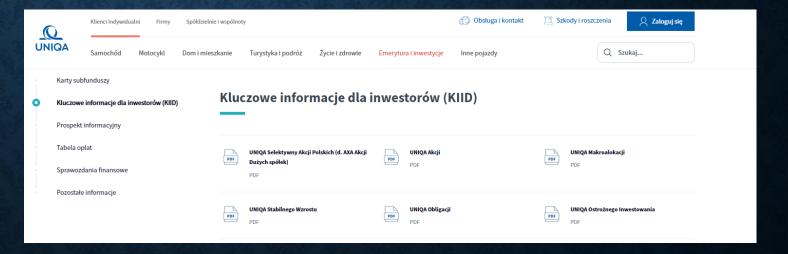
MAGIW #KARATEKIID PROJECT

Zespół MAGIW: Anna Wawrzyńczak, Wiktoria Śliwińska, Izabela Telejko, Marcel Witas, Grzegorz Zbrzeżny

WEBSCRAPPING

- BeatifulSoup
- Filtry (omijanie stron zewnętrznych, niektórych końcówek)
- Fuzzy Matching do wykrywania KIID-ów wśród PDF-ów
- Utworzenie tabeli z nazwami plików



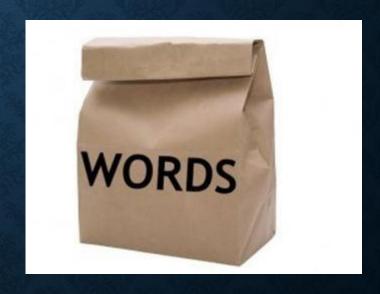
PODSTAWOWE INFORMACJE O KIIDACH

- pdfplumber
- Utworzenie kolumn ze specyfikacji
- Wiele warunków, wynikających z różnorodności dokumentów
- Wyrażenia regularne



BAG OF WORDS

- Spacy
- Lematyzacja



WERYFIKACJA OBECNOŚCI PODSTAWOWYCH WYRAŻEŃ

Fuzzy Matching Dokładność 95%

DZIĘKUJEMY ZA UWAGĘ!

