

Rapport du Projet Ingénierie cognitive



RÉALISÉ PAR :

- Alaoui Mdaghri Ahmed
- Khoulti Wissal

ENCADRÉ PAR :

- Mme. Maryem RHANOU
- Mme. Mounia MIKRAM

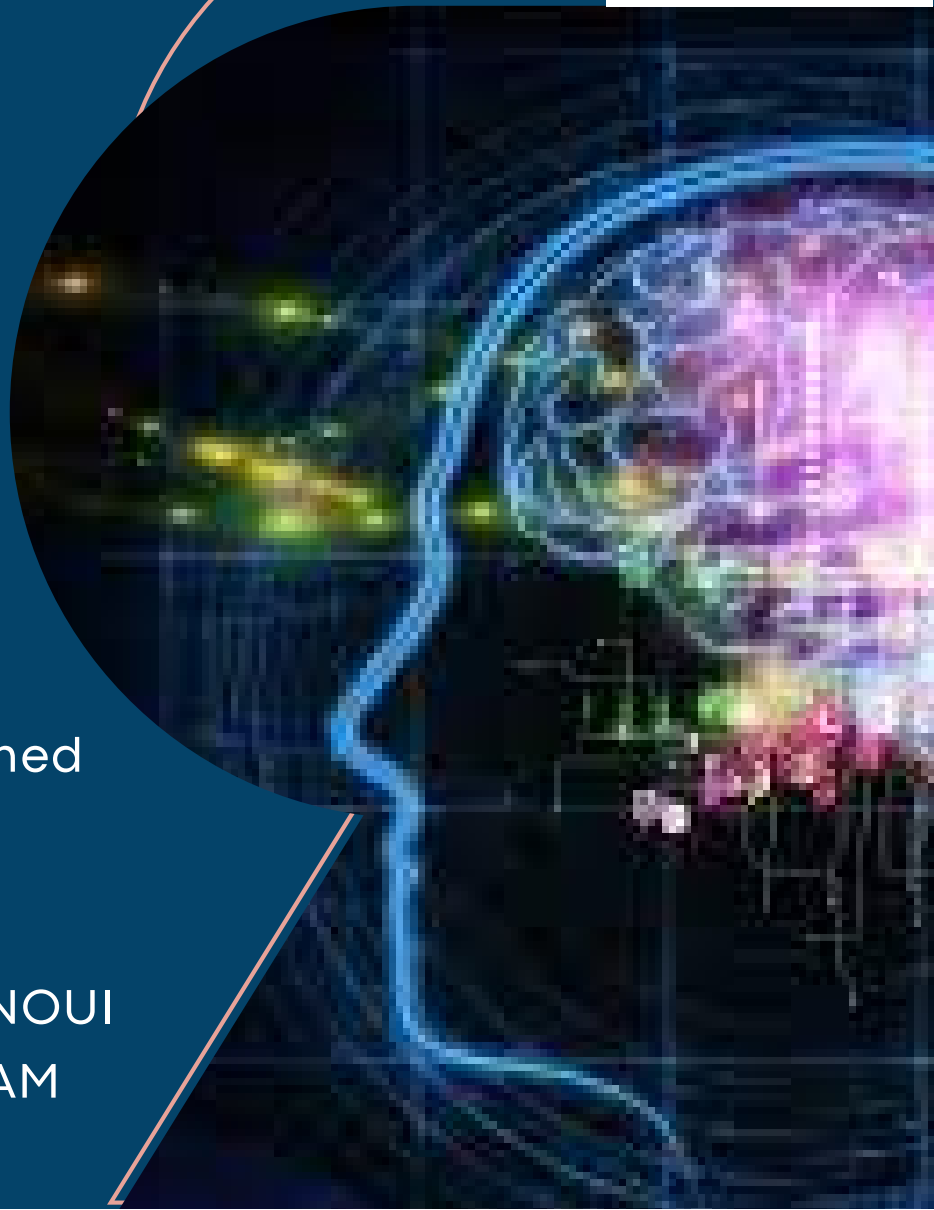


Table des matières

03

Introduction

04

Description du sujet

05

Sources de données

06

L'annotation et son application

10

Conclusion

Introduction

La recherche d'emploi fait face à un changement avec l'apparition de réseaux sociaux professionnels à titre d'exemple LINKEDIN conçu spécifiquement pour une communauté business permettant à l'utilisateur de documenter son propre network de personnes utilisant les compétences, expérience, postes et offres de recrutement.

Description du sujet

Dans notre cas on utilisera LinkedIn et Wikipédia comme source de données pour étudier le marché de travail au Maroc dans un domaine spécifique qui portera sur la science de données, pour cela on annotera les termes spécifiques à ce domaine là avec les différentes compétences requises et cela depuis les données scrapper de Wikipédia comme annotation, dans un autre temps, chaque poste scrapper de LinkedIn nous permettra d'en tirer la description du job avec le but de mieux s'informer et après se former en tenant compte de l'actualité des compétences techniques requises.

Sources de données

Nos données seront scrappées en totalité du web :

-LinkedIn pour alimenter notre base de données relative au marché de travail au Maroc avec l'intitulé du job et sa description.

	Job_title	Company	Location	Link	Description
0	Data Scientist	Société Générale Maroc	Casablanca, Casablanca-Settat, Morocco	https://ma.linkedin.com/jobs/view/data-scienti...	Le Data Scientist a pour mission la conception...
1	Data Scientist-(H/F)	Société Générale Maroc	Casablanca, Casablanca-Settat, Morocco	https://ma.linkedin.com/jobs/view/data-scienti...	Vos missions au quotidienLe Data Scientist a p...
2	AI Machine Learning Optimization Engineer H/F	Stellantis	Casablanca, Casablanca-Settat, Morocco	https://ma.linkedin.com/jobs/view/ai-machine-l...	Stellantis' AI team is expanding and o...
3	Data Scientist	Kenz'up	Casablanca, Casablanca-Settat, Morocco	https://ma.linkedin.com/jobs/view/data-scienti...	About Kenz'upJoin the new moroccan tech startu...

-Wikipédia pour les différentes annotations (langages de programmations, les plateformes, les bases de données et les frameworks).

languages.head()	frameworks_tools	platforms.head()	databases.head()
name	name	name	name
0 A.NET (A#/A sharp)	0 TensorFlow	0 Alibaba Cloud (1 P)	0 Oracle Database
1 A-0 System	1 PyTorch	1 Amazon Web Services (42 P)	1 MySQL
2 A+ (A plus)	2 Spark MLlib	2 Cloud databases (11 P)	2 Microsoft SQL Server
3 ABAP	3 Torch	3 Google Cloud (20 P)	3 PostgreSQL (free software)
4 ABC	4 Transformers	4 IBM cloud services (13 P)	4 IBM Db2

L'annotation et son application

Après extraction des données déjà scraper on peut en construire des patterns et les ajouter au modèle nlp «en_core_web_sm » et l'enregistrer pour utilisation ultérieure.

```
languages = pd.read_csv('languages.csv')
platforms = pd.read_csv('platforms.csv')
databases = pd.read_csv('databases.csv')
frameworks_tools = pd.read_csv('frameworks_tools_etc.csv')

patterns = []
for x in languages.name.tolist():
    patterns.append({"label": "PROG_LANG", "pattern": x, "id": "SKILLS"})

for x in databases.name.tolist():
    patterns.append({"label": "DB", "pattern": [{"lower": w.lower()} for w in str(x).split()], "id": "SKILLS"})

for x in platforms.name.tolist():
    patterns.append({"label": "PLATFORM", "pattern": [{"lower": w.lower()} for w in str(x).split()], "id": "SKILLS"})

for x in frameworks_tools.name.tolist():
    patterns.append({"label": "FRAMEWORKS", "pattern": [{"lower": w.lower()} for w in str(x).split()], "id": "SKILLS"})

nlp = spacy.load("en_core_web_sm")
ruler = nlp.add_pipe("entity_ruler", before="ner")
ruler.add_patterns(patterns)
ruler.to_disk("patterns.jsonl")
```

L'application du modèle préétablie on peut extraire les entités d'une description.

The image shows a resume with various skills and experiences highlighted by colored boxes representing NLP entity extraction. The entities are categorized as follows:

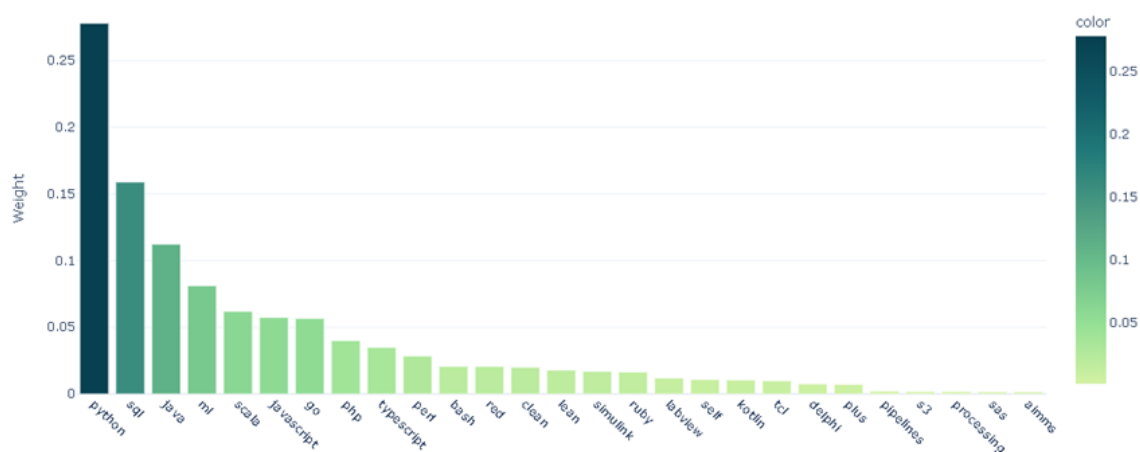
- ORG (Organizations):** Technologies, Python, Machine/Deep Learning.
- PROG_LANG (Programming Languages):** Julia, Java.
- FRAMEWORKS (Frameworks):** PyTorch, TensorFlow, NumPy, Pandas, Keras, Tensorflow.
- DB (Databases):** PostgreSQL, MySQL, Snowflake.
- GPE (Geographical Place Names):** JAX, GCP/AWS.
- DATE (Dates):** 3+ years.
- NORP (Nationalities or Regions):** BA.
- SKILLS (Skills):** experience with applied, ML packages, neural network architectures, Great communication skills, An appreciation for the connection between the product experience and models you build.

Avec une boucle for on pourra par la suite appliquer cela sur l’ensemble des descriptions déjà scrapper de Linkedin et stocker les résultats sous format de DataFrame comme suit :

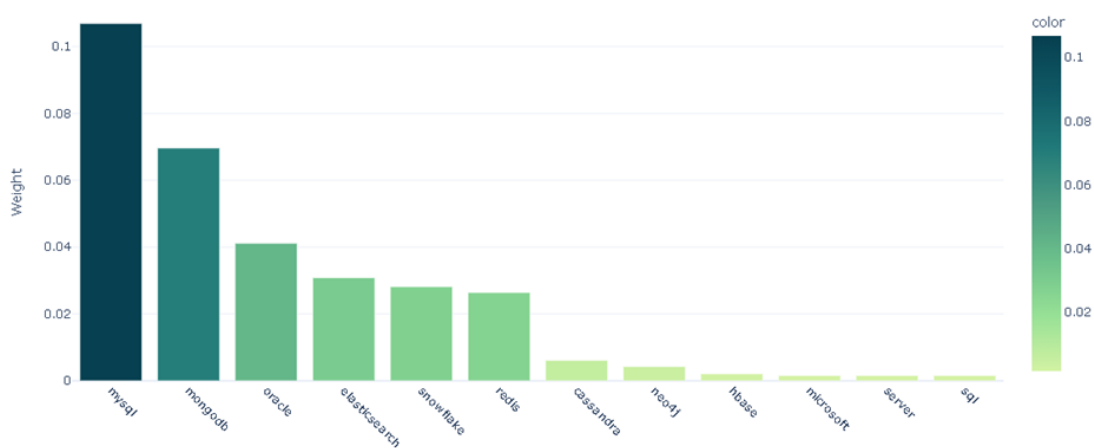
	skills_prog_langs	skills_platform	skills_databases	skills_frameworks
0				
1				
2	Python			Tensorflow Pytorch
3	Go ML SQL Python Scala B	Amazon Sagemaker		TensorFlow PyTorch
4	Python R SQL			
...
363	Lean Kotlin Ruby Go Python	Amazon Web Services		
364				
365	Lean Kotlin Ruby Go Python	Amazon Web Services		
366	Lean			
367	Delphi SQL			
368 rows × 4 columns				

Pour finaliser notre but était de remonter les outils les plus demandé dans le marché de travail marocain et c'est ce qu'on va visualiser pour chaque entité comme suit :

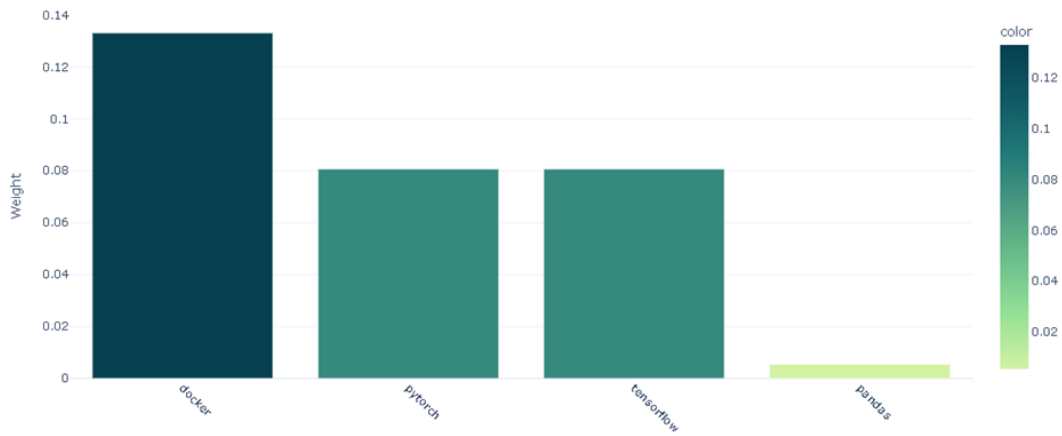
- Pour les langages de programmation :



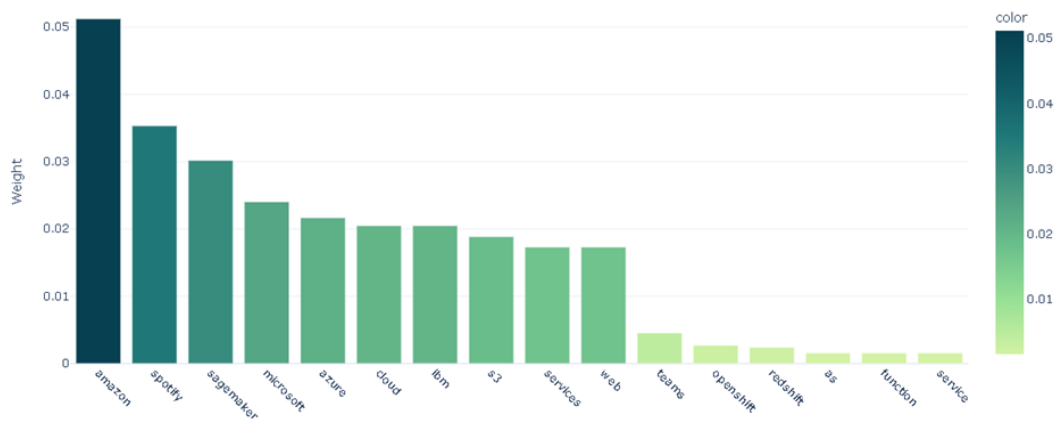
- Pour les bases de données :



-Pour Frameworks :



-Pour les plateformes :



-Vous pouvez consulter le code du projet sur ce lien !
https://github.com/AlaouiMdaghriAhmed/NER_data_jobs/

Conclusion

Ce projet nous a permis d'appliquer les connaissances acquises durant notre cursus à l'ESI, telles que l'utilisation de Python, Intelligence artificielle et la gestion de projet mais aussi la communication. Le projet nous apporte également une idée sur l'organisation dans la réalisation des projets et qui permettra de mieux s'adapter lors des prochains projets, stages et dans le monde professionnel d'une façon générale.

