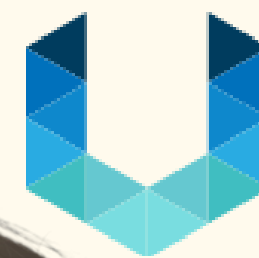


UNSTRUCTURED : SIMPLIFYING NLP TASKS AND DATA PIPELINES

INTEGRATIONS
WITH 8 ML
PLATFORMS



WHAT CHALLENGE?

CHALLENGE SOLVED : RAW TEXT CONVERTED TO DATA STRUCTURES

THERE ARE MODALITIES, TASKS AND AI MODELS

MODALITIES : TEXT, IMAGES, VIDEO, AUDIO AND GAMES

TASKS: EACH MODALITIES HAVE DIFFERENT TASKS

MODELS : MODELS ARE CREATED BASED ON MODALITIES AND THEIR TASKS

EACH OF THESE MODELS AND THE PLATFORMS HOSTING THEM REQUIRE **DIFFERENT DATA STRUCTURES**

CONVERTING TEXT INTO DATA STRUCTURES

MODE THAT IS PREVALENT IS TEXT

UNSTRUCTURED WORKS WITH TEXT

- HTML TO PLAIN TEXT
- TEXT PDF TO IMAGE EMBEDDED
PDF
- PRESENTATION TO EMAILS &
MSG
- AND MORE

EXTRACTING TEXT

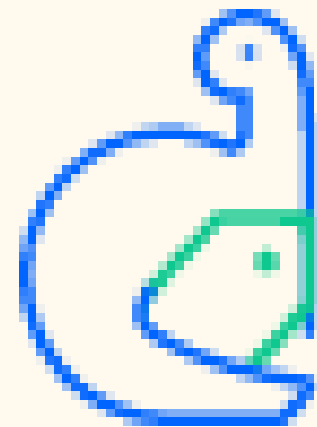
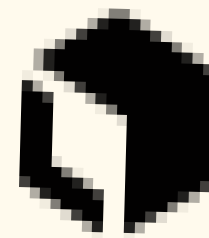
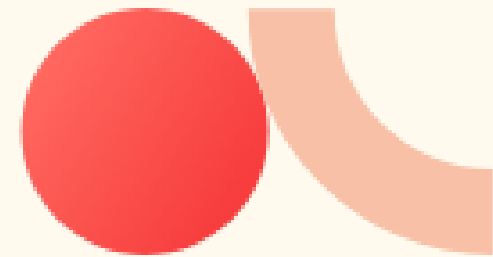
- IP ADDRESS/ MAIL ID/ PH NUM
- SELECTING WITH REGEX
- EXTRACTING DATES
- REPLACING UNICODE CHARS
- REMOVING SPACES + BULLETS
- TRANSLATING BETWEEN LANGS

INTRODUCING 8 MASS INTEGRATIONS

MASS: MODEL AS A SERVICE

INTEGRATIONS TO ML PLATFORMS

- ARGILLA
- DATASAU
- HUGGING FACE
- LABELBOX
- LABEL STUDIO
- LANGCHAIN
- PANDAS
- PRODIGY




HOW INTEGRATION IS DONE?

- DATA IS INGESTED VIA PARTITION
- IF CLEANING REQUIRED, ITS DONE AT CLEANING STEP
- STAGING CONVERTS THE TEXT INTO THE DATA STRUCTURE ACCEPTABLE BY VARIOUS ML PLATFORMS

LETS HEAD TO COLAB


NOW WE ARE TALKING!!! PRACTICE


insightbuilder/
python_de_learners_data





Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning

 2
Contributors

 0
Issues


 13
Stars

 4
Forks



python_de_learners_data/exploring_unstructured_part1.ipynb at main · insightbuilder/python_de_learners_data

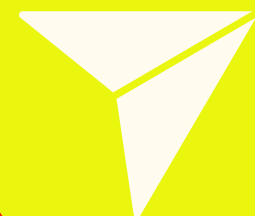
Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning - python_de_learners_data/exploring_unstructured_part1.ipynb at...

 GitHub

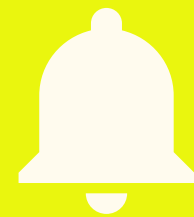
THANKS FOR WATCHING



LIKE



SHARE



SUBSCRIBE