# ISI Hakaton – Klasyfikator przepisów

# Raport nr 1

Lider: Maciej Brusiło

Pozostali członkowie: Emilia Cieszewska

Magdalena Karyś

Anna Hojan

## Ogólne założenia projektu:

Planujemy:

1. Scrawlować kilka do kilkudziesięciu tysięcy przepisów, korzystając z API udostępnianego przez http://www.bigoven.com/.
2. Składniki i treść przepisów będą tworzyć zbiór cech w naszym klasyfikatorze, kategoriami będą wyodrębnione w API: nazwa głównej kategorii i rodzaj kuchni danego przepisu.
3. Treść przepisów przed wrzuceniem do klasyfikatora będzie poddana odpowiedniej obróbce: usunięcie liczb, znaków interpunkcyjnych, linków itd.
4. Wynikiem działania dla każdego przepisu ze zbioru testowego będzie zwrócenie kilku kategorii najbardziej pasujących do danego przepisu.
5. W miarę możliwości na podstawie wyników zwracanych dla zbioru testowego będziemy ewaluować działanie klasyfikatora.

Klasyfikator nie będzie miał interfejsu graficznego, będzie działał w konsoli.

Wykorzystywane języki programowania:Python, Perl

Źródła repozytoriów:publiczne repozytoria na Githubie

## Co zrobiliśmy w tym tygodniu:

* utworzyliśmy repozytorium
* zapoznaliśmy się z api bigoven <https://api.bigoven.com>), poznaliśmy jego ograniczenia (w wersji darmowej maksymalnie 100 requestów na godzinę dla jednego api\_key)
* rozdzieliliśmy zadania
* napisaliśmy crawlera
* ściągnęliśmy, wstępnie przetworzyliśmy i zapisaliśmy ponad 1,5 tys przepisów.

## Co planujemy zrobić w następnym:

* Scrawlować tysiące przepisów.
* Stworzyć klasyfikator.
* Wytrenować klasyfikator.
* Sklasyfikować zbiór testowy i na tej podstawie zewaluować działanie klasyfikatora.

Repozytorium:https://github.com/AnnaHojan/Wyszukiwarka-przepisow