Energy Regulatory Co...	Direct Procure...	55,575,600	55,575,600.00	PROVISION OF STAFF RECR...	28/09/2022	View
CFAO MOTORS KENYA LIMI...	Kenya Electricity Gen...	Direct Procure...	PURCHASE ...	25,847.67	QUOTATION FOR MOTOR V...	31/10/2022	View

OCDS (JSON)   Back

Contract Details | Supplier Details

PROVISION OF CATERING SERVICES

PE Name:	THE KISUMU NATIONAL POLYTECHNIC	**PE Category:**	Public Colleges
Tender/RFP No:	TKNP/2022/23/24	**Procurement method:**	Open Tender
Title/Description:	PROVISION OF CATERING SERVICES		
Supplier Name:	The Alps Hotel Nakuru LTD	**PIN:**	P051659065W
Amount:	150000	**Award Date:**	Sep 5, 2022
Start Date:	Oct 3, 2022	**End Date:**	Oct 2, 2024
Expiry Date:	Jan 1, 1970		

https://tenders.go.ke/

Rules

1. Multiple individuals or entities may collaborate as a team. You may not participate on more than one Team. Each Team member must be a single individual,

2. Team membership may not exceed the Maximum of 3 members,

3. Freely & publicly available external data is allowed,

4. Participants must be aware of the terms of services, rules and licenses (for example CC 4.0 [6]) of the portals and respect scraping best practices.

5. Use of open source codes is allowed but should be acknowledged and their license respected,

6. All participants agree that they must share their winning solution's code with the challenge organizers for reproducibility of result and evaluation. The organizers do not claim to own your code and data.

7. MoroccoAI is not responsible for any abuse of the rules, use of illegal scraping methods or use of unauthorized sources.

8. Registered participants will be invited to an information session to explain the challenge in more details.

9. Don't cheat and have fun!

This is a Data Collection Competition. The participants are required to submit a google drive link that contains:

1. A submission csv file **submission.csv** (max size 10Mb). The csv file must follow the template file **template.csv**

2. A zip file **code.zip** that contains the scraping code. It is recommended to join a Jupyter notebook **main.ipynb (**Python 3) that can be used to test the scraping.

3. A **presentation.pdf** file that contains a presentation of the solution of the challenge.

4. Any other document that is considered relevant to the evaluation process.

Evaluation

The Jury of the challenge will evaluate participants work according to the following main criteria:

1. Quality of scrapped data
2. Amount of scrapped data
3. The adopted scraping approach (innovative or use of existing codes, etc...)
4. Presentation of the work
5. Bonus points for providing an additional notebook with some exploratory data analysis

6.   Timeline

- November 25, 2022 - Start Date
- December 02, 2022 - Registration Deadline
- December 15, 2022 - Final Submission Deadline
- December 21, 2022 - Winners Announcement.

All deadlines are at 11:59 PM UTC on the corresponding day unless otherwise noted. The competition organizers reserve the right to update the contest timeline if they deem it necessary.

7.   Prizes

Special prizes will be announced very soon. In addition, Nvidia is offering credits for training courses and certifications by Nvidia Deep Learning Institute (DLI).

- The prizes will be awarded on the basis of the Jury evaluation.
- Due to logistics considerations, the participants need to be residents in Morocco to be eligible for the prize !

Acknowledgements

References

[1] https://fintechnews.ae

[2] https://www.moroccoworldnews.com/

[3] https://www.marchespublics.gov.ma/

[4] https://tenders.go.ke/

[5] http://www.marchespublics.gov.tn/

[6] https://creativecommons.org/licenses/by/4.0/