

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

ARCHITEKTURA PROJEKTU „KOMPUTEROWA ANALIZA DYSKUSJI NA FORUM”

SPIS TREŚCI

1. Wstęp.....	1
2. Model Logiczny	1
3. Model wdrożeniowy	2
4. Komponenty.....	2
4.1. Frontend.....	2
4.2. Backend.....	3
4.3. Klasyfikator	4
5. Baza danych	5
6. Widoki.....	6
7. Wykorzystane technologie.....	7
8. Inne	Error! Bookmark not defined.

1. WSTĘP

Niniejszy dokument jest ogólną dokumentacją Narzędzia do komputerowej analizy dyskusji na forum (NKADF). Opisuje architekturę systemu, czynności potrzebne do instalacji, wdrożenia i utrzymania systemu oraz prezentację w postaci widoków.

2. MODEL LOGICZNY

NKADF jest aplikacją internetową typu klient-serwer. Głównym jej zadaniem jest udostępnienie użytkownikowi interfejsu, który umożliwia pracę nad forum dyskusyjnym, tj. import forum, wyświetlanie dyskusji, anotowanie i eksport zaanotowanego forum. By użytkownicy mogli zapisywać zmiany wprowadzono konta użytkowników, do których można przypisywać anotowane fora i wymiary anotacji.

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

3. MODEL WDROŻENIOWY

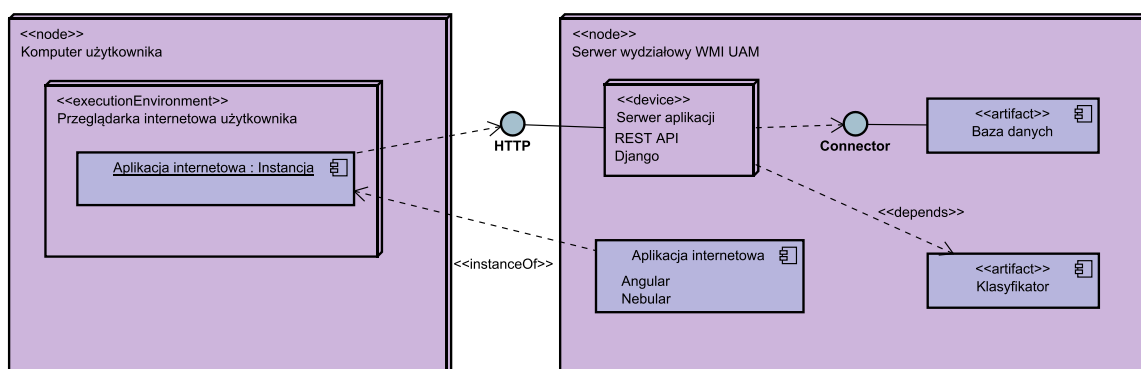


DIAGRAM 1 DIAGRAM WDROŻENIA

System składa się z serwera oraz klienta. Serwer znajduje się w domenie projektstudencki.pl należącej do Wydziału Matematyki i Informatyki UAM. Warstwa front-endowa uruchamiana jest w przeglądarce klienta po pobieraniu z serwera plików w wyniku zapytania HTTP. Usługa backendu działa po stronie serwera i obsługuje zapytania HTTP wysłane z klienta aplikacji. Baza danych znajduje się na serwerze i jest odpytywana przez aplikację. Aplikacja serwerowa wykorzystuje plik klasyfikatora do automatycznej anotacji dyskusji użytkowników, gdy zostanie wysłane odpowiednie zapytanie HTTP.

4. KOMPONENTY

4.1. FRONTEND

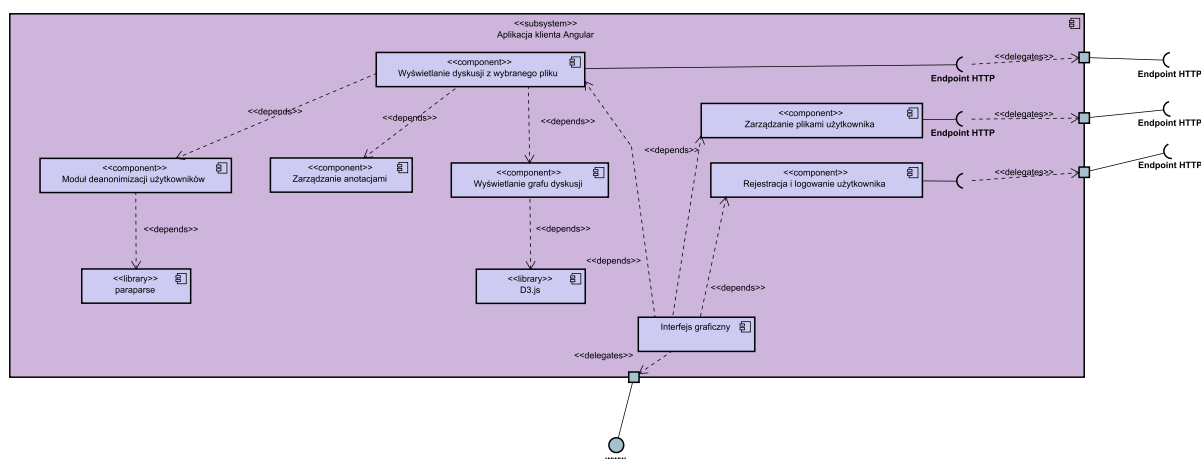


DIAGRAM 2 KOMPONENTY - FRONTEND

1. Moduł autoryzacji użytkownika - pozwala na rejestrację, logowanie oraz zarządzanie bieżącą sesją w przeglądarce

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

2. Moduł przeglądu pliku - umożliwia podgląd zawartości wybranego pliku z forum, wylistowanie dostępnych dyskusji, przypisanie wybranych przez użytkownika wymiarów oraz zanotowanie dyskusji za pomocą wbudowanych klasyfikatorów.
3. Moduł przeglądu dyskusji - odpowiedzialny za wylistowanie dyskusji, zawarte w nim są:
 - 3.1. Komponent paragrafu - umożliwia wyświetlenie danego paragrafu z edytowalną etykietą na podstawie wybranego wymiaru
 - 3.2. Funkcja eksportu danych do pliku TSV
 - 3.3. Funkcja wybrania wymiaru, w którym będzie przeglądana dyskusja
 - 3.4. Funkcja przekształcania ID autorów wypowiedzi na dane zawarte w pliku *users.xml* dostępnego z kopii zapasowej forum Moodle
 - 3.5. Komponent wizualizacji dyskusji
4. Moduł panelu głównego - odpowiedzialny z widok po zalogowaniu, zawarte w nim są:
 - 4.1. Komponent dostępnych dyskusji - umożliwia listowanie, dodawanie, usuwanie plików z forami
 - 4.2. Komponent dostępnych wymiarów - umożliwia listowanie, dodawanie oraz edycję wymiarów użytkownika

4.2. BACKEND

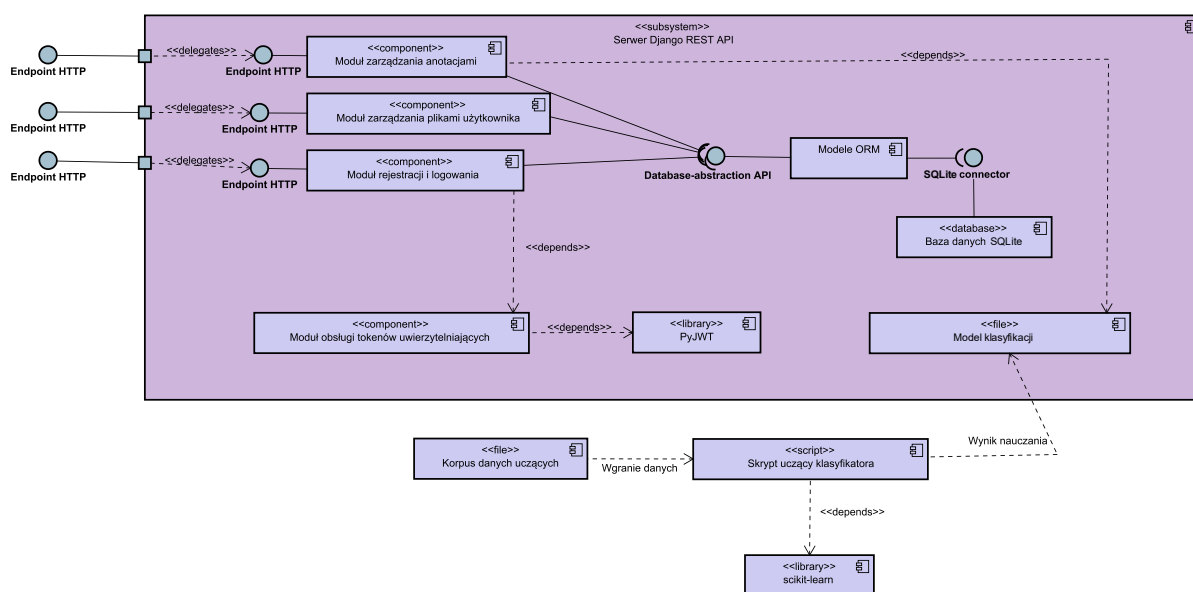


DIAGRAM 3 KOMPONENTY - BACKEND I KLASYFIKATOR

Na backend składają się następujące moduły:

1. Moduł zarządzania anotacjami - udostępnia funkcje związane z anotowaniem takie jak: stworzenie wymiaru, aktualizacja opisu wymiaru, wyświetlenie listy wymiarów,

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

wyświetlenie kategorii wymiaru, aktualizacja nazwy wymiaru, usunięcie wymiaru, przypisanie wymiaru do pliku użytkownika, wyświetlenie wymiarów przypisanych do pliku, usunięcie wymiaru z pliku, aktualizacja kategorii wymiaru, dodanie nowej kategorii, usunięcie kategorii, dodanie anotacji użytkownika, usunięcie anotacji, wygenerowanie anotacji przy użyciu klasyfikatorów

2. Moduł zarządzania plikami użytkownika - udostępnia funkcje związane z przetwarzaniem przesłanych plików takie jak: zapisanie informacji z pliku w bazie danych, wyświetlenie dyskusji z pliku, aktualizacja nazwy pliku, usunięcie pliku, eksport dyskusji do pliku .tsv.
3. Moduł rejestracji i logowania - udostępnia funkcje związane z zarządzaniem kontami użytkowników takie jak: rejestracja nowego użytkownika, logowanie użytkownika, sprawdzanie tokenu uwierzytelniającego, wylogowanie użytkownika.
4. Moduł obsługi tokenów uwierzytelniających - udostępnia następujące funkcje: sprawdzanie ważności tokenu, weryfikacja zawartości tokenu, wygenerowanie nowego tokenu, zastrzeżenie tokenu.
5. PyJWT – biblioteka języka Python udostępniająca funkcje do obsługi JSON-owych tokenów uwierzytelniających (JSON Web Token).
6. Modele ORM – definicje klas w języku Python odwzorowujące strukturę relacyjnej bazy danych. Na ich podstawie generowane są tablice w języku SQL.
7. Baza danych – instancja bazy danych Sqlite znajdująca się na serwerze.

4.3. KLASYFIKATOR

Klasyfikator regresji logistycznej – pliki wyuczonego modelu klasyfikatora oraz vectorizera do zamiany tekstu na postać wektorów. Pliki te ładowane są przez moduł zarządzania anotacjami celem automatycznej anotacji wskazanej dyskusji w wymiarze polaryzacji opinii.

Na naiwny klasyfikator beysowski składają się skrypt uczący oraz klasa predykcyjna napisane w Pythonie. Skrypt uczący służy do tworzenia i testowania klasyfikatora, jako parametr przyjmuje plik tekstowy zawierający korpus językowy oraz flagę, która służy do zapisywania wynikowego modelu. Wyjściem ze skryptu są informacje na temat uczonego modelu w trybie normalnym lub model oraz wektoryzer w trybie zapisu.

Klasa predykcyjna przy tworzeniu instancji ładuje model i wektoryzer do klasy oraz implementuje metodę, która dla zadanego tekstu zwraca anotację zasugerowaną przez model.

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

5. BAZA DANYCH

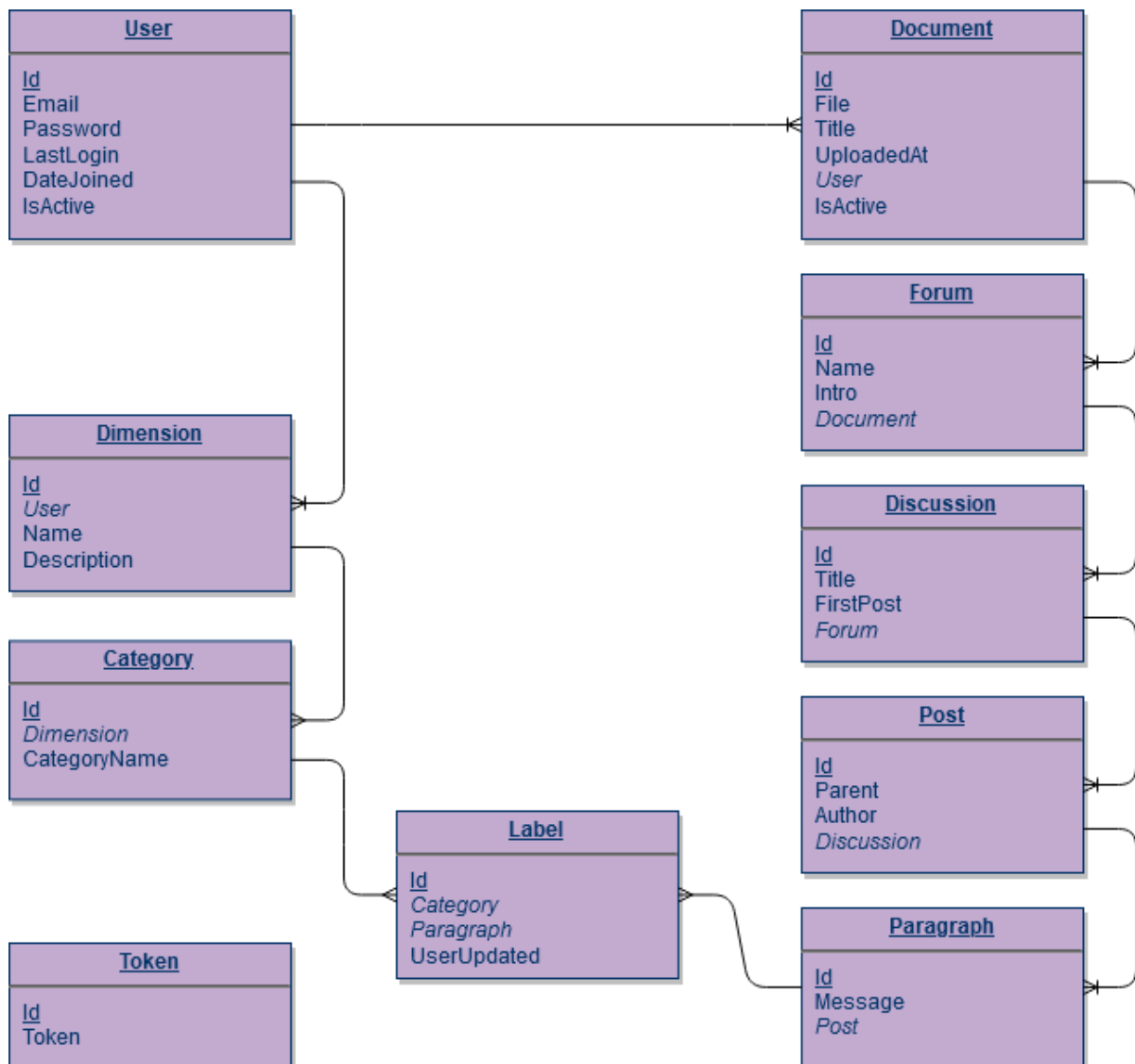


DIAGRAM 4 SCHEMAT BAZY DANYCH

- Tablica User – przechowuje informacje o zarejestrowanych użytkownikach
- Tablica Document – przechowuje informacje o przesłanych przez użytkowników plikach XML. Przypisuje każdy plik do jednego użytkownika.
- Tablica Forum – przechowuje informacje o forach zawartych w przesłanym pliku. Przypisuje każde forum do jednego pliku z tablicy Document.
- Tablica Discussion – przechowuje informacje o dyskusjach wyodrębnionych z forów przesłanych na serwer. Każda dyskusja jest przypisana do jednego forum, z którego pochodzi.

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

- Tablica Post – przechowuje informacje o postach użytkowników wyodrębnionych z przesłanego pliku XML. Każdy post jest przypisany do jednej dyskusji, z której pochodzi.
- Tablica Paragraph – przechowuje informacje o wypowiedziach użytkowników rozdzielone na paragrafy. Każdy paragraf jest przypisany do jednego postu użytkownika.
- Tablica Dimension – przechowuje informacje o wymiarach anotacji zdefiniowanych przez użytkownika. Każdy wymiar jest przypisany do jednego użytkownika, przez którego został stworzony.
- Tablica Category – przechowuje informacje o stworzonych przez użytkownika kategoriach anotacji w zadanym wymiarze. Każda kategoria jest przypisana do jednego wymiaru.
- Tablica Label – przechowuje informacje o anotacjach użytkownika. Jest tablicą pomocniczą relacji wiele do wielu, informując o wszystkich kategoriach anotacji przypisanych do danego paragrafu. Dodatkowo przechowuje informację, czy dana kategoria została wygenerowana przez klasyfikator, czy została ręcznie przypisana przez użytkownika.
- Tablica Token – przechowuje informacje o nieaktywnych tokenach uwierzytelniających wystawionych przez serwer.

6. WIDOKI

Widok logowania użytkownika

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

←

Rejestracja

Adres e-mail:

Hasło:

Powtórz hasło:

UTWÓRZ KONTO

[Posiadasz już konto? Zaloguj się](#)

Widok rejestracji użytkownika

Mój profil ⌵

Moje fora

documents/forum1_7d8HkYk.xml	×
documents/forum_rKl7aKA.xml	×
documents/forum1_E9ETbr.xml	×
documents/forum_gqjP9KM.xml	×
documents/forum_MmzFHC3.xml	×

Dodaj nowe forum

WYBIERZ PLIK **WYŚLIJ**

Moje wymiary

Schematy argumentacji	×
Polaryzacja opinii	×

DODAJ NOWY WYMIAR

Widok panelu głównego, widoczny po pomyślnym zalogowaniu. Podzielony jest on na komponent do zarządzania plikami w formacie XML z dostępnymi dyskusjami oraz na komponent do zarządzania wymiarami.

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

Przegląd wymiaru ⏻

[WRÓĆ DO STRONY GŁÓWNEJ](#)

Edytuj wymiar

Nazwa wymiaru

Opis wymiaru

ZAPISZ USUN WYMIAR

Kategorie

- hipoteza -
- rzeczowe -
- logiczne -
- emocjonalne -
- Inne -

[DODAJ NOWĄ KATEGORIĘ](#)

Widok edycji wybranego wymiaru. Możliwa jest zmiana nazwy, opisu, dodanie nowych bądź usunięcie wybranych kategorii oraz całkowite usunięcie wymiaru.

Wybór dyskusji ⏻

[WRÓĆ DO STRONY GŁÓWNEJ](#)

Wybierz dyskusję z *Najnowsze technologie*:

Najnowsze technologie - dyskusja

Czy komputer w lodówce to już współczesność?

Wymiary przypisane do pliku

Schematy argumentacji

Polaryzacja opinii

[DODAJ WYMIAR DO PLIKU](#)

[PRZYPISZ ANOTACJE KLASYFIKATORA POLARYZACJI OPINII](#)

[PRZYPISZ ANOTACJE KLASYFIKATORA SCHEMATU ARGUMENTACJI](#)

Widok przeglądu forum. W postaci listy widoczne są dostępne dyskusje. Z prawej strony widoczne są wymiary aktualnie przypisane do danego pliku. Poniżej opcji przypisania wymiaru do danego pliku, widoczne są przyciski odpowiedzialne za anotację dostępnych dyskusji z pomocą wbudowanego klasyfikatora.

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

Przegląd dyskusji

Dyskusje

ZŁADUJ DANE Z PLIKU USERS.XML Schematy argumentacji EKSPORTUJ DYSKUSJĘ WROĆ DO WYBORU DYSKUSJI

Najnowsze technologie - dyskusja

Czy komputer w lodówce to już współczesność?

Wizualizacje

Najnowsze technologie - dyskusja

Czy komputer w lodówce to już współczesność?

2

Kiedys żartobliwie mówiono, że niedługo komputer będzie już w każdej lodówce. Czy to "niedługo" jest już "dzisiaj"? Jakie zastosowania informatyki wdarły się do życia codziennego w taki sposób, że tego nie zauważamy?

Wybierz etykietę

hipoteza

7

Komputery w życiu codziennym to już norma. Mamy z nimi do czynienia w kuchni, łazience oraz salonie. Firmy technologiczne dążą do stworzenia inteligentnych domów, w zasadzie już takie domy istnieją, ale zwykły śmiertelnik może o nich tylko pomarzyć.

Zacznę więc może od wspomnianej lodówki. Na targach CES w 2011 roku firma LG zaprezentowała inteligentną lodówkę, która sama przeszukiwała Internet w poszukiwaniu ciekawych przepisów, bazując przy tym na produktach, które znajdowały się w jej wnętrzu. Sprawdzanie terminów ważności produktów również nie było problemem. Wszystkie niezbędne informacje, dotyczące lodówki i produktów z jej wnętrza, użytkownik mógł odebrać przy pomocy komputera, bądź urządzenia mobilnego, podłączonego do internetu.

Samsung w tym roku zaprezentowały lodówkę z wyświetlaczem dotykowym oraz możliwością dokonywania zakupów z poziomu lodówki.[\[url\]](http://www.youtube.com/watch?v=yu3keu48hz1)

Samsung i LG nie poprzestają tylko na konstruowaniu inteligentnych urządzeń AGD, chcą również zapewnić swobodną komunikację pomiędzy urządzeniami bez naszej ingerencji.

Dwa tygodnie temu miałem okazję obejrzeć film dokumentalny o inteligentnych domach. Prezentowane technologie mieściły się w centrum badawczym firmy Samsung, których oglądanie jest dostępne dla zwiedzających. Tak więc system rozpoznający domowników wybiera dla konkretnej osoby poziom oświetlenia w pomieszczeniu, załącza odpowiednią muzykę, itp.

Nie będę już wspominał o automatycznym nalewaniu wody do wanney o określonej godzinie, kabinach prysznicowych z telewizorem, łącznością telefoniczną i internetową czy podgrzewanych deskach sedesowych. Część tych rozwiązań jest już dostępna na rynku, a część będzie wdrażana w tym roku.

Zapewne część z Was słyszała o posiadłości Billa Gatesa, która jest w całości naszpikowana elektroniką. Jeśli nie to polecam poniższy artykuł:[\[url\]](http://mt.interia.pl/gadzetynews-jak-mieszka-bill-gates,nid.690078?url)

Wybierz etykietę

emocjonalne

Wybierz etykietę

rzeczowe

Wybierz etykietę

rzeczowe

Wybierz etykietę

rzeczowe

Wybierz etykietę

inne

Wybierz etykietę

inne

Wybierz etykietę

rzeczowe

Widok przeglądu dyskusji. Każdy post dyskusji jest odpowiednio podzielony na paragrafy, które są oznaczone edytowalną kategorią z wybranego wymiaru. Istnieje opcja natychmiastowej zmiany wymiaru poprzez wybranie go z listy. Możliwe jest również dodanie pliku *users.xml* z kopii zapasowej, który umożliwi przekształcenie numerów ID autorów danego postu na dane z forum Moodle. Do użycia jest też przycisk umożliwiający eksport danej dyskusji wraz z etykietami do pliku w formacie TSV. Z lewej strony widoczne są dostępne dyskusje z danego pliku oraz ich wizualizacje.

Wizualizacja dyskusji

Wiadomość:

Sam zgadzam się z tym, że należy postawić na nowe rozwiązania i nowoczesną technologię, która w wielu przypadkach jest równoznaczna z ułatwieniem życia programisty i zapewnieniu wielu komfortowych rozwiązań, które odciągają twórcę oprogramowania i pozwalają skupić się na innych aspektach tworzonego systemu, lecz nie jestem do końca przekonany. Głównie dlatego, że nowoczesna technologia musi się odpowiednio wdrożyć i osiągnąć pewien poziom stabilności. W ostatecznym rozrachunku uważam, że należy dobrać narzędzia adekwatnie do problemu, który chcemy rozwiązać.

```

graph TD
    30((30)) --- 31((31))
    30 --- 32((32))
    30 --- 35((35))
    32 --- 33((33))
    32 --- 34((34))
  
```

Widok wizualizacji dyskusji. Dyskusja zaprezentowana jest w formie drzewa, w którym

Komputerowa analiza dyskusji na forum		14.01.2021
Architektura		1.0
K. Boczoń	M. Armacki	M. Romaszkin

wierzchołki są postami w dyskusji i są odpowiedzią do rodzica danego wierzchołka. Gdy najdziemy na wybrany wierzchołek wyświetla się cała odpowiedź.

7. WYKORZYSTANE TECHNOLOGIE

Backend:

- Python
- Sqlite3
- Django
- 8. scikit-learn
- 9. nginx
- 10. UWSGI

Frontend:

- Angular
- NgRx
- Nebular
- xml2js
- D3.js