

Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

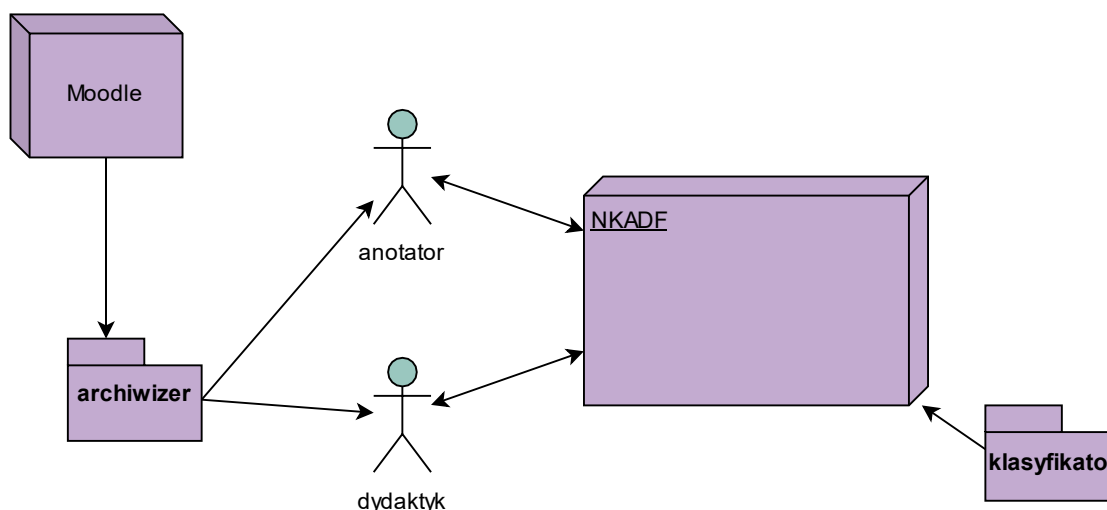
WYMAGANIA DLA PROJEKTU „KOMPUTEROWA ANALIZA DYSKUSJI NA FORUM”

0. WERSJE DOKUMENTU

wersja	data	zakres zmian
1.0	25.10.2020	połączono dokumentu wymagań i zakresu projektu; dostosowano go do wymaganego szablonu
1.1	03.11.2020	rozszerzono poszczególne sekcje dokumentu; przeredagowano sekcje 1, 3 i 4; zamieniono przypadki użycia na historyjki użytkownika w sekcji 3
1.2	10.11.2020	ponownie przeredagowano sekcje 1, 3 i 4; rozbudowano historyjki użytkownika; dodano diagramy
1.2.1	17.11.2020	przerysowano niektóre diagramy; zmieniono historyjki użytkownika na przypadki użycia w sekcji 3
1.3	24.11.2020	przemianowano i rozbudowano sekcję 1, przeniesiono do niej część podsekcji z innych sekcji; dodano diagram przypadków użycia w sekcji 3
1.4	21.12.2020	przeniesiono diagramy do dokumentu architektury systemu, zredagowano sekcję 1; uaktualniono kamienie milowe w sekcji 10; udoskonalono dokument

1. ELEMENTY SKŁADOWE PROJEKTU

1.1. KONTEKST FUNKCJONOWANIA



Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

Narzędzie do Komputerowej Analizy Dyskusji na Forum będzie wykorzystywane przez dydaktyka lub anotatora. Do rozpoczęcia pracy potrzebne będą dane z zewnętrznego serwisu Moodle, który pozwala pobrać kopię zapasową kursu, a więc również pliki forum. Z uwagi na ochronę danych osobowych, użytkownik nie może przekazać całej kopii zapasowej – jeśli nie została ona zanonimizowana – a jedynie pojedyncze pliki. Do ich uzyskania potrzebny jest program do archiwizacji, który rozpakuje kopię zapasową.

1.2. WYSZCZEGÓLNIONE PRODUKTY PROJEKTU

1.2.1. OPROGRAMOWANIE

W czasie trwania projektu zostanie wytworzone następujące oprogramowanie:

- **Klient aplikacji webowej**
 - źródło strony w języku HTML5
 - wygląd w oparciu o framework Nebulari własne stylowanie za pomocą Sass
 - zarządzanie logiką aplikacji za pomocą biblioteki Angular2+
 - wyświetlanie grafów oraz wykresów za pomocą biblioteki D3.js
- **Serwer aplikacji webowej** w języku Python oparty o framework Django
 - skrypty logiki biznesowej
 - parser pilków forum Moodle'a
- Relacyjna **baza danych** oparta o MySQL
- **Klasyfikator tekstu** stworzony z użyciem biblioteki Python Scikit-learn
 - parser plików forum Edumatic
 - skrypt uczenia maszynowego
 - skrypt predykcji
 - model uczenia maszynowego do polaryzacji opinii
 - model uczenia maszynowego do klasyfikacji argumentów retorycznych

1.2.2. ZBIORY DANYCH

- czysty korpus tekstowy wypowiedzi na forum
 - v1: 6000 wszystkich paragrafów
 - v2: 2000 wyselekcjonowanych paragrafów
- otagowany pod kątem polaryzacji wypowiedzi korpus 1500 paragrafów
- otagowany pod kątem argumentów retorycznych korpus 800 paragrafów
- przykładowe forum Moodle'a na potrzeby demo

1.2.3. DOKUMENTACJA

Zostaną stworzone następujące dokumenty:

- wizja projektu
- wymagania projektowe
- raport z uczenia maszynowego (prezentacja)

Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

- dokumentacja serwera REST API
- dokumentacja klienta aplikacji
- opis architektury system

1.3. DEFINICJE, AKRONIMY I SKRÓTY

- *anotacja* – przyporządkowanie pewnego określenia (klasy) wycinkowi tekstu
- *anotator* – typ użytkownika,
- *anonimizacja danych* – usunięcie danych osobowych z kopii zapasowej Moodle'a
- *archiwizator plików, program do archiwizacji* – zewnętrzny program do „pakowania i rozpakowywania plików”, np. 7-Zip, WinRAR
- *backup* – patrz: kopia zapasowa
- *demo* – wersja klienta aplikacji dostępna dla niezalogowanego użytkownika z predefiniowanym forum
- *dydaktyk* – typ użytkownika, zwykle zarządzający forum na Moodle
- *kopia zapasowa* – plik z rozszerzeniem .mbz, zawierający zarchiwizowane dane na temat kursu, możliwy do pobrania z Moodle'a
- *Moodle* – środowisko nauczania zdalnego
- *NKADF* - Narzędzie do Komputerowej Analizy Dyskusji na Forum
- *Plik forum* – plik .xml pochodzący z Moodle'a, zawierający opis forum
- *tagowanie* – patrz: anotacja
- *użytkownik* - osoba korzystająca z narzędzia NKADF
- *uczestnik* – osoba biorąca udział w tworzeniu forum na Moodle'u
- *wymiar anotacji* – zbiór klas na potrzeby anotacji
- *wytyczne* – opis wszystkich klas w ramach wymiaru
- *Wydział, WMI* – Wydział Matematyki i Informatyki UAM

2. GRANICE PROJEKTU

- Wczytane do systemu będą mogły być tylko pliki forum w formacie XML wygenerowane przez Moodle'a (lub takie o identycznej strukturze). Nie jest przewidziane wsparcie forum z innych platform e-learningowych (np. Edumatic lub Blackboard).
- Nie będzie możliwe adnotowanie poszczególnych fragmentów tekstu (np. pojedynczych słów), a jedynie całych paragrafów.

3. WYMAGANIA FUNKCJONALNE

3.1. SKRÓCONE PRZYPADKI UŻYCIA

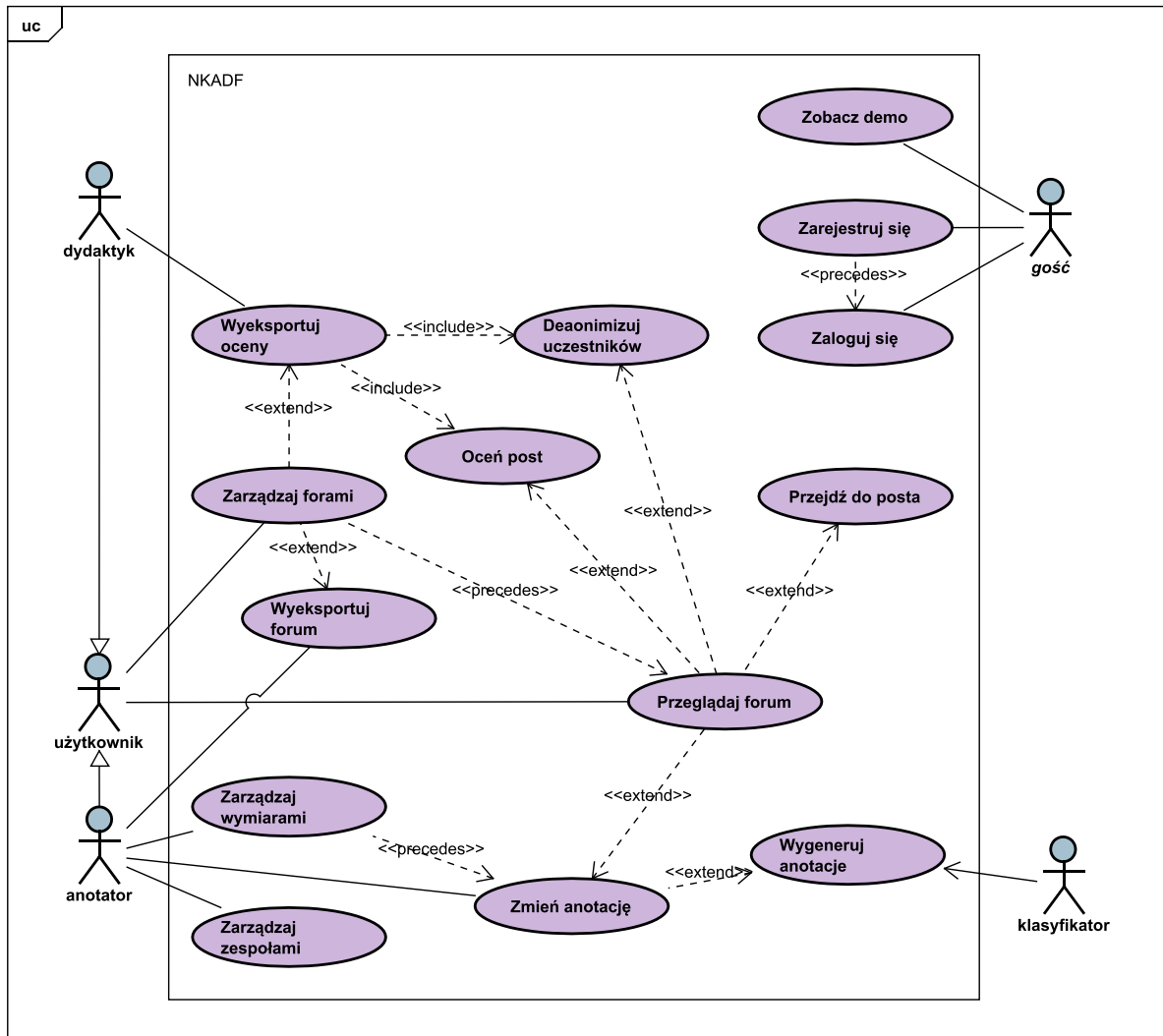
Pełne przypadki użycia znajdują się w dokumencie *Przypadki użycia*.

nazwa Opis

Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

<i>Zobacz demo</i>	Niezalogowany użytkownik chce wypróbować produkt. System udostępnia przykładowe forum do pracy.
<i>Zarejestruj się</i>	Użytkownik rejestruje się w systemie. Podaje login i hasło. System tworzy konto użytkownika.
<i>Zaloguj się</i>	Użytkownik chce się zalogować. Podaje login i hasło. System sprawdza poprawność danych i loguje użytkownika. W przypadku niepoprawnych danych, system podaje komunikat o błędzie.
<i>Zarządzaj forami</i>	Użytkownik może dodawać, usuwać i aktualizować pliki forum, dodawać do nich wymiary anotacji oraz zespoły. System waliduje przesłane fora pod względem poprawności pliku, w razie błędu w parsowaniu wyświetla komunikat.
<i>Zarządzaj wymiarami</i>	Użytkownik może tworzyć i edytować wymiary anotacji oraz dodawać do nich wytyczne.
<i>Zarządzaj zespołami</i>	Użytkownik może tworzyć i edytować zespoły anotatorów.
<i>Przełóż forum</i>	Użytkownik wybiera forum do pracy. System załaduje listę wątków dostępnych na forum. Użytkownik wybiera wątek. System wyświetla posty w ramach wątku. Użytkownik może wybrać kolejny wątek w bocznym menu.
<i>Przejdź do postu</i>	Użytkownik wybiera węzeł postu w mapce wątku. Widok forum przewija się na wysokość wybranego postu.
<i>Deanonimizuj uczestników</i>	Użytkownik przygotowuje listę identyfikatorów i nazwisk, i przekazuje ją do systemu. System podmienia identyfikatory na nazwiska uczestników w widoku forum.
<i>Oceń post</i>	Użytkownik przypisuje ocenę liczbową do postu w wątku.
<i>Wyeksportuj oceny</i>	System przygotowuje zestawienie uczestników forum i ocen przypisanych do ich postów w ramach forum.
<i>Wygeneruj anotacje</i>	System generuje anotacje w predefiniowanym wymiarze i taguje wypowiedzi w ramach forum.
<i>Zmień anotacje</i>	Użytkownik wybiera wymiar w ramach którego będzie anotował forum, następnie przypisuje jedną z dostępnych kategorii paragrafowi tekstu. Może ją potem nadpisać. System automatycznie zapisuje zmiany.
<i>Wyeksportuj forum</i>	System sprawdza, czy wszystkie wypowiedzi zostały otagowane i pyta użytkownika o potwierdzenie akcji. System tworzy plik zawierający wypowiedzi oraz przypisane im tagi.

3.2. DIAGRAM PRZYPADKÓW UŻYCIA



4. WYMAGANIA NIEFUNKCJONALNE

- Responsywność klienta aplikacji webowej
- Skalowalność
- Rozszerzalność
- Łatwość wdrożenia i utrzymania
- Hostowanie aplikacji na subdomenie Wydziału
- Szyfrowanie HTTPS
- Zabezpieczenie przed SQL injectem dzięki modelowi ORM

Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

- Autoryzacja logowania, hash/salt na dane logowanie przechowywane w bazie danych
- Komunikacja wyłącznie na linii serwer-klient, z brakiem możliwości połączenia z zewnątrz
- Obsługa danych wejściowych wyłącznie w ściśle zdefiniowanym formacie
- Uwierzelnianie użytkownika i sesji poprzez JSON Web Token
- Zabezpieczenie przed Forced Browsing po stronie klienta aplikacji

5. MIERZALNE WSKAŹNIKI WDROŻENIOWE

- Wersja pierwotna aplikacji zostanie przedstawiona klientowi pod koniec pierwszego semestru i przez niego zaakceptowana
- Wersja ostateczna aplikacji zostanie udostępniona w domenie wydziałowej w czasie trwania drugiego semestru i otrzyma pięć recenzji od użytkowników.
- W systemie zarejestruje się przynajmniej kilku unikalnych użytkowników.
- Klasyfikator osiągnie wyniki zbliżone do podobnych modeli tego typu.

6. KRYTERIA AKCEPTACJI PROJEKTU DLA I SEMESTRU PRAC

6.1. WYMAGANE

- klient i serwer aplikacji działają lokalnie
- przejście użytkownika przez następujące kroki: wgranie pliku, wyświetlenie forum, zmiana anotacji dla kilku wybranych paragrafów
- korpus przygotowany z co najmniej $\frac{1}{3}$ dostępnych danych
- model klasyfikacji polaryzacji opinii
- wyświetlanie predykcji klasyfikatora w aplikacji

6.2. OCZEKIWANE

- zaimplementowana wizualizacja dyskusji w formie drzewa dla wybranego wątku
- deanonimizacja uczestników forum z poziomu przeglądarki

6.3. PLANOWANE

- użycie sieci neuronowych w uczeniu maszynowym

7. KRYTERIA AKCEPTACJI PROJEKTU DLA II SEMESTRU PRAC

7.1. WYMAGANE

- klient i serwer aplikacji są dostępne w domenie wydziałowej
- możliwość rejestracji, logowania i zarządzania panelem użytkownika
- klasyfikator rodzajów argumentów o dowolnej precyzji

Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

7.2. OCZEKIWANE

- przygotowane demo dla niezarejestrowanych użytkowników
- tworzenie zespołów do pracy nad jednym forum

7.3. PLANOWANE

- dodawanie zewnętrznych modeli klasyfikacji

8. ORGANIZACJA PRACY ZESPOŁU

Zespół wykonawczy:

- [Marcin Armacki](#) – back-end developer, projekt aplikacji serwerowej, administracja i projekt bazy danych, Machine Learning,
- [Karolin Izabel Boczoń](#) – liderko zespołu, analiza biznesowa, przetwarzanie tekstu
- [Michał Romaszkin](#) – front-end developer, Design UI/UX, Web Development

Zadania współdzielone: opracowywanie dokumentacji, przygotowanie danych do uczenia maszynowego

Interesariusze:

- [Prof. UAM dr hab. Jacek Marciniak](#) – klient, wsparcie w zakresie zarządzania projektem, dostarczenie danych
- [Dr Marek Kubis](#) – wsparcie merytoryczne w przygotowaniu klasyfikatora
- [Pracownicy WMI zainteresowani produktem](#) - treści, funkcjonalności, testowanie

9. RYZYKA PROJEKTOWE

- Oferowane przez nas modele do klasyfikacji tekstu mogą nie spełnić oczekiwań użytkowników końcowych.
- Zmiana koncepcji wynikająca z badawczego charakteru projektu.
- Brak zainteresowania projektem wśród społeczności akademickiej wynikający z bardzo wąskiego zastosowania produktu.

10. KAMIENIE MIŁOWE

I.p.	kamień milowy	termin
1.	Dostarczenie pierwszej wizji projektu	03.04.2020
2.	Prezentacja pierwszego przyrostu: <ul style="list-style-type: none"> ▪ wgranie i walidacja pliku XML w aplikacji webowej ▪ parser forum XML -> JSON 	24.04.2020
3.	Otrzymanie danych językowych do budowy klasyfikatorów	02.05.2020
4.	Prezentacja drugiego przyrostu:	23.05.2020

Komputerowa analiza dyskusji na forum		22.12.2020
Wymagania		1.4
K. Boczoń	M. Armacki	M. Romaszkin

	<ul style="list-style-type: none"> ▪ zaadnotowane 1500 paragrafów ▪ pierwszy klasyfikator: polaryzacja opinii 	
5.	Prezentacja trzeciego przyrostu: <ul style="list-style-type: none"> ▪ wyświetlenie forum z pliku ▪ automatyczna adnotacja domyślnym klasyfikatorem ▪ możliwość zmiany adnotacji dla konkretnego paragrafu 	19.06.2020
6.	Zaprezentowanie prototypu aplikacji przed komisją	02.07.2020
7.	Dodanie obsługi kont użytkowników	10.11.2020
8.	Prezentacja czwartego przyrostu: <ul style="list-style-type: none"> ▪ mockupy 	24.11.2020
9.	Prezentacja piątego przyrostu: <ul style="list-style-type: none"> ▪ generowanie tagów z klasyfikatora (polaryzacja opinii) ▪ deanonimizacja uczestników 	22.12.2020
10.	Stworzenie baseline'u klasyfikatora rodzajów argumentacji	31.12.2020
11.	Udostępnienie aplikacji na wydziale	04.01.2021
12.	Zaprezentowanie projektu na seminarium wydziałowym	05.01.2021
13.	Zakończenie zbierania informacji zwrotnej od użytkowników	17.01.2021
14.	Dopuszczenie do obron	19.01.2021
15.	Obrona projektu	25.01.2021