Міністерство освіти і науки України

Тернопільський НАЦІОНАЛЬНИЙ технічний Університет

імені Івана Пулюя

ФАКУЛЬТЕТ КОМП’ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ І ПРОГРАМНОЇ ІНЖЕНЕРІЇ

КАФЕДРА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

**кОРНЯТ МИКОЛАЙ РОМАНОВИЧ**

УДК 004.4

**ІНТЕГРАЦІЯ МАШИННОГО НАВЧАННЯ ДЛЯ ПОКРАЩЕННЯ РУЧНОГО ПЕРЕКЛАДУ В СЕРВІСІ “CROWDIN”**

121 «Інженерія програмного забезпечення»

**Автореферат**

дипломного проекту на здобуття освітнього ступеня «магістр»

Тернопіль

2018

|  |  |
| --- | --- |
| Проект виконано на кафедрі програмної інженерії Тернопільського національного технічного університету імені Івана Пулюя. | |
| **Керівник проекту:** | кандидат технічних наук, доцент кафедри програмної  інженерії  **Михалик Дмитро Михайлович**,  Тернопільський національний технічний університет  імені Івана Пулюя |

Захист відбудеться 20 лютого 2018 р. о 9.00 годині на засіданні екзаменаційної комісії №31 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 101

**ЗАГАЛЬНА ХАРАКТЕРИСТИКА ПРОЕКТУ**

**Актуальність теми проекту**. Необхідність покращення функціональності сервісу, збільшення швидкості перекладу.

**Мета проекту.** Побудова системи для інформаційної підтримки перекладачів в сервісі локалізації.

**Об’єкт, методи та джерела дослідження.** Система машинного навчання на основі існуючих даних і методи та засоби її реалізації.

**Практичне значення отриманих результатів.** Система використовується для покращення швидкості перекладу тексту та покращення зв’язності контексту у перекладах за рахунок використання існуючих перекладів.

**Структура проекту.** Проект складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 4 частин, висновків, переліку посилань. Обсяг проекту: розрахунково-пояснювальна записка – арк. формату А4, графічна частина – слайдів.

**ОСНОВНИЙ ЗМІСТ ПРОЕКТУ**

**У вступі** проведено огляд сучасних досягнень науки і техніки у сфері локалізації програмного продукту, описано загальну специфіку тематики та завдання розробки.

В розділі **«Розробка програмної системи»** описано предметну область та специфіку галузі. Описано аналіз вимог до інтегрованої системи. Проаналізовано існуючі розробки та їх переваги і недоліки. Коротко описано використані технології та чому вони були обрані в процесі розробки. Описано процес реалізації програмної системи та прийняті технічні рішення

В розділі **«Навчання та тестування програмної системи»**, розглянуті методи та технології за допомогою яких можна здійснювати навчання та тестування. Описано процеси та їх результати.

В розділі **«Обґрунтування економічної ефективності**» здійснено розрахунок витрат на оплату праці, норм часу на виконання дипломного проекту, витрат на електроенергію, суму амортизаційних відрахувань та ціну дослідження, економічну ефективність і термін окупності капітальних вкладень.

В розділі **«Охорона праці та безпека в надзвичайних ситуаціях»** розглянуто питання забезпечення електробезпеки користувачів ПК для покращення безпеки праці на підприємстві, розглянуті правила безпеки експлуатації електронно−обчислювальних машин та вплив виробничого середовища на працездатність та здоров’я користувачів комп’ютерів.

**У загальних висновках щодо дипломного проекту** описано результати розробки системи покращення перекладу для сервісу локалізації «Crowdin».

В графічній частині приведено результати розробки інтегрованої системи з машинним навчанням.

**ВИСНОВКИ**

В результаті виконаної роботи розроблено програмний продукт, що відповідає заявленим вимогам згідно поставленого проекту.

До переваг побудованої системи можна віднести модульність і розширюваність. Система побудована на мові Python, що дає значний приріст до продуктивності системи та розгортається за допомогою контейнера Docker, що дозволяє не звертати увагу на тип системи та виносити інтеграцію на інший сервіс в разі потреби, збільшувати цим продуктивність. Система містить простий та зручний інтерфейс взаємодії для обробки запитів і орієнтується на продуктивність.

До недоліків можна віднести те, що поточна реалізація потребує значної кількості даних для коректної роботи, і її використання не ефективне для малих проектів.

**АНОТАЦІЯ**

Актуальність теми роботи полягає в необхідності покращення функціональності сервісу, покращенні швидкості та якості перекладання проектів.

Об’єктом дослідження є система машинного навчання, побудована на основі алгоритму word2vec, та методи і засоби для її реалізації.

Метою роботи є побудова системи для інформаційної підтримки перекладачів в сервісі локалізації.

Система розроблена за допомогою мови Python з використанням Docker контейнера в якості основи. Система дозволяє покращити швидкість перекладу тексту, за рахунок підказок інтерфейсу використовуючи існуючі переклади, покращує зв’язність контексту у перекладах. Інтеграція створена як окремий мікросервіс, що надає свою гнучкість в використанні і не впливає на основну роботу сервісу Crowdin.

Ключові слова: PYTHON, WORD2VEC, RNN, LSTM, ІНТЕГРАЦІЯ, CROWDIN, ЛОКАЛІЗАЦІЯ ПРОДУКТУ.