

Raport z Negocjacji

Kierownik GL03: Dominika Scisłowska

26 marca 2022

Spis treści

1	Opis zajęć - przeprowadzonych negocjacji	2
1.1	Jakiej aplikacji dotyczy projekt?	2
1.2	Skróty i definicje	2
1.3	Wstępny wybór technologii	2
1.4	Założenia i ograniczenia	3
2	Opis aplikacji	3
2.1	Spis widoków	4
2.2	Logika biznesowa	4
2.3	Mechanizmy	4
2.4	Możliwe problemy i rozwiązania	6
3	Streszczenie artykułu	6
4	Słowniczek do artykułu	7
5	Pytania do zleceniodawcy i odpowiedzi	7
6	Osoby w projekcie - na podstawie zajęć nr 3	8
6.1	Opis zajęć	8
6.2	Zadania poszczególnych osób	9

1 Opis zajęć - przeprowadzonych negocjacji

1.1 Jakiej aplikacji dotyczy projekt?

Tłem projektu jest działalność instytutu w Skierniewicach - The National Institute of Horticultural Research, który to specjalizuje się w badaniach dotyczących gleby - jej poziomu pH oraz obecności w niej organizmów żywych.

Mikrobiolodzy bazują między innymi na indykatorach, analizie DNA i zdjęciach mikroskopowych. Projekt ma na celu ułatwienie oraz zautomatyzowanie ich pracy, poprzez wyeliminowanie konieczności przeprowadzania badań chemicznych w procesie rozróżniania rodzajów bakterii/grzybów. Skutkiem realizacji tego projektu będzie program, a dokładniej kompleksowa aplikacja spełniająca wyżej wymienione zadanie (*więcej na temat aplikacji w rozdziale: "Opis aplikacji"*).

1.2 Skróty i definicje

- The National Institute of Horticultural Research - w skrócie InHort.
- poziom pH gleby - odczyn gleby wskazujący na jej kwasowość lub zasadowość.
- indykator - chemiczna nazwa wskaźnika.
- zdjęcie mikroskopowe - rodzaj fotografii wykonany przy użyciu mikroskopu, podłączonego do aparatu.
- FCBF - Fast Correlation Based Filter.
- Python - wysoko poziomowy język programowania ogólnego przeznaczenia, cechujący się czytelnością i klarownością kodu źródłowego.
- C Sharp - wieloparadygmatowy obiektowy język programowania.
- Matlab - środowisko do wykonywania obliczeń naukowych oraz inżynierskich, a także do tworzenia symulacji komputerowych.

1.3 Wstępny wybór technologii

Zastosowane mechanizmy będą opierały się na porównaniu kluczowych cech bakterii/grzybów przedstawionych na wprowadzonych zdjęciach mikroskopowych, takich jak:

1. kształt,
2. rozmiar,
3. kolor,

4. liczebność populacji.

z dostępnymi już w bazie, dlatego wśród zalecanych technologii znalazły się:

- Python
- Mathlab

Zawierają one gotowe metody, które pozwolą na sprawniejszą i bardziej ergonomiczną realizację projektu. Kluczowa jest również dokładność i trafność przeprowadzonych dopasowań, także jeśli mówimy o cechach alternatywnych dla koloru.

Wstępnie wybrana technologia to: Python.

1.4 Założenia i ograniczenia

1. Napisany w wyniku realizacji projektu kod powinien być przejrzysty oraz zawierać komentarze.
2. Projekt powinien być pracą grupową.
3. Kluczowymi elementami projektu są:
 - (a) analizy i badania,
 - (b) wybór technologii,
 - (c) wybór kierownictwa,
 - (d) adekwatna dokumentacja.
4. Zdjęcia mikroskopowe wykorzystywane do stworzenia bazy, a także nowe, wprowadzane do rozpoznania będą wykonane na tym samym przybliżeniu.

2 Opis aplikacji

Funkcja aplikacji:

- rozróżnianie poszczególnych rodzajów bakterii/grzybów na podstawie zdjęć mikroskopowych załączanych przez użytkownika.

Podstawowe cechy aplikacji podane przez zleceniodawcę:

- aplikacja okienkowa - łatwa w obsłudze i przyjazna użytkownikowi,
- możliwość sledzenia jaką część obliczeń wykonano,
- elegancka prezentacja wyników - właściwe ich formatowanie,
- możliwość rozróżniania organizmów na poziomie rodzajów,
- prezentacja ile procent bakterii/grzybów zostało prawidłowo przypisanych do odpowiedniej klasy.

2.1 Spis widoków

Aplikacja będzie zawierała kilka widoków i trybów, a także przyciski do przełączania między nimi. Tryby i widoki:

1. wprowadzania danych/odczytu wyników
2. ułatwiający korzystanie z aplikacji osobom z poważnymi wadami wzroku
- żółto/niebieski (opcjonalnie, dodatkowo: jasny/ciemny),
3. uproszczony (opcjonalnie, dodatkowo: szczegółowy)

Interface powinien być jak najprostszy (bez opisów i linków) oraz wystarczająco intuicyjny, by móc obsługiwać całą aplikację bez wcześniejszego obejrzenia samouczka.

Wstępny projekt GUI: Po uruchomieniu aplikacji pojawi się miejsce służące do wczytania plików, następnie po ich załadowaniu pokaże się przycisk do zatwierdzenia. Po zatwierdzeniu aplikacja będzie przenosiła użytkownika do okna, w którym na bieżąco będą pokazywały się gotowe wyniki. W tym widoku, na samym dole będzie znajdował się pasek ładowania, wraz z przyciskiem do ewentualnego zatrzymania procesu. Ostateczną prezentację wyników użytkownik będzie mógł zobaczyć w formie siatki, bądź w formie listy. Będzie również dostępna opcja zapisu wyników do pliku .csv lub .pdf.

2.2 Logika biznesowa

Narzędzie na podstawie posiadanych wcześniej danych (bazy zdjęć mikroskopowych, cech konkretnych rodzajów bakterii/grzybów) i przy wykorzystaniu zastosowanych mechanizmów przyporządkowuje przesłane przez użytkownika zdjęcia prezentujące bakterie/grzyby do odpowiednich rodzajów. Zwraca ono wynik przeprowadzonych obliczeń zgodny z wymaganiami podanymi przez zleceniodawcę (*"Opis aplikacji"* - *"Podstawowe cechy aplikacji..."*). Jest ono dedykowane mikrobiologom pracującym w InHort. W projekcie używamy: przetwarzania obrazu, statystyki i wybranych obszarów sztucznej inteligencji.

2.3 Mechanizmy

Schemat działania będzie opierał się na czterech głównych składowych:

Segmentacja Rejonu Zainteresowania

Przy użyciu maski binarnej można odróżnić bakterie/grzyby od tła. Wtedy organizmy można oznaczyć na przykład na białą, a ich otoczenie na czarno.

Wyliczanie Cech

1. wariancja koloru czerwonego na obrazku,
2. wariancja koloru niebieskiego na obrazku,

3. wariancja koloru w skali grayscale,
4. krzywizna bakterii,
5. ilość bakterii na zdjęciu przy danym przybliżeniu (z założenia stałe),
6. cechy związane z teksturą.

Wybór Cech

Na przykład przy użyciu:

- maska, filtr LaPlace'a - korzystamy z zinterpolowanych wielomianów, a następnie wytyczamy krzywizny.
- metoda k-średnich - przyporządkowujemy punkty do danej grupy w następujący sposób:
 1. wyznaczamy centroidy,
 2. mierzymy odległości punktów od centroid i przypisujemy je do tych bliżej położonych,
 3. wyliczamy średnie (x i y) i w te punkty przesuwmy centroidy,
 4. wyliczmy nowe odległości punktów od centroid i zmieniamy przyporządkowania.
- FCBF - algorytm dobierający cechy wyróżniające daną klasę (rodzaj mikroorganizmów) oraz w najmniejszym stopniu powiązane z innymi cechami.

Rozpoznanie Rodzajów

- K-Najbliższych Sąsiadów - metoda w której:
 1. mierzymy odległości punktów i numerujemy je od tych najbliższych od danego do tych najdalszych,
 2. sprawdzamy dane k i na podstawie przewagi elementów z danej grupy ustalamy grupę dla całości,
 3. UWAGA! Dla takiej samej liczby grupę przydzielamy losowo.
 4. korzystamy tu z odległości euklidesowej:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (1)$$

gdzie:

- x,y - dwa punkty w euklidesowej przestrzeni n,
- n - przestrzeń n.

- Perception
- Extreme Machine Learning - metoda, w której "uczymy" na bazie dostępnych zdjęć dopasowywać nowe do konkretnych rodzajów.

2.4 Możliwe problemy i rozwiązania

- zbyt długi czas realizacji projektu - podział zespołu na grupę deweloperską i testową, jednoczesna praca,
- brak organizacji - wybór kierowników obu grup oraz jednego głównego,
- zbyt skomplikowane metody do implementacji - użycie odpowiednich technologii z gotowymi metodami.

3 Streszczenie artykułu

Artykuł opublikowany w: *"Applied Mathematics and Information Sciences 2016"*.
Pt. *"Computerized Classification System for the Identification of Soil Microorganisms"*

Artykuł skupia się na, innym niż dotychczas znane, rozwiązaniu dotyczącym obecności mikroorganizmów w glebie. Obecność bakterii/grzybów wpływa na ilość azotu, a zatem również na żyzność gleb, ich kompleksowe wyszukiwanie na podstawie zdjęć mikroskopowych, bez konieczności prowadzenia szczegółowych badań chemicznych może zautomatyzować i usprawnić pracę mikrobiologów w instytucie.

Artykuł powstał we współpracy osób reprezentujących trzy ośrodki:

1. SGGW,
2. Politechnikę Warszawską,
3. Zakład Agrotechniki Instytutu Ogrodnictwa w Skierniewicach.

W badaniach uznano 12 rodzin mikroorganizmów.

W rozwiązaniu problemu rozpoznawania klas organizmów zastosowano metodę składającą się z czterech etapów:

1. rozdzielenie tła zdjęcia mikroskopowego od badanych obiektów,
2. generowanie, wyszukiwanie charakterystycznych cech,
3. wybieranie i ocenianie cech,
4. przyporządkowanie rodzin na podstawie wybranych cech.

Każdy z etapów został precyzyjnie opisany.

Dane użyte w eksperymentach zostały znormalizowane poprzez podzielenie każdej kolumny movierzy przez wartość maksymalną. W końcowych akapitach zostały zaprezentowane otrzymane wyniki i wnioski, a także zapowiedź kontynuacji badań. Warto wspomnieć, że najlepszy losowy klasyfikator uzyskał dokładność 98,41 procenta na bazie zawierającej 441 zdjęć.

4 Słowniczek do artykułu

- SGGW - Szkoła Główna Gospodarstwa Wiejskiego w Warszawie,
- cecha - element odróżniający lub charakteryzujący (w tym przypadku mikroorganizmy)

5 Pytania do zleceniodawcy i odpowiedzi

1. Jaka ma być maksymalna liczba zdjęć przy pojedynczym przesyłaniu?
ODPOWIEDŹ: Nie ma podanej maksymalnej liczby zdjęć.
2. Czy aplikacja ma być w jakiejś konkretnej szacie graficznej?
ODPOWIEDŹ: Nie, ale aplikacja powinna być przejrzysta.
3. Na jakim systemie operacyjnym aplikacja będzie używana najczęściej?
ODPOWIEDŹ: Windows/Linux
4. Czy przewidziane jest rozbudowywanie aplikacji o kolejne funkcje w przyszłości?
ODPOWIEDŹ: Jeśli będzie sensownie zrobiona to tak.
5. Czy powinna być dostępna możliwość zapisywania wyników do pliku?
ODPOWIEDŹ: Tak.
6. Czy aplikacja ma być tylko i wyłącznie w języku polskim? Jeśli nie to w jakim jeszcze?
ODPOWIEDŹ: Jeśli będzie czas to może być jeszcze wersja anglojęzyczna.
7. Jaki jest czas na wykonanie projektu?
ODPOWIEDŹ: Do ostatnich zajęć.
8. Czy powinna być dostępna baza z wynikami do ich odczytu na różnych komputerach?
ODPOWIEDŹ: Nie, wystarczy wersja lokalna. Wszystkie pliki mają być przechowywane lokalnie i wgrywane za każdym razem od nowa.

9. Czy jest możliwość zaprezentowania wersji testowej celem jak najlepszego dostosowania ostatecznej wersji do potrzeb użytkowników?

ODPOWIEDŹ: Nie, oddajemy cały projekt.

10. Jaki jest planowany budżet na realizację tego projektu?

ODPOWIEDŹ: 0zł.

11. Czy na zdjęciu może być przedstawiony więcej niż jeden rodzaj bakterii/grzybów?

ODPOWIEDŹ: Nie.

12. Czy na zdjęciu może nie być żadnych bakterii/grzybów? Co w taki przypadku?

ODPOWIEDŹ: Aplikacja zawsze ma nam podać dopasowany rodzaj bakterii/grzybów, nawet jeżeli na zdjęciu będzie przedstawione coś innego.

13. Jak wiele różnych rodzajów mikroorganizmów będziemy uwzględniać?

ODPOWIEDŹ: Prawdopodobnie około 5 klas/rodzajów.

14. Jaki procent dokładności powinniśmy osiągnąć dla bakterii?

ODPOWIEDŹ: Jak największy. Dla cech związanych z kolorem 95 procent, natomiast dla pozostałych - 85 procent.

6 Osoby w projekcie - na podstawie zajęć nr 3

6.1 Opis zajęć

Oprócz prezentacji szczegółów dotyczących ról w zespole, na zajęciach wybrani studenci prezentowali swoje raporty. Otrzymaliśmy także odpowiedzi na pytania oraz cenne źródła dodatkowej wiedzy:

- artykuł: "Machine Learning and Deep Learning Based Computational Approaches in Automatic Microorganisms Image Recognition: methodologies, challenges and developments."
- kanał na YouTube: "StatQuest"

W pracy grupowej kluczowy jest odpowiedni podział zadań, który pozwoli na sprawną i możliwie jak najmniej problemową realizację całego projektu. Proces wyboru poszczególnych ról może wyglądać następująco:

1. Wybór kierowników.
2. Podział pozostałych członków zespołu na dwie grupy:
 - deweloperską,
 - testową.

6.2 Zadania poszczególnych osób

Grupa Deweloperów

Ta grupa może zostać podzielona na dwie podgrupy:

1. **Grupa eksperymentalna:** testuje ona modele, przeprowadza crossvalidację, wyznacza dokładności dla poszczególnych modeli, itd.

Przykładowo: 10 różnych modeli np. topologia i wagi, badanie która z metod daje największą dokładność np. RF, równanie granicy decyzyjnej.

W tej grupie kluczowymi technologiami/paczkami mogą być: Python, Tensor Flow, Pytorch, Numpy.

Praca tej grupy może być podzielona na kilka etapów, z których większość może być wykonywana równolegle, co bez wątpienia przyspieszy realizację projektu:

- (a) Preprocessing zdjęć - nakładanie filtrów, zmiana kontrastów.
- (b) Segmentacja Regionu Zainteresowania - zastosowanie masek.
- (c) Wyliczanie cech zdjęć - zarówno dla zdjęć przetworzonych jak i nieprzetworzonych.
- (d) Wybór cech (np. przy użyciu FCBF) i klasyfikacja (np. RF, MLP, kNN).

2. **Grupa programowa:** tworzy program na podstawie najlepszych modeli, wybranych przez grupę pierwszą. Wylicza cechy (z grupy eksperymentalnej) dla danych wejściowych. Na podstawie modelu klasyfikuje nowe dane wejściowe - nowe zdjęcia.

Polecane metody: uśrednianie wagi, najlepsze wagi, RF.

W tej grupie kluczowymi technologiami/paczkami mogą być: C Sharp, Python - na tę technologię się zdecydowaliśmy.

Praca tej grupy może być podzielona na kilka etapów:

- (a) Stworzenie GUI.
- (b) Otrzymanie modelu od grupy eksperymentalnej i zadbanie o parametry, żeby mogły być wykorzystane w języku programowania.
- (c) Stopniowa praca nad mechanizmem przetwarzania zdjęć:
 - jedno zdjęcie,
 - wiele zdjęć - jedno po drugim,
 - wiele zdjęć - asynchronicznie (na samym końcu)
- (d) Praca nad mechanizmem pokazującym ile zdjęć zostało już przetworzonych.
- (e) Praca nad prezentacją wyników i zapisem ich do pliku np. do .csv, .pdf

Grupa Testerów

Ta grupa będzie skupiała się na wykonywaniu testów. Dzięki znajomości specyfiki aplikacji może ona pracować równolegle z grupą deweloperską - nie musi, a nawet nie może czekać na rezultat wyżej wspomnianej grupy, ponieważ wydłużyłoby to cały proces realizacji projektu.

Do zadań tej grupy zaliczamy:

1. Testowanie działania GUI.
2. Sprawdzanie czy wyniki z pierwszej grupy deweloperskiej pasują do założeń drugiej grupy deweloperskiej.
3. Tworzenie scenariuszy testowych.

Kierownicy Grup

Do zadań kierowników grup zaliczamy:

1. Koordynowanie pracy zespołu.
2. Rozdzielanie zadań poszczególnym osobom - sami również mogą wykonywać niektóre działania.
3. Organizacja spotkań grupy.
4. Zdawanie raportów do kierownika głównego.

Kierownik Główny

Do zadań kierownika głównego zaliczamy:

1. Koordynowanie pracy zespołów, zlecanie zadań.
2. Tworzenie dokumentacji - to zadanie może zostać częściowo lub w całości oddelegowane, ale wtedy konieczna jest kontrola postępów.
3. Organizacja spotkań z kierownikami i odbieranie od nich raportów.
4. Zdawanie raportów do Prowadzącego przedmiot, zawierających:
 - postępy prac,
 - co zostało zrobione,
 - co jeszcze nie zostało zrobione,
 - ewentualne problemy, nad których rozwiązaniem aktualnie pracuje zespół.

Uwagi

Należy pamiętać, że wszelkie kontakty między testerami i deweloperami powinny odbywać się przy udziale kierowników. Tylko w taki sposób będą mogły być one prawidłowo wprowadzone do dokumentacji.

Dodatkowo, warto wspomnieć, że każda osoba będzie uzupełniała cotygodniowy, krótki raport (około 5 zdań) odnośnie swojej pracy wykonanej w danym okresie.