

ЗАТВЕРДЖЕНО  
1116130.00901-01-ЛЗ

## СИСТЕМА ДОСТУПУ ДО ЕНЦИКЛОПЕДИЧНИХ ЗНАНЬ НА ПРИРОДНІЙ МОВІ

Опис програми

1116130.00901-01 13 01  
Листів 19

## АНОТАЦІЯ

Документ 1116130.00901-01 13 01 «Система доступу до енциклопедичних знань на природній мові. Опис програми» входить до складу програмної документації до дипломного проекту.

Даний документ містить опис серверної частини системи, веб клієнта та їх функціональних можливостей. Серверна частина написана на мові програмування Python. Веб клієнт написаний на мовах JavaScript, HTML, CSS. Об'єм пам'яті, що займає програма комплексу та конфігураційні файли, складає 2 Мб. Конфігурація комп'ютера стандартна, на ньому повинно бути встановлено середовище розробки PyCharm 5.0.1 Community Edition.

Інв. № подл.	Підпис і дата				Замість інв. №	Інв. № дубл.	Підпис і дата	
Зм..	Лист	№ докум.	Підпис	Дата	1116130.00901-01 13 01			Аркуш
								2

## ЗМІСТ

1 Загальні відомості.....	4
2 Функціональне призначення.....	5
3 Опис логічної структури.....	6
3.1 Алгоритм роботи програми.....	6
3.2 Структура програми.....	7
3.3 Використані методи.....	11
3.4 Зв'язки програми з іншими програмами.....	11
4 Технічні засоби.....	12
5 Запуск та завантаження.....	13
6 Вхідні дані.....	14
7 Вихідні дані.....	15
8 Опис інтерфейсу користувача.....	16
9 Повідомлення.....	19

Інв. № подл.	Підпис і дата				Інв. № дубл.	Підпис і дата				Замість інв. №	Підпис і дата				Інв. № дубл.	Підпис і дата							
Зм...	Лист	№ докум.			Підпис	Дата	1116130.00901-01 13 01										Аркуш						
																	3						

## 1 ЗАГАЛЬНІ ВІДОМОСТІ

Найменування програмного засобу: «Система доступу до енциклопедичних знань на природній мові».

Для функціонування веб клієнту потрібен веб-браузер останньої версії (Google Chrome 50.00, Internet Explorer 11.0, Mozilla Firefox 46, Opera 37).

Для функціонування серверної частини потрібна операційна система Ubuntu 15 або Windows 7 (або вище) із встановленим інтерпретатором мови програмування Python 3.5.1.

Програму «Система доступу до енциклопедичних знань на природній мові» реалізовано у середовищі PyCharm Community Edition на мовах Python, JavaScript, HTML та CSS.

Інв. № подл.	Підпис і дата				Інв. № дубл.	Замксть. інв. №	Підпис і дата	
Зм..	Лист	№ докум.	Підпис	Дата	1116130.00901-01 13 01			Аркуш
								4

2 ФУНКЦІОНАЛЬНЕ ПРИЗНАЧЕННЯ

Функціональне призначення продукту полягає у наданні користувачу можливості за текстовим або голосовим запитом отримати інформацію, що міститься в енциклопедичних системах.

Інв. № подл.	Підпис і дата				Інв. № дубл.	Замість інв. №	Підпис і дата	Інв. № подл.	
Зм..	Лист	№ докум.	Підпис	Дата	1116130.00901-01 13 01				Аркуш
									5

## 3 ОПИС ЛОГІЧНОЇ СТРУКТУРИ

### 3.1 Алгоритм роботи програми

На рис. 3.1 приведено алгоритм роботи програми, що відображає послідовність дій при користуванні програмою.



Рисунок 3.1 – Алгоритм роботи програми

Підпис і дата	Інв. № дубл.	Замість інв. №	Підпис і дата	Інв. № подл.

Зм..	Лист	№ докум.	Підпис	Дата

1116130.00901-01 13 01

### 3.2 Структура програми

Для розробки програмного продукту був обраний метод об'єктно-орієнтованого проектування. Центральне місце в об'єктно-орієнтованому підході до проектування займає розробка діаграм прецедентів, компонентів та діаграм класів.

Система, що проектується, представлена у вигляді діаграми прецедентів (варіантів використання). Користувач представлений у вигляді одного актора (Actor), взаємодія з системою представлена за допомогою варіантів використання (Use Case). У програмі «Система доступу до енциклопедичних знань на природній мові» у якості актора виступатиме кінцевий користувач системи. Варіанти використання надають опис можливостей, які система надає актору.

На діаграмах використан тип відношення між варіантами використання та актором “асоціація”, що відображається лінією зі стрілкою між актором і варіантом використання.

Кінцевий користувач системи може виконати наступні варіанти використання (рис. 3.2):

- задання питання ГОЛОСОМ;
- задання питання ТЕКСТОМ;
- задання питання за запропонованим прикладом;
- оцінка правильності відповіді.

Структуру програми клієнтської частини можна представити за допомогою діаграми компонентів. Діаграма компонентів дозволяє визначити архітектуру системи, що розробляється, встановивши залежності між програмними компонентами. У багатьох середовищах розробки модуль або компонент відповідає файлу. Пунктирні стрілки, що сполучають модулі, показують стосунки взаємозалежності, аналогічні тим, які мають місце при компіляції вихідних текстів програм.

Інв. № подл.	Підпис і дата	Замість інв. №	Інв. № дубл.	Підпис і дата	<p>використання (рис. 3.2):</p> <ul style="list-style-type: none"> <li>– задання питання голосом;</li> <li>– задання питання текстом;</li> <li>– задання питання за запропонованим прикладом;</li> <li>– оцінка правильності відповіді.</li> </ul> <p>Структуру програми клієнтської частини можна представити за допомогою діаграми компонентів. Діаграма компонентів дозволяє визначити архітектуру системи, що розробляється, встановивши залежності між програмними компонентами. У багатьох середовищах розробки модуль або компонент відповідає файлу. Пунктирні стрілки, що сполучають модулі, показують стосунки взаємозалежності, аналогічні тим, які мають місце при компіляції вихідних текстів програм.</p>
Зм..	Лист	№ докум.	Підпис	Дата	<div>1116130.00901-01 13 01</div> <div>Аркуш 7</div>

Була розроблена діаграма компонентів для опису клієнтської частини програми «Система доступу до енциклопедичних знань на природній мові» (рис. 3.3).



Рисунок 3.2 – Діаграма варіантів використання програми кінцевим користувачем

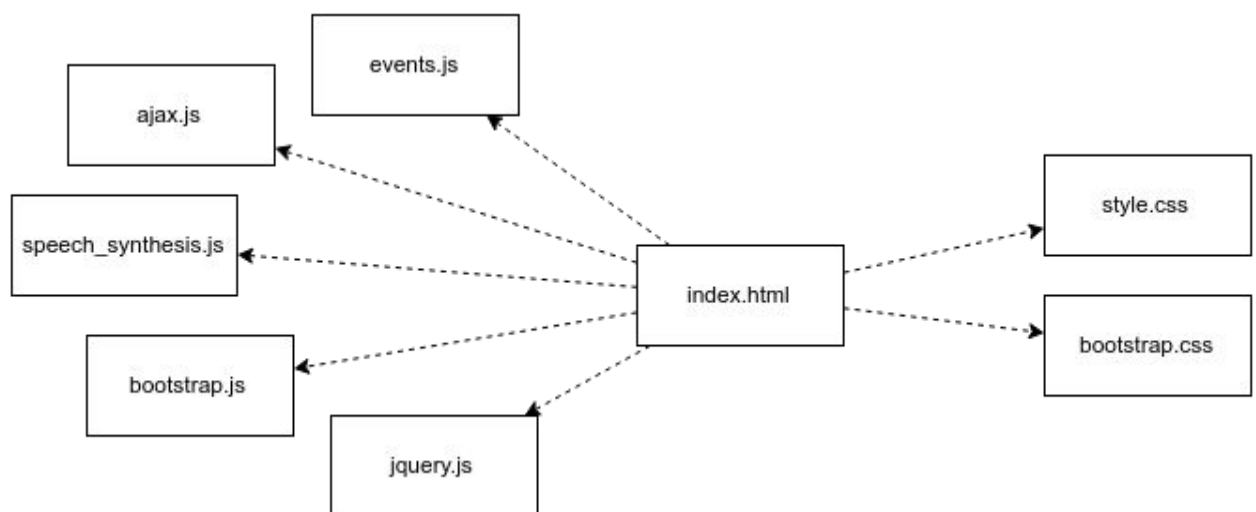


Рисунок 3.3 – Діаграма компонентів клієнтської частини програми

Детальніше розглянемо кожний модуль:

- flask\_app/templates/index.html (модуль, що містить розмітку головної сторінки);
- flask\_app/static/css/style.css (модуль, що містить стилі до головної сторінки);

Підпис і дата	
Інв. № дубл.	
Замість інв. №	
Підпис і дата	
Інв. № подл.	

Зм..	Лист	№ докум.	Підпис	Дата

1116130.00901-01 13 01

Аркуш

8



- flask\_app/static/js/ajax.js (модуль, що містить асинхронні запити до сервера);
- flask\_app/static/js/events.js (модуль, що містить логіку обробки подій, викликаних користувачем);
- flask\_app/static/js/speech\_synthesis.js (модуль, що містить логіку озвучення текстової відповіді).

Структуру програми серверної частини можна представити за допомогою діаграми класів. Діаграма класів служить для представлення статичної структури моделі системи в термінології класів об'єктно-орієнтованого програмування. На цій діаграмі показують класи, інтерфейси, об'єкти й кооперації, а також їхні відносини.

На рис. 3.5 показано діаграму класів серверної частини програми.

Детальніше розглянемо кожний клас:

- QuestionCategorizer - клас, що забезпечує віднесення питання до однієї із категорій (питання щодо опису сутності, питання щодо властивості сутності, неправильно поставлене питання);
- Question - абстрактний клас, що описує загальні властивості питання;
- DescribeQuestion - клас, що відповідає за питання щодо опису сутності;
- PropertyQuestion - клас, що відповідає за питання щодо властивості сутності;
- WrongQuestion - клас, що відповідає за неправильно поставлене питання;
- Entity - клас сутності;
- Property - клас властивості сутності;
- DB - клас, що відповідає за зв'язок з базою даних;
- DBPediaKnowledgeBase - клас, що відповідає за зв'язок з базою знань DBPedia;
- QATokenizer - клас, що забезпечує розбір тексту на слова;
- PatternMatcher - клас, що забезпечує співставлення тексту із шаблоном;
- SubjectFinder - клас, що забезпечує пошук головної сутності в питанні;

Підпис і дата		Інв. № дубл.		Замість інв. №		Підпис і дата		Інв. № подл.	
Зм..	Лист	№ докум.	Підпис	Дата	1116130.00901-01 13 01				Аркуш 9

- EntityNotFoundError - клас винятку, що описує ситуацію, коли задана сутність не знайдена;
- LowAnswerConfidenceError - клас винятку, що описує ситуацію, коли відповідь має низьку ймовірність бути правильною;
- EmptyPropertyDescriptionsError - клас винятку, що описує ситуацію, коли задана сутність не має властивостей;
- UnknownQuestionTypeError - клас винятку, що описує ситуацію, коли задане питання не можна віднести до жодної з існуючих категорій.

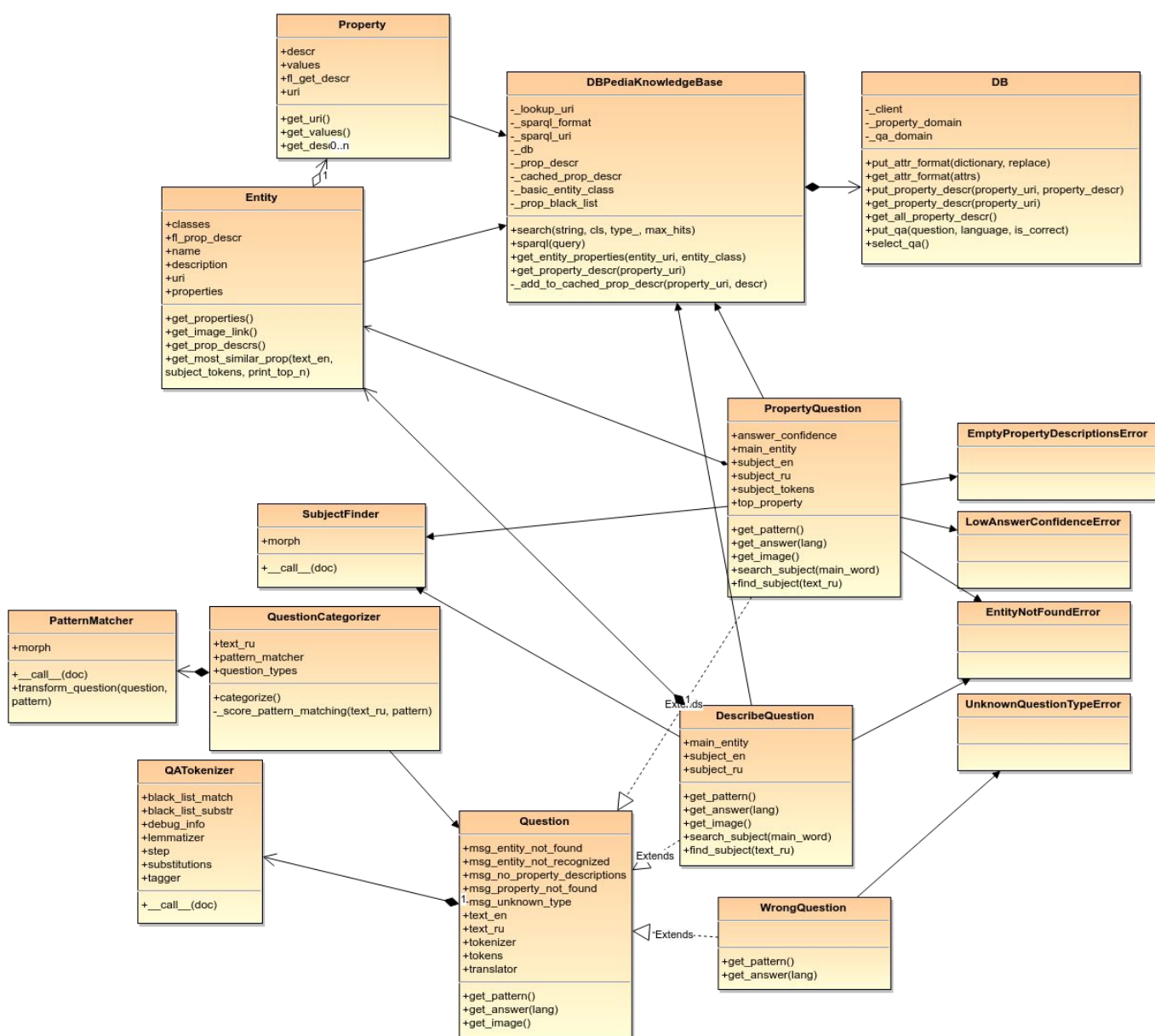


Рисунок 3.5 – Діаграма класів серверної частини програми

Інв. № дубл.	Підпис і дата
Замість інв. №	Підпис і дата
Інв. № подл.	

Зм..	Лист	№ докум.	Підпис	Дата
------	------	----------	--------	------

1116130.00901-01 13 01

### 3.3 Використані методи

У ході розробки були використані наступні методи:

- кешування результатів виконання функцій за допомогою декораторів;
- проектування веб-інтерфейсу, який працює без перезавантажень сторінки;
- асинхронні запити до серверу за допомогою технології AJAX.

### 3.4 Зв'язки програми з іншими програмами

Програма використовує наступні веб-сервіси:

- DBpedia Lookup API, що надає доступ до пошукової системи;
- DBpedia SPARQL, що надає доступ до бази знань через SPARQL-запити;
- Microsoft Translator API, що надає можливість перекладу тексту на будь-які мови.

Також у програмі використовується бібліотека boto3, яка надає доступ до Amazon Web Services, а саме до хмарної бази даних SimpleDB, де зберігається певна інформація з бази знань для пришвидшення роботи програми та статистика щодо правильності отриманих користувачами відповідей.

Інв. № подл.	Підпис і дата				Інв. № дубл.	Замість інв. №	Підпис і дата	що до правильності отриманих користувачами відповідей.
</								

## 4 ТЕХНІЧНІ ЗАСОБИ

Для коректного функціонування програмного продукту вимагається наявність ЕОМ під управлінням ОС 64-bit Ubuntu 14.04+, Windows 7+ або OS X Mavericks 10.9+, що має наступні мінімальні системні вимоги:

- процесор з тактовою частотою 2 ГГц або вищий;
- оперативна пам'ять не менш ніж 1 Гб;
- вільне місце на диску від 1 Гб;
- монітор з роздільною здатністю 1024\*768 або більшою;
- маніпулятор типу «миша»;
- клавіатура;
- мікрофон;
- можливість підключення до мережі Internet.

Інв. № подл.						Підпис і дата	Замксть.інв.№	Інв.№ дубл.	Підпис і дата	
Зм..	Лист	№ докум.	Підпис	Дата	1116130.00901-01 13 01					Аркуш
										12

## 5 ЗАПУСК ТА ЗАВАНТАЖЕННЯ

Оскільки програма виконана у формі веб-додатку, користувач здійснюватиме її запуск через веб-браузер указавши в адресному рядку певну адресу веб-сервера на якому буде розгорнуто цей веб-клієнт (localhost:5000, якщо додаток розгорнуто на локальному HTTP-сервері).

Після того як сервер обробить запит, до браузера буде завантажено HTML, CSS та JavaScript файли, які будуть відображені браузером у вигляді веб-інтерфейсу.

Програма має єдину точку входу (головна сторінка сайту).

					Підпис і дата		Інв. № дубл.		Замксть. інв. №		Підпис і дата		Інв. № подл.		
					1116130.00901-01 13 01										Аркуш
Зм..	Лист	№ докум.	Підпис	Дата											13

## 6 ВХІДНІ ДАНІ

Вхідними даними є:

- запит у форматі тексту, що є назвою певної сутності або має структуру питального речення російською мовою, яке починається із питальної конструкції;
- аналогічний запит у форматі аудіозапису;
- зворотний зв’язок у вигляді оцінки правильності відповіді системи за бінарною шкалою (“так” або “ні”);
- інформація зі сторонніх енциклопедичних систем, а саме: короткий текстовий опис та картинка.

Інв. № подл.	<div> <div>Підпис і дата</div> <div>Інв. № дубл.</div> <div>Замість інв. №</div> <div>Підпис і дата</div> </div>				<div> <div>1116130.00901-01 13 01</div> <div>Аркуш</div> <div>14</div> </div>
Зм...	Лист	№ докум.	Підпис	Дата	

## 7 ВИХІДНІ ДАНІ

Вихідними даними є:

- відображення голосового запиту у текстовому вигляді;
- відповідь на запит у вигляді тексту;
- рисунок, який доповнює відповідь на запит;
- озвучення тексту відповіді на запит;
- статистика результатів оцінювання якості відповідей, що містить інформацію про відсоток правильних відповідей та загальну кількість оцінок;
- запит пошуку сутностей до енциклопедичних систем;
- запит до енциклопедичних систем у форматі SPARQL.

Інв. № подл.					Підпис і дата		Замкст. інв. №	Інв. № дубл.	Підпис і дата
Зм..	Лист	№ докум.	Підпис	Дата	1116130.00901-01 13 01				Аркуш
									15

8 ОПИС ІНТЕРФЕЙСУ КОРИСТУВАЧА

На початку роботи користувач заходить на стартову сторінку системи, вигляд якої наведено на рис. 8.1.

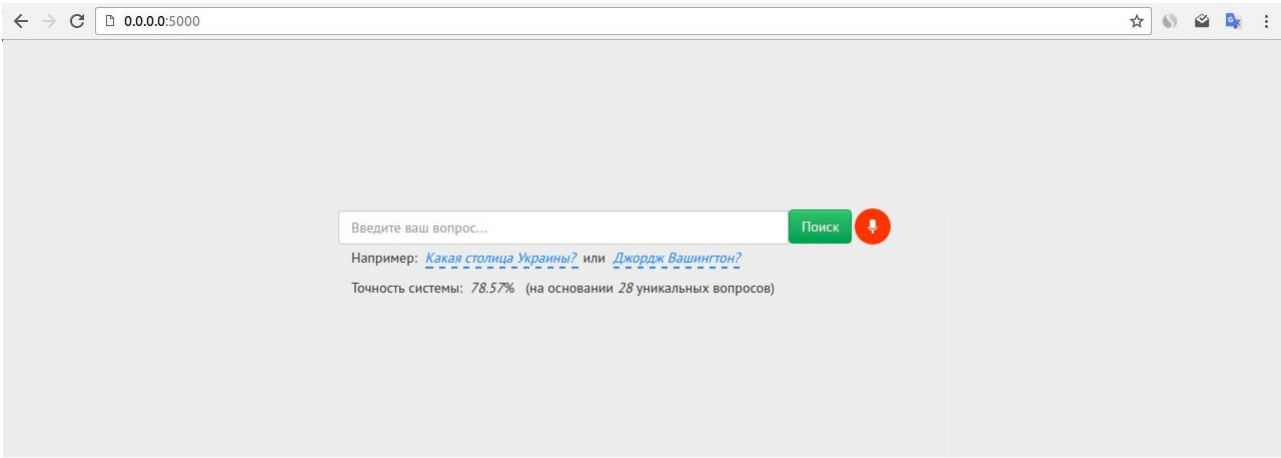


Рисунок 8.1 – Стартова сторінка

Для того, щоб задати запитання голосом, користувачу потрібно натиснути на іконку мікрофона. Після цього з’явиться вікно із запитом про використання браузером мікрофону (рис. 8.2).

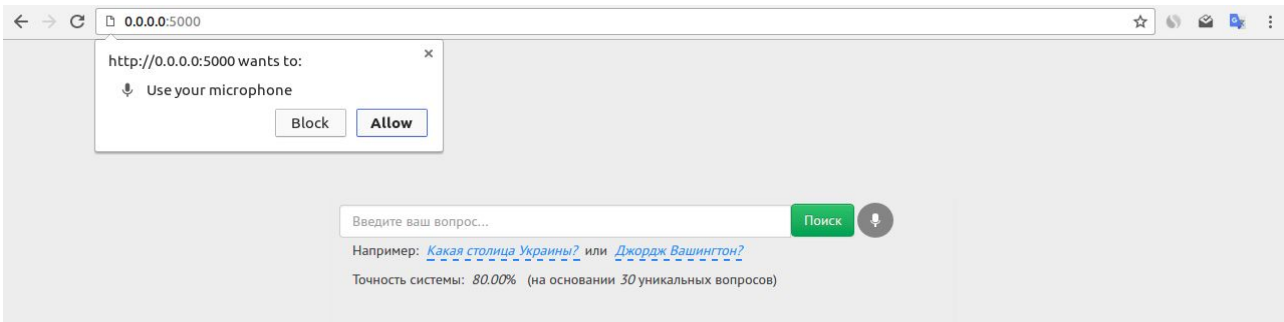


Рисунок 8.2 – Запит на використання мікрофону

Після задання питання голосом або текстом, перед користувачем з’явиться анімація очікування відповіді (рис. 8.3). Після певного часу очікування з’явиться відповідь з фреймом оцінки її правильності (рис. 8.4.). Коли користувач оцінить правильність відповіді, спочатку до бази даних надійде запит на запис цієї оцінки, а потім фрейм зникне з інтерфейсу користувача (рис. 8.5).

Інв. № подл.	Підпис і дата	Замість інв. №	Інв. № дубл.	Підпис і дата	1116130.00901-01 13 01					Аркуш 16
					Зм..	Лист	№ докум.	Підпис	Дата	



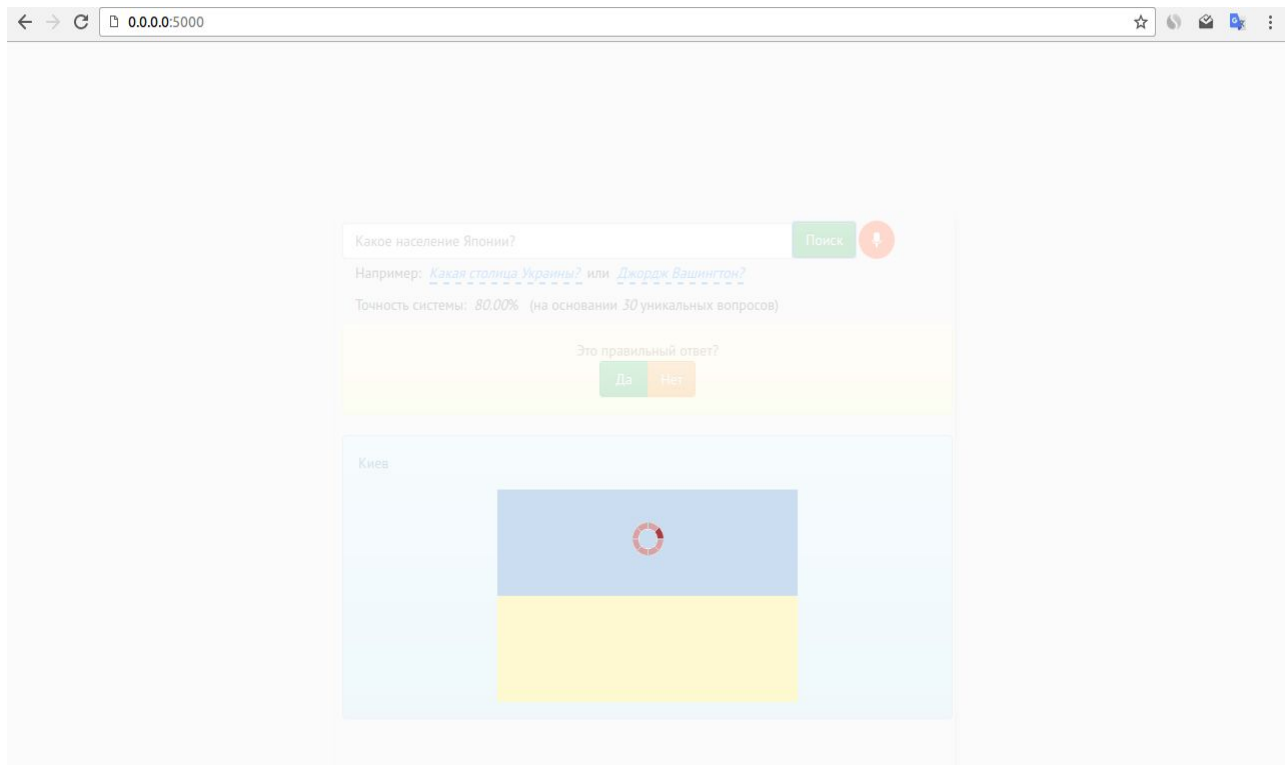


Рисунок 8.3 – Очікування відповіді

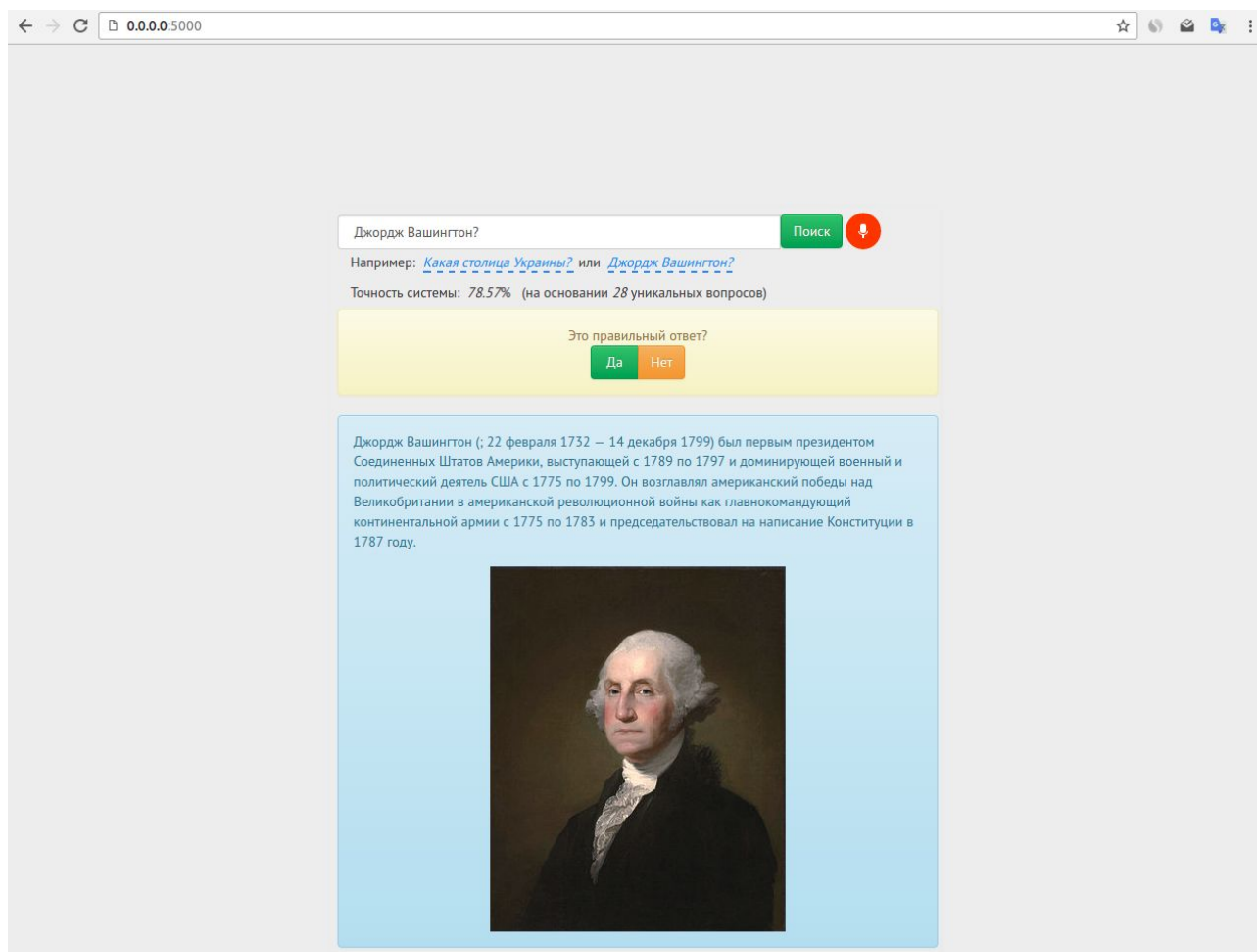


Рисунок 8.4 – Відповідь на питання з фреймом оцінки правильності відповіді

Підпис і дата	
Інв. № дубл.	
Замість інв. №	
Підпис і дата	
Інв. № подл.	

Зм..	Лист	№ докум.	Підпис	Дата
------	------	----------	--------	------

1116130.00901-01 13 01

Аркуш

17

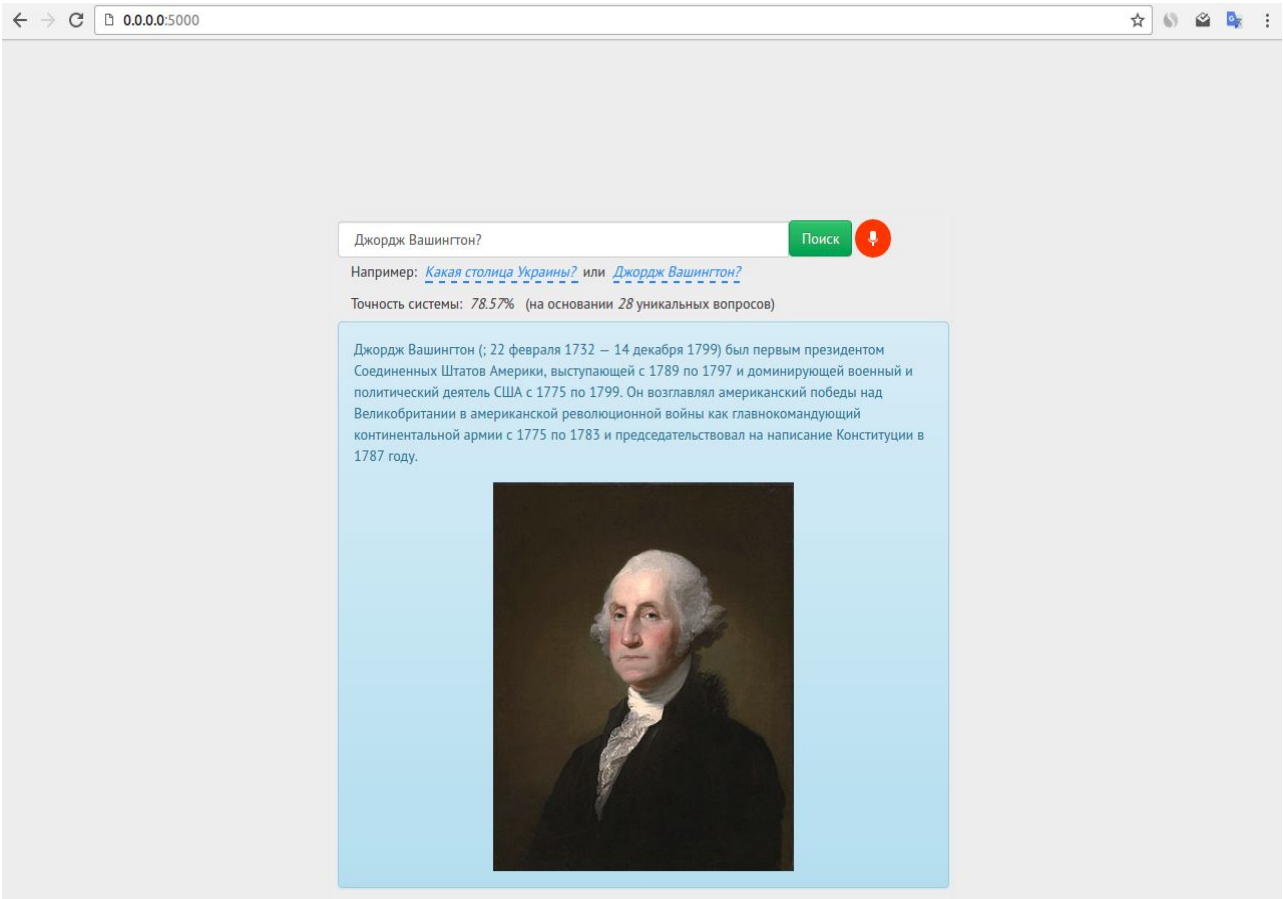


Рисунок 8.5 – Відповідь на питання без фрейму оцінки правильності відповіді

Підпис і дата					1116130.00901-01 13 01	Аркуш
Інв. № дубл.						
Замість інв. №						
Підпис і дата					1116130.00901-01 13 01	Аркуш
Інв. № подл.						
Зм..	Лист	№ докум.	Підпис	Дата		

## 9 ПОВІДОМЛЕННЯ

Повідомлення, які можуть виникнути під час користування програмою, наведені в табл. 9.1.

Таблиця 9.1 – Повідомлення користувачу

Текст повідомлення	Опис ситуації	Рекомендовані дії
«Указанная сущность не была найдена в базе. Пожалуйста, перефразируйте вопрос!»	Вказана користувачем сутність не була знайдена в базі знань.	Перефразувати питання.
«Указанная сущность не была распознана в вопросе. Пожалуйста, перефразируйте вопрос!»	Вказана користувачем сутність не була розпізнана у запитанні.	Перефразувати питання.
«Указанное свойство сущности не было найдено. Пожалуйста, перефразируйте вопрос!»	Властивість вказаної сутності не була знайдена у базі знань.	Перефразувати питання.
«Тип вопроса не распознан. Спросите о какой-нибудь сущности либо её свойстве.»	Тип питання не розпізнано.	Сформулювати питання про певну сутність або її властивість.
«Указанная сущность не имеет свойств. Пожалуйста, задайте вопрос о другой сущности.»	Вказана сутність не має властивостей.	Задати питання про іншу сутність.

Підпис і дата	
Інв. № дубл.	
Замість інв. №	
Підпис і дата	
Інв. № подл.	

					1116130.00901-01 13 01	Аркуш
Зм..	Лист	№ докум.	Підпис	Дата		19