

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Західноукраїнський національний університет  
Факультет комп'ютерних інформаційних технологій  
Кафедра комп'ютерної інженерії

Жаборинська Ірина Василівна

**Програмний додаток перекладу тексту на основі API-  
інтерфейса онлайн перекладача / Text translation software  
application based on the online translator API**

спеціальність: 123 – Комп'ютерна інженерія  
освітньо-професійна програма – Комп'ютерна інженерія

Кваліфікаційна робота

Виконав: студентка групи КІ-41  
Жаборинська Ірина Василівна

---

Науковий керівник  
К.т.н., Батько Ю.М.

Кваліфікаційну роботу

Допущено до захисту

«\_\_\_» \_\_\_\_\_ 20\_\_ р.

Завідувач кафедри

\_\_\_\_\_ О.М. Березький

ТЕРНОПІЛЬ - 2021

## РЕЗЮМЕ

Кваліфікаційна робота на тему «Програмний додаток перекладу тексту на основі API-інтерфейса онлайн перекладача» зі спеціальності 123 «Комп'ютерна інженерія» освітнього ступеня «бакалавр» містить 77 сторінок пояснюючої записки, 11 рисунків, 10 таблиць, 2 додатки. Обсяг графічного матеріалу 2 аркуші формату А3.

Метою кваліфікаційної роботи є розробка програмного додатку для перекладу текстової інформації API-інтерфейса онлайн перекладача.

Методи досліджень базуються на теорії алгоритмів (для аналізу розроблених методів та алгоритмів), алгоритмах семантичного аналізу (для обробки запитів користувачів), технологій об'єктно-орієнтованого програмування (для програмної реалізації спроектованої структури програмного додатку).

В кваліфікаційній роботі на основі аналізу існуючих алгоритмів обробки інформації та алгоритмів перекладу текстової інформації розроблений програмний додаток для перекладу тексту на основі API-інтерфейса онлайн перекладача.

Проведено тестування розробленого програмного додатку на мові програмування C#, що підтвердило доцільність використання існуючих цифрових бібліотек для обробки текстової інформації та розробки «клієнт-серверних» додатків.

Розроблений програмний продукт є ефективним засобом з простим інтерфейсом, що дозволяє вирішувати проблему отримання, аналізу та перекладу інформації в глобальній мережі інтернет на основі запитів користувачів.

Ключові слова: ІНТЕРНЕТ, СИСТЕМА ПЕРЕКЛАДУ, КОРИСТУВАЦЬКИЙ ЗАПИТ, СЕМАНТИЧНИЙ АНАЛІЗ.

## RESUME

Qualification thesis “Text translation application based on the online translator API” in the specialty 123 "Computer Engineering" "bachelor" education degree contains 77 pages of explanatory notes, 11 figures, 10 tables, 3 appendixes. The volume of graphic material is 2 sheets of A3 format.

The purpose of the qualification work is to develop a software application for translating text information of the online translator API.

Research methods are based on algorithm theories (for analysis of developed methods and algorithms), semantic analysis algorithms (for user record blocks), object-oriented programming technologies (for software implementation of the designed structure of the software application).

In the qualification work, based on the analysis of existing information processing algorithms and text information translation algorithms, a software application for text translation based on the online translation API was developed.

The developed software application was tested in the C # programming language, which confirms the feasibility of using existing digital libraries for text processing and development of "client-server" applications.

The developed software product is an effective tool with a simple interface that allows you to solve the problem of obtaining, analyzing and translating information on the global Internet based on user requests.

Keywords: INTERNET, TRANSLATION SYSTEM, USER REQUEST, SEMANTIC ANALYSIS.

## ЗМІСТ

Перелік умовних скорочень.....	9
Вступ.....	10
1 Машинний переклад текстових даних в системах обробки інформації.....	12
1.1 Технології перекладу текстів та їх класифікація .....	12
1.2 Машинний переклад текстової інформації.....	16
1.3 Програмні засоби аналізу та перекладу текстових даних .....	20
1.4 Постановка задач кваліфікаційної роботи та шляхи її вирішення .....	25
2. Алгоритми обробки текстових даних на основі статистичного аналізу .....	26
2.1 Алгоритми статистичного аналізу текстових документів .....	26
2.2 Методи автоматичного оцінювання якості машинного перекладу.....	30
2.3 Алгоритм перекладу документів на основі API онлайн перекладачів ..	35
3. Програмна система онлайн перекладу тексту .....	38
3.1 Структура автоматизованої системи перекладу тексту .....	38
3.2 Підсистеми додатку онлайн перекладу текстів.....	46
3.3 Тестування додатку онлайн перекладу текстів .....	49
4 Техніко-економічний розділ.....	52
4.1 Розрахунок витрат на розробку програмного додатку .....	52
4.2 Визначення експлуатаційних витрат .....	57
4.3 Визначення економічної ефективності та терміну окупності .....	61
Висновки.....	63
Список використаних джерел.....	64
Додаток А Вихідний текст головного класу програмного додатку .....	68
Додаток Б Довідка про впровадження результатів роботи.....	74
Додаток В Світлокопія виданої публікації .....	75

					КР.КІ.07097/17.00.00.000 ПЗ		
Змн.	Лист	№ докум.	Підпис	Дата			
Розробив		Жаборинська І.В.			Літ.	Арк.	Акрушів
Перевір.		Батько Ю.М.			8	77	
Консульт.					ЗУНУ, ФКІТ, КІ-41		
Н. Контр.		Мельник Г.М.					
Затвердив		Березький О.М.					

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

КС	–	Комп'ютерна система
БД	–	База даних
ЕС	–	Елементарне слово
СС	–	Слово синонім
ОЯП	–	Оцінка якості перекладу
МП	–	Машинний переклад

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

## ВСТУП

У світі існує понад 7000 мов, 4000 з яких написані. Проте лише 100 або близько того можна перекласти за допомогою автоматизованих інструментів, таких як Google Translate. Нові дослідження обіцяють нам також поспілкуватися з іншими. Якщо повідомлення на французькій або іспанській мовах, набравши його в машину автоматичного перекладу, миттєво можна отримати необхідний текст та можна надати адекватну відповідь на зрозумілій мові. Але багато інших мов все ще кидають виклик машинному перекладу, включаючи мови, якими розмовляють мільйони людей, такі як Волоф, Луганда, Тві та Еве в Африці. Це тому, що алгоритми, що працюють на цих ядрах, вивчають переклади людей - в ідеалі мільйони слів перекладеного тексту.

Такого матеріалу достатньо для таких мов, як англійська, французька, іспанська та німецька, завдяки багатомовним інституціям, таким як канадський парламент, ООН та Європейський Союз. Їхні людські перекладачі видають потоки перекладених стенограм та інших документів. Лише Європейський Парламент за десятиліття виробляє дані з 1,37 млрд слів 23 мовами.

Однак таких даних не існує для мов, якими можна широко говорити, але не настільки перекладено. Вони відомі як малоресурсні мови. Запасний матеріальний навчальний матеріал для цих мов складається з релігійних видань, зокрема перекладеної Біблії. Але це становить вузький набір даних і цього недостатньо для підготовки точних, широких роботів-перекладачів.

В даний час Google Translate пропонує можливість спілкуватися приблизно на 108 різних мовах, тоді як Bing Translator від Microsoft пропонує близько 70 мов. Проте по всьому світу існує понад 7000 розмовних мов і принаймні 4000 із системою письма.

Навчання людини-перекладача або аналітика розвідки новій мові може зайняти роки. Навіть тоді цього може бути недостатньо для виконання поставленого завдання. Наприклад, в Нігерії розмовляють понад 500 мовами це

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		10

призводить до того, що найвідоміші в світі експерти в цій країні можуть зрозуміти лише незначну частину цих, якщо такі є.

Щоб подолати цей бар'єр, проводяться широкі роботи та фінансуються дослідження для розробки системи, яка може знаходити, перекладати та узагальнювати інформацію з будь-якої мови з низьким рівнем ресурсів, будь то в тексті чи мові.

Ідея машинного перекладу природних мов вперше з'явилася у XVII столітті, але стала реальністю лише наприкінці XX століття. Сьогодні комп'ютерні програми широко використовуються для автоматизації процесу перекладу. Хоча в галузі машинного перекладу було досягнуто значного прогресу, повністю автоматизовані переклади далеко не ідеальні. Тим не менше, країни продовжують витратити мільйони доларів на різні програми автоматичного перекладу. На початку 1990-х уряд США спонсорував змагання серед систем машинного перекладу. Можливо, одним із цінних результатів цього підприємства був корпус чисельних оцінок якості машинного перекладу, виготовлених вручну, щодо набору довідкових перекладів. Розвиток систем машинного перекладу дало поштовх великій кількості досліджень, тим самим спонукаючи багатьох дослідників шукати надійні методи для автоматичної оцінки якості машинного перекладу.

Таким чином, задача розробки програмного забезпечення та аналіз алгоритмів перекладу тексту з однієї мови є актуальною, та поширеною, зокрема в областях, що характеризуються обробкою великої кількості інформації, що надходить на різних мовах світу.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						11
Змн.	Арк.	№ докум.	Підпис	Дата		

# 1 МАШИНИЙ ПЕРЕКЛАД ТЕКСТОВИХ ДАНИХ В СИСТЕМАХ ОБРОБКИ ІНФОРМАЦІЇ

## 1.1 Технології перекладу текстів та їх класифікація

Переклад - це послуга, яка змінює слова або текст з однієї мови на іншу. Глибоке розуміння обох використовуваних мов життєво важливо, щоб це було точно. Завдяки глобальній пов'язаності Інтернету цей процес популярний на веб-сайтах, у пошукових системах та навіть у соціальних мережах. Але системи перекладу та послуги в різних країнах відрізняються. Правильно перекладений текст або слова передаватимуть той самий зміст, емоції та наміри, що і оригінальне повідомлення. Він також враховуватиме будь-які потенційні відмінності в культурі та формулюванні двох мов. Люди іноді використовують терміни переклад та тлумачення як синоніми. Але перекладач працюватиме з письмовим текстом, а перекладач - усно. Зараз багато компаній пропонують автоматизовані послуги, які зроблять роботу за вас. Але якість перекладів може відрізнятися. Різні компанії будуть виконувати цю послугу за різними стандартами, і ціни залежать від цього. Наприклад, документи, що перекладаються та надсилаються до державних органів, можуть вимагати засвідченого перекладу. Письмовий твір, що перекладається для публікації, потенційно потребуватиме вищого стандарту перекладу, ніж твір, який використовується лише для інформації.

На сьогоднішній день використання машинного перекладу значно прискорює процес комунікації між різними людьми, адже можливість отримання перекладу в реальному часі цьому сприяє. Людям не потрібно читати навчальні записи за роки, щоб вивчити мову.

Дослідницькі групи використовують технологію нейронних мереж для вирішення проблеми – форми штучного інтелекту, що імітує деякі аспекти людського мислення. Моделі нейронних мереж революціонізували обробку мови в останні роки. Замість того, щоб просто запам'ятовувати слова та речення, вони можуть дізнатися їх значення. Вони можуть зрозуміти з контексту, що такі слова,

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дата		



як "собака", "пудель" та англійське "dog", виражають подібні поняття, навіть якщо вони зовні дуже різняться.

Однак для цього моделям зазвичай потрібно пройти мільйони сторінок навчального тексту. Завдання полягає в тому, щоб змусити їх вчитися на менших обсягах даних, як це роблять люди. Врешті-решт, людям не потрібно читати парламентські записи за роки, щоб вивчити мову.

Основні компоненти систем машинного перекладу – це автоматичний пошук, розпізнавання мови, переклад та узагальнення тексту, всі вони адаптовані до мов з низьким рівнем ресурсів. Одним з новітніх напрямів збирання баз даних текстової інформації є моніторинг тексту та мовлення з інтернету у вигляді статей новин, блогів та відео. Завдяки користувачам у всьому світі, які розміщують вміст рідними мовами, зростає маса онлайн-даних для багатьох мов з низьким рівнем ресурсів. Ці дані в інтернеті, як правило, одномовні, це означає, що сомалійські статті чи відео написані цією мовою і не мають паралельного перекладу.

Існує думка, що під час попередньої підготовки нейронні моделі вивчають певні структури та особливості людської мови загалом, які потім можна застосувати до завдання перекладу. Але після попередньої підготовки з багатьох мов нейронні моделі можуть навчитися перекладати між окремими мовами, використовуючи дуже мало двомовних навчальних матеріалів, відомих як паралельні дані. Досить кількох сотень тисяч слів паралельних даних – приблизно тривалість кількох романів.

Багатомовний пошуковий механізм зможе аналізувати людську мову, а також текст, що представляє ще один комплекс складних проблем. Наприклад, технологія розпізнавання мови та транскрипції, як правило, бореться зі звуками, іменами та місцями, з якими раніше не стикалася. Одним із рішень є повернення до слів, які спочатку переписувались із мірою невизначеності, вказуючи на те, що машина не була з ними знайома. Під час повторної перевірки одним із них може виявитись незрозуміле раніше, маловідоме ім'я політика.

Після пошуку та перекладу відповідної інформації пошукова система підсумовує її для користувача. Цей мовний бар'єр може створити проблему для

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

кожного, кому потрібно поспішати збирати точну глобальну інформацію, включаючи спецслужби.

Переклади можна класифікувати в основному на дві великі категорії перекладу: для бізнесу та не для бізнесу. Однак деякі типи перекладів начебто збігаються в обох категоріях. Дванадцять найважливіших типів перекладів, які виконуються в наш час перелічено на рисунку 1.1.

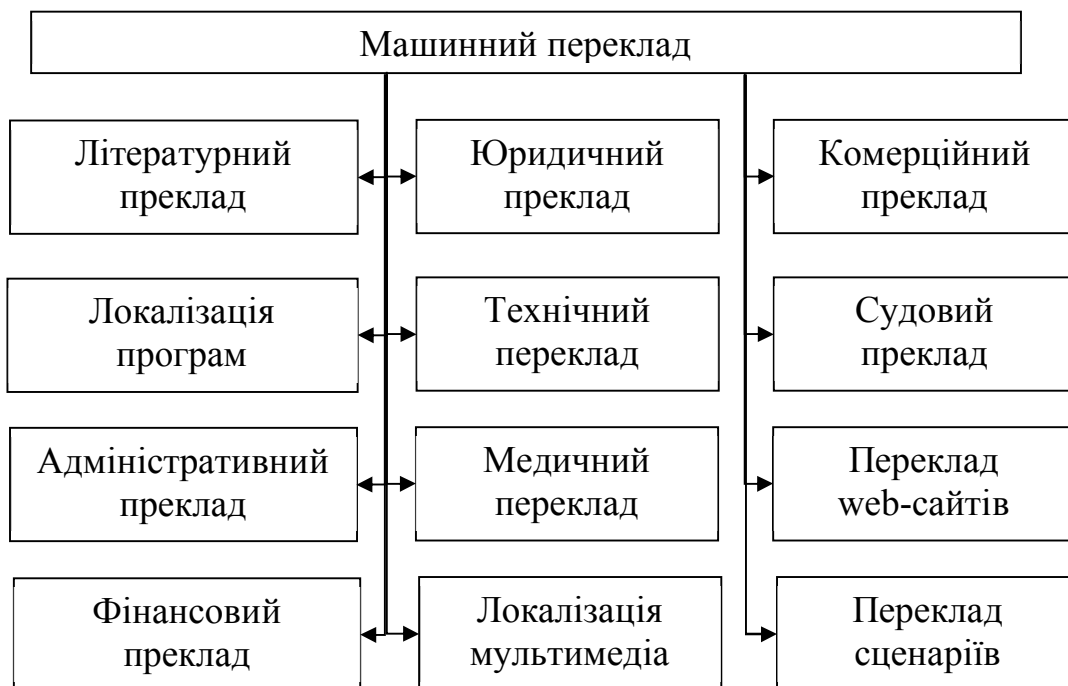


Рисунок 1.1 – Класифікація типів машинного перекладу

Літературний переклад стосується перекладу літературних творів, таких як оповідання, романи, вірші та п'єси. Його часто вважають найвищою формою перекладу, оскільки художній переклад – це набагато більше, ніж просто передача значення та контексту документа мовою джерела на мову перекладу. Він передбачає включення відповідних культурних нюансів, переклад гумору, почуттів, емоцій та інших тонких елементів певного твору.

При проведенні локалізації програм головним для перекладу є користувальницький інтерфейс, але на цьому локалізація програмного забезпечення не закінчується. Вона включає повідомлення про помилки, системні повідомлення, файли довідки тощо. При цьому слід зазначити, що зміст потрібно адаптувати відповідно до місцевої (цільової) культури.

Комерційний переклад – це різновид перекладу, який потребує типів перекладачів, які володіють спеціалізованими навичками, такими як знання ділового жаргону та галузі, до якої належить бізнес. Типи текстів при перекладі комерційних документів можуть включати ділову кореспонденцію, звіти, тендерну документацію, рахунки компаній, службові записки тощо. Іноді це може збігатися з юридичним перекладом, якщо компанія займається юридичними документами.

Юридичний переклад – один із найскладніших перекладів, що включає переклади свідоцтв про народження, про шлюб, переклади договорів, угод, договорів, меморандумів, заповітів тощо. Хороший перекладач повинен розуміти різні контексти, що лежать в основі документів, і контексти двох регіонів чи країн, для яких призначені документи, соціально-культурні аспекти, а також політико-правові аспекти. Потім їм потрібно буде перекласти його таким чином, щоб цільова аудиторія легко схопила текст.

Технічний переклад включає будь-який технічний вміст, який потребує перекладу – посібники користувача, монографії, онлайн-довідковий текст, буклети інструкцій, навчальні матеріали та відео, маркетингові матеріали для таких галузей, як виробництво, наука чи техніка – все це входить до сфери технічного перекладу.

Під судовим перекладом, мається на увазі діяльність з перекладу судових документів, таких як доповіді, протоколи засідань, показання експертів, показання свідків, судові рішення, судові листи, співбесіди та багато іншого – в основному, діяльність, пов'язана зі справами.

Адміністративний переклад – у сфері перекладу адміністративний відноситься до перекладу управлінських текстів, що використовуються в організаціях – будь то величезні корпорації чи регіональний бізнес. Хоча адміністративний переклад можна назвати підмножиною комерційного перекладу, усі комерційні переклади не обов'язково є адміністративними.

Медичний переклад включає будь-який медичний вміст, пов'язаний з пацієнтом, наприклад, етикетки, упаковки, інструкції чи програмне забезпечення, а вміст, пов'язаний з продуктом, як наукові роботи, документи з

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дата		

клінічних випробувань, сертифікати управління якістю тощо, як правило, потребує перекладу. Абсолютно важливо, щоб постачальники послуг перекладу мали досвід, мали необхідні знання та були професіоналами своєї країни.

Переклад веб-сайту – це копія веб-сайту, субтитри до відео на ваших веб-сторінках та будь-які документи. Тут також потрібно буде змінити речі, такі як валюти, формати адрес та макети, щоб звернутися до різних місцевих аудиторій.

Сьогодні переклад сценаріїв є життєво важливим, оскільки багато популярних фільмів і телешоу виходять з Голлівуду, дублюються на кілька мов і виходять у всьому світі.

Відео, графіка, анімація, GIF-файли, інфографіка – все це можна згрупувати в розділі локалізація мультимедіа і це дуже важливо сьогодні, оскільки все більше і більше компаній створюють мультимедійний вміст, щоб розширити охоплення своєї аудиторії та залучити їх. Локалізація цього вмісту може стати досить складною, хоча зовні може виглядати просто – оскільки вона повинна відповідати місцевій культурі та приваблювати клієнтів у цьому регіоні.

«Фінансовий переклад» - це фінансові документи, такі як банківські записи, виписки, виписки з рахунків тощо – які іноді потребують перекладу, щоб полегшити сприйняття цільовою аудиторією. Тут фактичний вміст, який потрібно перекласти, може бути меншим, але це повинно бути зроблено точно.

## 1.2 Машинний переклад текстової інформації

Хоча перекладачі з бюро перекладів або фрілансери були найбільш улюбленим та найпоширенішим способом здійснення перекладів, використання машин активно впроваджували в цю сферу впродовж останніх кількох років.

Машинний переклад має широке комерційне, військове та політичне застосування. Наприклад, все частіше до інтернету отримують доступ неанглійські мови, які читають неанглійські сторінки. Можливість чітко знаходити відповідну інформацію не повинна обмежуватися нашими мовними

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		16

можливостями. Крім того, ми можемо не мати достатньої кількості мовознавців якоюсь мовою, яка нас цікавить, щоб впоратися з величезним обсягом документів, які ми хотіли б перекласти. Введіть автоматичний переклад. Машинний переклад ставить ряд цікавих проблем машинного навчання: набори даних, як правило, дуже великі, як і відповідні моделі; навчальний матеріал часто буває галасливим та страждає скупою статистикою; простір пошуку можливих перекладів є достатньо великим, щоб вичерпний пошук був неможливим. Дослідження машинного навчання, такі як методи з максимальною маржею, часто з'являються в дослідженнях перекладу. Системи SMT на сьогодні вже достатньо зрілі, щоб їх можна було застосовувати у виробничих системах. Хорошим прикладом цього є онлайн-переклад арабсько-англійської мови від Google, який базується на методах SMT.

Використання машинних перекладачів зростає з моменту його впровадження кілька років тому. Це можна проілюструвати за популярністю Google Translate і подібних інструментів, таких як Ginger Translate, Skype Translator, Systran (спеціально розроблений для перекладу ділових комунікацій), Parago та інших. Google додав 13 нових мов у свій додаток для перекладу, а великі компанії інвестують мільйони у розробку програмного забезпечення для перекладу. Це робиться для більш дешевого перекладу, а також в виведення людей із процесу перекладу, що в довгостроковій перспективі виявиться економічно вигідним, проте це збільшує певні ризики.

Порівнюючи машинний переклад та людський переклад, з 1950-х років вчені експериментували із програмним забезпеченням, що виконує машинний переклад і з тих пір було зроблено багато вдосконалень, і сьогодні ця технологія набагато потужніша.

Тим не менше, навіть сьогодні, результати цих інструментів далекі від досконалості, що негативно впливає на їх розповсюдження в сучасних реаліях та зменшує відсоток їх використання в автоматизованих системах обробки текстових даних. Серед основних недоліків та ризиків використання програмних продуктів даного типу є як важливі так і не зовсім (рисунок 1.2).

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дата		



Рисунок 1.2 – Переваги та недоліки машинного перекладу

Неправильні інтерпретації. По-перше, одне слово може мати різні інтерпретації в залежності від ситуації або контексту. Наприклад, слово «замок»: «я побудував прекрасний замок» або «цей замок виглядає надійно» - тут слово «замок» має різне значення в кожному реченні. В першому варіанті це будівля, а в другому технічний пристрій. Машини чи програмне забезпечення ще недостатньо розумні, щоб розрізнити значення слів на основі контексту, і можуть забезпечити веселий переклад.

Тільки переклад від слова до слова. Знову ж таки, при виборі кращого перекладача між машинним або людським перекладом то слід врахувати, що ідіоми, тон, культурні посилання, на яких будується мова, що передаються повідомленням на сьогоднішній день не може бути зрозумілим жодною машиною чи програмним забезпеченням. На сьогодні машини можуть виконувати лише буквальний переклад або переклад від слова до слова.

Відмінності в побудові речень. Манера будівництва також відрізняється від мови до мови. Наприклад, французьке речення: "J'ai cinq chemises bleus",

якщо перекласти слово в слово, звучить так по-англійськи: "У мене є 5 сорочок блюзових", замість відповідного, "У мене є 5 блакитних сорочок". Це пов'язано з тим, що прикметник кольору ставиться після іменника французькою мовою, а також змінюється відповідно до числа. Тож програмне забезпечення може дати вам граматично неправильне речення.

Нерозуміння контексту. Досконале розуміння контексту, нюанси, тон необхідно для створення ідеального перекладу - що може зробити тільки людина. Так само в мові можуть бути слова, які означають щось неприємне чи незручне в іншій - і машина не розпізнає, це особливо часто трапляється у випадку імен.

Слова не знайдені. Ще одним недоліком цього методу перекладу є те, що коли машина не знаходить еквівалентний термін у цільовій мові, то вона зберігає вихідне слово (на мові джерела) – це означає, що взагалі не перекладає його, що не відбувається щоб бути дуже корисним.

Переваги машинного перекладу. Не можна сказати багато переваг, говорячи про 100% машинний переклад. Такі інструменти працюють набагато краще разом із людськими ресурсами в сукупності, яка називається машинним перекладом, але ось основні переваги суто машинного перекладу:

Простота перекладу типізованих проектів. Так, це корисно для формульної мови, проектів, що не мають вирішального значення, де потрібен переклад фрагментів інформації і де немає шансів змінити значення, як у згаданих вище прикладах.

Переклад малого тексту. Якщо необхідно просто отримати суть невеликого тексту, скажімо, як електронне повідомлення, або те, що хтось говорить мовою, яку ви взагалі не знаєте, або перекласти прості речення чи фрази – тоді так, машинний переклад може бути досить зручно.

Швидкість. І так, машинний переклад дуже швидкий – набагато швидший, ніж людський перекладач.

Можна сказати, що людський переклад має набагато більше переваг, ніж для машинний переклад. Серед головних перевагах людського перекладу можна навести такі:

Природний стиль і тон. Людина-перекладач, як правило, володіє

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дата		

принаймні двома мовами. Володіння носієм мови або вільне володіння двома мовами дає їм можливість перекладати з однієї мови на іншу таким чином, щоб перекладений текст здавався природним – як і вихідний текст.

Диктований переклад тексту. Люди також можуть перекладати диктовані роботи. Слухаючи тон мовця, і паузи, і гучність тощо, перекладач-людина може легко вловити настрій мови, контекст і зможе виявити справжнє значення та емоції того, що говорить призначений; щось машина просто не може сподіватися зробити.

Контекст та культурне розуміння. Не лише в продиктованих текстах, але і в письмових, перекладачі-люди можуть прекрасно розуміти контекст, нюанси та тон будь-якого письмового тексту. Вони також можуть адаптуватися до мови або діалекту

Перевірена робота. Знову ж таки, як правило, перекладач-людина, принаймні хороший та ефективний спеціаліст, перегляне свою роботу ще раз, щоб визначити, що в роботі, яку вони щойно закінчили, немає помилок або невідповідностей.

### 1.3 Програмні засоби аналізу та перекладу текстових даних

На сьогоднішній день в світі ІТ-технологій широко розповсюджені програмні додатки та цифрові бібліотеки функцій для аналізу та машинного перекладу. Класифікацію категорій засобів перекладу можна провести за такими критеріями:

Програмне забезпечення для пам'яті перекладів. Це, мабуть, найвідоміший та найуживаніший інструмент САТ. Програмне забезпечення розрізає документ, який слід перекласти на менші частини, які називаються "сегментами". По мірі того, як перекладач прогресує у перекладі документа, зберігає перекладений текст в окремій базі даних, до якої він може отримати доступ пізніше. У момент,

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		20



коли програмне забезпечення натрапляє на текст, подібний до частини, яка вже перекладена, воно рекомендує повторно використовувати попередній переклад.

Деякі програми працюють дещо по-іншому, як, наприклад, із довідковими документами, завантаженими раніше. Прикладами програмного забезпечення є Trados Workbench, Similis, Star Transit, SDLX тощо.

Програмне забезпечення пошукової системи мови. Ці пошукові системи працюють майже так само, як Google, Bing або Yahoo; лише замість того, щоб шукати у всесвітній мережі, вони перебирають величезні бази даних пам'яті перекладів. Основною функцією цих пошукових систем є пошук фрагментів збережених текстів, які раніше були перекладені, які відповідають новому тексту, що перекладається. Linguee - приклад цього інструменту.

Програмне забезпечення для управління термінологією. Засоби перекладу програмного забезпечення для управління термінологією. Це дозволяє перекладачу здійснювати автоматичний пошук термінів у певному документі в базі даних, що зберігає переклади. Деякі інструменти онлайн-перекладу дозволяють додавати нові відповідні слова та терміни та використовують різні методи перевірки.

Пізніше перекладач може здійснити перевірку та визначити, чи були терміни точно перекладені чи ні, і чи термін, який неодноразово з'являється, був перекладений послідовно протягом усього документа. LogiTerm, SDL MultiTerm тощо.

Програмне забезпечення для перевірки. Програмні засоби перевірки дозволяють перекладачу створювати пам'ять перекладів, використовуючи як джерело, так і ціль одного і того ж документа. Інструмент розрізає документи на сегменти та намагається визначити сегменти, які відповідають один одному.

Результати цих операцій можна імпортувати в інший інструмент перекладу, щоб він міг бути використаний для перекладів у майбутньому. YouAlign, Bitext2 та LF Aligner – це приклади програм для перевірки.

Інтерактивний машинний переклад. Це щось на зразок програмного програмного забезпечення на смартфонах, воно аналізує текст який вводиться користувачем, а далі надає варіанти перекладу в залежності від попередньої

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		21

інформації. Окрім цих конкретних інструментів онлайн-перекладу, в додаток до них надають програмні засоби з додатковими функціями:

- перевірка правопису для коректури;
- програмне забезпечення для перевірки граматики;
- інтернет-словники та бази даних термінології;
- індексація та засоби повнотекстового пошуку: що дозволяють шукати раніше перекладені тексти та довідкові документи;
- програмне забезпечення, що узгоджується/узгоджується: довідкові інструменти, які допомагають шукати слово разом із його контекстом.

Було розглянуто декілька найбільш відомих програм для роботи з зображеннями у вигляді цифрових файлів різного типу (таблиця 1.1).

Таблиця 1.1 – Порівняння програмних засобів перекладу тексту

Назва системи	Доступність мов	Перевірка орфографії	Додаткові словники	Вартість
Linguee	100+	+	+	Платний продукт
The Free Dictionary	50+	+	+	Платний продукт
ProZ	45-70	+	+	Платний продукт
SDL Trados	80+	+	+	Платний продукт
MemoQ	75	+	+	Платний продукт
WordFast Pro	50	+	+	Платний продукт
Fluency Now	80+	+	+	Умовно платний
Zanata	60+	+	+	Умовно платний
YouAlign	70	+	+	Freeware
Memsources	60	+	+	Freeware

Linguee. Один із найпопулярніших інструментів на сьогоднішній день на ринку, це унікальний інструмент, який є поєднанням словника та пошукової системи. Це означає, що ви можете легко шукати двомовний текст - терміни та

фрази різними мовами, перевіряти значення та їх контекстні переклади. Він також здійснює веб-пошук раніше перекладених документів, які мають значення, і відображає спосіб перекладу певного слова скрізь в Інтернеті. Перекладачі та вивчачі мови часто використовують цей інструмент перекладу разом із Google Images, щоб полегшити роботу.

The Free Dictionary. Це всебічний інструмент онлайн-перекладу: поєднання енциклопедії, тезаурусу та словника. Більше того, він доступний кількома мовами, такими як англійська, португальська, французька, іспанська та японська. На додаток до мовних нюансів, таких як скорочення, цитати та ідіоми, цей інструмент має багаті медичні, юридичні та фінансові терміни. Кілька розділів енциклопедії регулярно оновлюються і пропонують користувачам щоденні акції: статтю чи слово. Додаток Безкоштовний словник доступний для Android, а також мобільних пристроїв iOS.

ProZ. Замість програми, це портал, який з'єднує перекладачів та лінгвістів та дозволяє їм обмінюватися знаннями про термінологію, співпрацювати над перекладами, навчатись, ділитися словниками тощо. Ця інтерактивна платформа для краудсорсингу має форуми для обговорення, до яких перекладачі можуть приєднатися та шукати відповіді на свої запитання.

Студія SDL Trados. Цей інструмент перекладу настійно рекомендується професійними перекладачами. Він поєднує термінологію, машинний переклад, ТМ та локалізацію програмного забезпечення. Це потужний інструмент, який може бути дуже корисним для збільшення вашої клієнтської бази. Існує безкоштовна 30-денна демо-версія, після якої вам потрібно заплатити, щоб користуватися нею.

МетоQ. Це ідеальний варіант для фрілансерів, це один з найкращих інструментів онлайн-перекладу, який має кілька потужних функцій, що дозволяють просто повторно використовувати попередні переклади. Ви також можете використовувати МетоQ для перевірки правопису, перевірки використання правильних термінів, забезпечення узгодженості та покращення загальної якості. Хоча це платне рішення, існує 45-денна демо-версія, яку ви можете перевірити спочатку.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						23
Змн.	Арк.	№ докум.	Підпис	Дата		

WordFast Pro. Це багатоплатформене програмне забезпечення ТМ допомагає покращити загальний процес перекладу. Навіть із демо-версією, перекладачі можуть експортувати та імпортувати ТМ та використовувати їх без обмежень за часом - за умови обсягу пам'яті 500 одиниць.

Fluency Now. Цей преміальний інструмент в основному є програмним забезпеченням ТМ і популярний серед перекладачів-фрілансерів. Він сумісний з Windows, Linux і Mac і коштує 9,95 доларів на місяць. У версії Enterprise є ще кілька недоліків, таких як модуль управління проектами, який називається Fluency Flow. Статистика документів та вбудована функція коректури - інші функції, доступні в цьому інструменті перекладу.

Zanata. Це веб-система для управління проектами локалізації та корисна для творців вмісту та перекладачів. Він дбає про повний робочий процес, що дозволяє перекладачам зосередитись на своїй основній роботі перекладу, а не турбуватися про формати та інше. Zanata має ТМ, яка шукає та рекомендує найкращі збіги для перекладу в усій системі, а її редактор сумісний майже з усіма популярними веб-браузерами. Найкраще те, що декілька перекладачів можуть працювати над редактором та спілкуватися в режимі реального часу в чатах.

YouAlign. Цей інструмент програмного забезпечення для вирівнювання в Інтернеті безкоштовний для використання. Це дозволяє швидко та легко створювати вирівнювання з заархівованих документів на рівні речення. Це дозволяє завантажувати вирівнювання в HTML, якщо ви хочете публікувати в Інтернеті або шукати терміни та фрази за допомогою повнотекстової пошукової системи. Ви також можете завантажити вирівнювання у форматі TMX, щоб додати до вашого ТМ. За допомогою цього інструменту можна обробляти понад 100 форматів файлів.

Memsource. Ця хмарна база даних ТМ інтегрована з Gengo і забезпечує ефективний та дієвий інструмент управління проектами. Редагування та управління термінологією - це інші операції, доступні перекладачам для підвищення їх загальної продуктивності. Редактор Memsource - це настільна версія, яку можна завантажити безкоштовно.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		24

#### 1.4 Постановка задач кваліфікаційної роботи та шляхи її вирішення

У цьому розділі аналізуються сучасні методи опису та класифікації інформації. Розглянуто підходи до проведення перекладів текстових масивів даних на різні мови світу. Проведено порівняльний аналіз сучасних програмних засобів для аналізу та обробки великих масивів текстової інформації.

Для досягнення цієї мети необхідно вирішити наступні завдання:

- 1) провести аналіз та класифікацію інформації, канали передачі та її обробки;
- 2) проаналізувати підходи до перекладу текстових масивів даних на різні мови світу;
- 3) провести аналітичний огляд існуючих програмних систем для обробки та перекладу текстової інформації;
- 4) дослідити алгоритми статичного поділу текстової інформації на окремі складові;
- 5) розробити алгоритм перекладу текстової інформації за допомогою API-інтерфейса онлайн перекладача;
- 6) розробити блок-схему програмного додатку для опису та перекладу текстових даних, програмно реалізувати запропоновані алгоритми, провести порівняльний аналіз з програмами-аналогами.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		25

## 2. АЛГОРИТМИ ОБРОБКИ ТЕКСТОВИХ ДАНИХ НА ОСНОВІ СТАТИСТИЧНОГО АНАЛІЗУ

### 2.1 Алгоритми статистичного аналізу текстових документів

Перші ідеї статистичного машинного перекладу були представлені Уорреном Вівером ще в 1947 році. Він пояснив, що мові властива логіка, яку можна трактувати так само, як будь-який логічний математичний виклик. Він стверджував, що логічний висновок можна використовувати для ідентифікації "висновків" цільовою (неперекладеною) мовою на основі того, що вже існувало в мові-джерелі (перекладі). З появою хмари та доступністю потужних комп'ютерів теорія статистичного машинного перекладу (SMT) стала практичним варіантом.

Отже, SMT є парадигмою машинного перекладу, де переклади формуються на основі статистичних моделей, параметри яких впливають із аналізу двомовних текстових корпусів (текстових тіл) - вихідного тексту перекладеного матеріалу та цільового тексту неперекладеного матеріалу.

Статистичний машинний переклад (SMT) має справу з автоматичним відображенням речень однією людською мовою (наприклад, українською) на іншу людську мову (наприклад, англійську). Перша мова називається вихідною, а друга - цільовою. Цей процес можна сприймати як стохастичний процес. Існує багато варіантів SMT, залежно від того, як моделюється переклад. Деякі підходи стосуються відображення рядків до рядків, деякі використовують дерева-до-рядків, а деякі використовують моделі від дерева до дерева. Усі спільно поділяють центральну ідею автоматичного перекладу, причому моделі оцінюються з паралельних корпусів (пари джерело-ціль), а також з одномовних корпусів (приклади цільових речень).

Статистичний машинний переклад починається з дуже великого набору даних затверджених попередніх перекладів. Це відоме як корпус (корпуси множини) текстів, який потім використовується для автоматичного виведення статистичної моделі перекладу. Потім цю модель застосовують до

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		26

неперекладених цільових текстів, щоб встановити збіг на основі ймовірності, щоб запропонувати розумний переклад.

Ну, протягом багатьох років більшість найбільших глобальних організацій, наприклад, ЄС, ООН, Світовий банк, Світова організація охорони здоров'я тощо, розробляють величезні корпоративні корпорації у різних поєднаннях мов джерела та цільової мови. Багато з них виникли як перекладені людиною тексти. Потім вони стають доступними для користувачів/розробників машинного перекладу для подальшого вдосконалення та використання для цілей машинного перекладу. Процес є еволюційною моделлю, при цьому кожен існуючий корпус постійно вдосконалюється, додається та оновлюється.

Ця парадигма SMT базується на тому, що відоме як "імовірнісна математична теорія". Така теорія передбачає шанси (ймовірність) чогось відбутися залежно від різних змінних, які можуть вплинути на подію. Наприклад, кидання кубика дає ймовірність того, що одна кістка буде певним числом, дорівнює  $1/6$ . Отже, у азартного гравця є кожен шостий шанс приземлитися на обраному ним номері. Ймовірність того, що дві кубики будуть однаковими, становить  $1/6 \times 1/6 = 1/36$ . У покері професійний азартний гравець відстежує карти, на які грається, і використовуючи теорію ймовірностей, він/вона вирішує, чи має азартний гравець хороші шанси на перемогу, виходячи з шансів, які вони розраховали в своїй голові.

У SMT інженер машинного перекладу створює модель перекладу, використовуючи частоту фраз, що відображають навчальний базис у таблиці. Ця таблиця зберігає фразу та кількість повторень цього циклу протягом усього навчального базису. Чим частіше фраза повторюється у навчальному базисі, тим більш вірогідним є правильний переклад цілі. Кожна фраза (що зберігається у таблиці фраз) може складати від одного до п'яти слів. Ця таблиця фраз називається "Модель перекладу".

Отже, інженер машинного перекладу використовує модель ймовірності, щоб потрапити в правильну комбінацію перекладу джерело/ціль. Процес еволюційний, оскільки корпус вдосконалюється та коригується після кожного перекладу для усунення/коригування будь-яких аномалій. Чим частіше

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		27

використовується базис, тим досконалішим він стає. Розвиток якості базису – це постійний органічний процес досліджень та розробок високоцінного перекладацького майна.

Крім того, інженер машинного перекладу створює вторинну модель, використовуючи цільові дані перекладу. Ця модель допомагає визначити порядок, у якому інженеру потрібно збирати фрази (з таблиці фраз), щоб оптимізувати плавність перекладу, тобто надати перекладеному тексту природний мовний потік. Бездоганність гарантує, що буквальні переклади (тобто всі слова є, але сенс речення - ні) замінюються більш природними звуковими перекладами.

Для того, щоб перекласти вихідне речення, інженер машинного перекладу проходить такий процес декодування (як зазначено на діаграмі нижче):

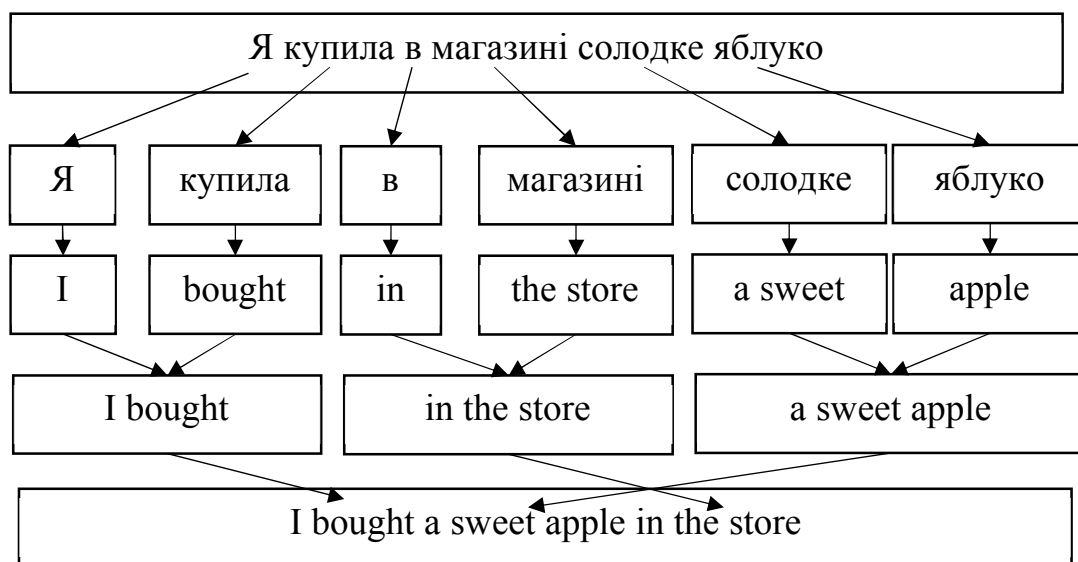


Рисунок 2.1 – Приклад кроків роботи алгоритму машинного перекладу

Алгоритм розбиває речення мовою джерела на фрази.

Потім шукає кожну з цих фраз у таблиці фраз/модель перекладу та формує переклади цільовою мовою.

Потім інженер використовує таблицю фраз/модель перекладу, щоб змінити порядок цих фраз, щоб оптимізувати плавність перекладу.



Формально переклад можна описати як пошук найбільш ймовірного цільового речення  $e^*$  для деяких вихідних речень  $f$ :

$$e^* = \operatorname{argmax}_e P(f|e)P(e),$$

де  $e$  – мова перекладу;

$f$  – мова оригіналу.

Цей підхід має три основні аспекти:

Модель перекладу ( $P(f|e)$ ), яка визначає набір можливих перекладів для певного цільового речення. Модель перекладу також призначає ймовірності цих перекладів, представляючи їх відносну правильність.

Мовна модель ( $P(e)$ ), яка моделює плавність запропонованого цільового речення. Це призначає розподіл за рядками, з більшою ймовірністю призначається реченням, які є більш репрезентативними для природної мови. Мовні моделі, як правило, є згладженими  $n$ -грамовими моделями, як правило, зумовленими двома (або більше) попередніми словами при прогнозуванні ймовірності поточного слова.

Процес пошуку (операція  $\operatorname{argmax}$ ), який стосується навігації у просторі можливих цільових перекладів. Це називається декодуванням. Розшифровка  $\operatorname{smt}$  є важкою для NP, тому в більшості підходів використовується пошук відповідного променя.

Це називається підходом Source-Channel до перекладу. Більшість сучасних систем SMT замість цього використовують лінійно-часову модель, оскільки вона є більш гнучкою та дозволяє збалансувати різні аспекти перекладу:

$$e^* = \operatorname{argmax}_e (\sum f_i(e,f) \lambda_i),$$

де  $f_i(e,f)$  – функціональні функції які фіксують певний аспект перекладу;

$\lambda_i$  – вага функціональної функції.

Коли присутні дві функціональні функції  $P(f|e)$  та  $P(e)$ , то маємо модель Source-Channel. Вагові коефіцієнти є масштабними факторами (збалансовані

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						29
Змн.	Арк.	№ докум.	Підпис	Дата		

внески, що вносить кожна функційна функція) і оптимізовані щодо певної функції втрат, яка оцінює якість перекладу. Часто це стосується метрики оцінки bleu. Як правило, поверхня помилки не є опуклою, а функція втрат недиференційована, тому слід застосовувати методи пошуку, які не використовують похідні першого порядку. Варто зазначити, що оцінка машинного перекладу є складною проблемою, і що такі методи, як bleu, не позбавлені критики.

Системи SMT зазвичай розкладають цілі речення на послідовність рядків, що називаються фразами. Потім завдання моделювання стає завданням визначення того, як розбити вихідне речення на послідовність суміжних фраз і як вказати, яку вихідну фразу слід асоціювати з кожною цільовою фразою. Системи, засновані на фразах, випереджають попередні моделі на основі слів, оскільки переклад на основі фраз може фіксувати місцевий (у межах фрази) порядок слів. Крім того, підходи до перекладу на основі фраз повинні приймати менше рішень, ніж моделі на основі слів. Це означає, що помилок допускається менше.

## 2.2 Методи автоматичного оцінювання якості машинного перекладу

Оцінка машинного перекладу слугує двом цілям: відносна оцінка дозволяє виявити, чи одна система машинного перекладу краща за іншу, а абсолютна оцінка (що має значення від 0 до 1) дає абсолютний показник ефективності (наприклад, коли дорівнює одиниці, це означає ідеальний переклад).

Однак розробка відповідних методів для чисельної оцінки якості машинного перекладу є складним завданням. У багатьох галузях науки існують вимірювані показники ефективності, такі як, наприклад, різниця між прогнозованими та фактично спостережуваними результатами. Оскільки природні мови складні, то оцінити правильність перекладу надзвичайно складно. Дві абсолютно різні послідовностей слів (пропозиції) може бути повністю

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дата		

еквівалентні (наприклад “ваза стоїть на столі” та “на столі є ваза”), і дві послідовності, які відрізняються невеликою деталлю може мати абсолютно різні значення.

Традиційно підставами для оцінки якості машинного перекладу є адекватність (переклад передає те саме значення, що й оригінальний текст), та милозвучність (переклад правильний з граматичної точки зору). Більшість сучасних методів оцінки якості машинного перекладу спираються на довідкові переклади. Раніше підходи до оцінювання тексту кандидата щодо посилального тексту базувались на ідеї схожості тексту кандидата (текст перекладений системою машинного перекладу) та довідкового тексту (текст перекладений професійним перекладачем), тобто, оцінка подібності повинна була бути пропорційною кількості відповідних слів. Приблизно в той же час була висунута інша ідея. Вона базувалася на тому, що відповідні слова в правильному порядку в кандидатурі та посилальних реченнях повинні мати вищі бали, ніж відповідні слова не в порядку.

Можливо, найпростіша версія цієї самої ідеї полягає в тому, що текст кандидата повинен бути винагороджений за те, що він містить довші суміжні послідовності відповідних слів. Конкретна версія цієї ідеї, яку вони називають, дуже корелює з людськими судженнями. Інший варіант цієї ідеї, який зараз широко відомий як бал "NIST". Хоча показники BLEU та NIST можуть бути корисними для порівняння відносної якості різних виходів МТ, важко отримати розуміння від таких заходів.

На сьогоднішній день основний підхід до оцінки якості мовних моделей для систем машинного перекладу спирається на використання статистичних методів. У цьому випадку модель насправді є розподілом ймовірностей на множині всіх речень мови. Природно, застосовувати модель таким чином неможливо; отже, використовуються більш компактні алгоритми. Коротко розглянемо, які моделі в даний час використовуються в комерційних та експериментальних системах оцінки якості машинного перекладу з необмеженими словниками.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

Метод приблизного збігу рядків. В інформатиці приблизна відповідність рядків (часто розмовно називається нечітким пошуком рядків) - це техніка пошуку рядків, які приблизно відповідають шаблону (а не точно). Проблема пошуку приблизного збігу рядків, як правило, ділиться на дві підзадачі: пошук приблизного підрядка всередині даного рядка та пошук рядків словників, які приблизно відповідають шаблону.

Коефіцієнт помилок слів (wer) - це показник, заснований на цьому підході. wer обчислюється як сума вставок, видалень та підстановок, нормована на довжину посилального речення. Якщо wer дорівнює нулю, переклад ідентичний тексту посилання. Основна проблема полягає в тому, що отримана оцінка не завжди знаходиться в діапазоні від 0 до 1. У деяких випадках, коли переклад помилковий, wer може бути більше 1.

Іншою версією wer є метрика werg, в якій сума вставок, видалень та заміщ нормується на відстань Левенштейна, тобто тривалість редагувань. У теорії інформації та обчислювальній лінгвістиці відстань Левенштейна (редакційна відстань або відстань редагування) між двома рядками визначається як мінімальна кількість редагувань, необхідних для перетворення одного рядка в інший, при цьому допустимими операціями редагування є вставка, видалення або заміна один символ. Перевага цієї метрики полягає в тому, що значення якості перекладу завжди знаходитиметься в діапазоні від 0 до 1 (навіть у гіршому випадку збігу обставин або за відсутності перекладу значення не перевищуватиме одиницю).

Експерименти, проведені показали, що метрика werg не є надійною і не узгоджується з оцінками, отриманими при аналізі машинного перекладу людьми.

Частота помилок, незалежна від позиції (per), нехтує порядком слів в операції зіставлення рядків. У цьому випадку обчислюється різниця між текстом кандидата та посиланням, нормована на довжину перекладу посилання.

Ще однією метрикою, яка широко використовується при оцінці якості перекладу, є коефіцієнт помилок перекладу (ter). Ця метрика дозволяє виміряти кількість редагувань, необхідних для зміни вихідних даних системи, в один із поданих посилань на переклади.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		32

Насправді, будь-яка метрика відповідності рядків може бути використана для оцінки якості машинного перекладу. Одним із таких прикладів є «ядро рядків», яке дозволяє враховувати різні рівні природної мови (наприклад, морфологічний, лексичний тощо) або взаємозв'язок між синонімами.

*N*-грамні моделі. У мовних моделях *n*-грамів використовується явне припущення, що ймовірність наступного слова в реченні залежить від попередніх *n*-1 слів. На практиці використовуються моделі з *n*=1, 2, 3 і 4. Для англійської мови найбільш вдалими є триграмові або чотириграмові моделі. Сьогодні майже всі системи оцінки якості машинного перекладу покладаються на *n*-грамові моделі. У цьому випадку ймовірність цілого речення обчислюється як добуток ймовірностей складових *n*-грам.

Основними перевагами *n*-грамних моделей є їх відносна простота та можливість побудови моделі, яку можна навчити на досить великому корпусі мови. Однак такі моделі не позбавлені недоліків. У моделі *n*-грам зробити це неможливо змодельовати семантичні і прагматичні відносини в мові. Насправді, якщо словник містить *N* слів, кількість можливих пар слів буде  $N^2$ . Навіть якщо лише 0,1% з них насправді трапляються в мові, мінімальний обсяг мовного корпусу, необхідний для отримання статистично достовірних оцінок, становитиме 125 мільярдів слів або приблизно 1 терабайт. Для триграмових моделей мінімальний корпус досягне сотень тисяч терабайт.

Для подолання недоліків використовується добре розроблена техніка згладжування, яка дає змогу оцінити параметри моделі в умовах недостатньої або неіснуючої інформації.

Основними показниками на основі *n*-грам є BLEU, NIST, F-міра та METEOR.

BLEU (Bilingual Evaluation Understudy) - це алгоритм автоматичного оцінювання якості машинного перекладу, який порівнюється з еталонним перекладом, використовуючи *n*-грам.

Вимірювання якості перекладу є складним завданням, насамперед через відсутність визначення "абсолютно правильного" перекладу. Найбільш поширеною методикою оцінки якості перекладу є порівняння результатів

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		33

автоматизованих та людських перекладів того самого документа. Але це не так просто, як може здатися: переклад одного перекладача може відрізнятись від перекладу іншого перекладача. Ця невідповідність між різними довідковими перекладами представляє серйозну проблему, особливо коли різні довідкові переклади використовуються для оцінки якості автоматизованих рішень перекладу.

Документ, перекладений спеціально розробленим автоматизованим програмним забезпеченням, може збігатися на 60% з перекладом, виконаним одним перекладачем, і на 40% збігатися з перекладом іншого перекладача. Хоча обидва професійні переклади є технічно правильними (вони граматично правильні, вони мають однакове значення тощо), 60% перекриття слів є ознакою вищої якості машинного перекладу. Таким чином, хоча для порівняння використовуються посиальні переклади, вони не можуть бути цілком об'єктивним та послідовним виміром якості машинного перекладу.

Метрика BLEU оцінює якість машинного перекладу за шкалою від 0 до 1. Чим ближче оцінка до одиниці, тим більшим є перекриття з еталонним перекладом і, отже, тим кращою є система машинного перекладу. Якщо коротко, метрика BLEU вимірює, скільки слів збігається в одному рядку, причому найкраща оцінка дається не відповідним словам, а послідовностям слів. Наприклад, рядок із чотирьох слів у перекладі, який відповідає перекладу з людським посиланням (у тому самому порядку), матиме позитивний вплив на оцінку BLEU і зважуватиметься важче (і оцінюватиметься вище), ніж одно- або двословне збіг.

Точність вимірювання NIST (Національний інститут стандартів і технологій) - це показник, що використовується для оцінки варіантів машинного перекладу. NIST був задуманий як вдосконалена версія BLEU. У цьому випадку обчислюється середнє арифметичне  $n$ -грам. Важливою відмінністю від метрики BLEU є той факт, що NIST також покладається на частотну складову (точність та відкликання). Якщо BLEU просто обчислює точність  $n$ -грам, додаючи однакову вагу для кожного точного збігу, NIST також обчислює, наскільки інформативним є кожен  $n$ -грам, що відповідає.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						34
Змн.	Арк.	№ докум.	Підпис	Дата		

Наприклад, навіть якщо біграма "на" збігається з тією ж фразою в тексті посилання, переклад все одно отримує нижчий бал, ніж правильне узгодження біграми "розподіл за розміром", оскільки остання фраза менш вірогідна.

F-міра - це метрика, яка обчислює середнє гармонічне значення точності та відклику. Показник базується на пошуку найкращої відповідності між кандидатом та посиланням на переклади (співвідношення між загальною кількістю відповідних слів до довжини перекладу та довідкового тексту). Іноді корисно поєднувати точність і виклик того самого усередненого значення.

Показник для оцінки перекладу з явним замовленням (METEOR) є вдосконаленою версією F-міри. Ця система була розроблена для усунення деяких слабких місць метрики BLEU. METEOR оцінює результати, підбираючи автоматизовані та довідкові переклади буквально до слова. Коли доступно більше одного посилального перекладу, автоматизований переклад порівнюється з кожним із них і повідомляється про найкращий результат.

Можна по-різному ставитись до різних метрик, але на даний момент BLEU, METEOR та NIST використовуються найбільш широко. Саме ці показники порівнюються з усіма іншими системами оцінки якості машинного перекладу. Розробники F-міри стверджують, що їх показник найкраще узгоджується з оцінкою, зробленою людиною. Однак це не завжди так. F-міра погано працює з найменшою середньою відстанню редагування. Емпіричні дані показують, що слід приділяти більше уваги повноті (згадуванню) перекладу. Дослідження показують, що найчастіше відкликання є параметром, що дозволяє визначити якість перекладу.

### 2.3 Алгоритм перекладу документів на основі API онлайн перекладачів

Для реалізації програмного додатку аналізу текстів та їх подальшого перекладу було запропоновано алгоритм в основі якого закладени принципи статистичного аналізу текстів та подальшого використання онлайн сервісів для

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		35

перекладу в реальному часі. Оскільки швидкість при перекладі відіграє важливу роль, тому під час проектування алгоритму значна увага буде приділена до того, щоб алгоритм швидко та з достатньою точністю могли перекладати тексти з однієї мови на іншу, при цьому інформативні втрати при перекладі були мінімальними.

Запропонований алгоритм містить наступні етапи роботи:

Етап 1. Отримання/завантаження тексту.

Етап 2. Встановлення параметрів роботи алгоритму (вибір мови джерела, вибір мови перекладу, вибір напрямку перекладу).

Етап 3. Проведення первинної оцінки отриманого тексту (перевірка чи введений текст відповідає обраній мові).

Етап 4. Розбиття вхідного тексту на множину окремих слів.

Етап 5. Формування словосполучень на основі коефіцієнтів зв'язності між окремими словами які знаходяться поруч. Відстань між сусідніми словами може бути 0, 1 або 2 слова.

Етап 6. Формування запиту для онлайн сервісу перекладу текстових даних, Проводиться тестовий запит, якщо під час повернення відповіді на запит виникла помилка невідповідності шаблону, то вносимо зміни в шаблон запиту.

Етап 7. Отримання відповіді від онлайн сервісу перекладу текстових даних.

Етап 8. Проведення орфографічної оцінки отриманого перекладу на основі правил відповідної мови, що задають масивом шаблонів.

Етап 9. Вивид отриманих результатів машинного перекладу на екран.

Розроблений алгоритм базується на основних принципах та сучасних наукових розробках в галузі машинного перекладу та враховує всі особливості при роботі з автоматичними системами машинного перекладу. В ньому передбачено етап попередньої обробки, що дозволяє зменшити час обробки тексту на сервісі онлайн перекладу, дозволяє провести групування словосполучень для більш точного перекладу тексту, встановлення напрямку перекладу для більш точного підбору слова або словосполучення з множини синонімів. Графічно запропонований алгоритм можна представити у вигляді наступної блок-схеми (рисунок 2.3):

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						36
Змн.	Арк.	№ докум.	Підпис	Дата		





Рисунок 2.3 – Блок схема алгоритму перекладу текстової інформації на основі використання онлайні сервісів

Серед переваг запропонованого алгоритму є:

- можливість швидкої реалізації;
- автоматична перевірка результатів машинного перекладу;
- можливість роботи з текстами на різних мовах.

Серед недоліків є:

- труднощі при аналізі текстів, що містять багато граматичних та орфографічних помилок;
- складність роботи алгоритму, якщо введений текст містить слова на різних мовах.

Змн.	Арк.	№ докум.	Підпис	Дата

### 3. ПРОГРАМНА СИСТЕМА ОНЛАЙН ПЕРЕКЛАДУ ТЕКСТУ

#### 3.1 Структура автоматизованої системи перекладу тексту

Для написання тестового програмного забезпечення було використано програмне середовище Visual Studio 2019 та мову програмування C#. Вибір даного набору засобів для розробки був пов'язаний з тим, що під час навчання саме на таких програмних розробках було отримано максимальний досвід роботи. Окрім того, обрана мова є широко відома та розповсюджена. Мова C# - це сучасна, об'єктно-орієнтована мова програмування загального призначення, розроблена корпорацією Майкрософт в рамках її .Net та за ініціативи Андреса Хейлсберга. Мова програмування C# - це дуже проста мова для вивчення. Вона повністю заснований на мовах C та C++. Синтаксис мови C# є дуже виразним, проте він також простий і легкий у вивченні. Будь-хто може миттєво розпізнати синтаксис фігурних дужок C#, хто знає про мову C, C++ або Java. Існує багато важливих особливостей мови C#, які роблять її більш корисною та унікальною в порівнянні з іншими мовами.

Швидка швидкість. Мова C# дуже швидка, її компіляція та час виконання занадто швидкі.

Простота. C# - це проста мова. Це дає структурований підхід до розбиття проблеми на частини. Крім того, він має багатий набір функцій бібліотеки та типів даних. Код мови C# не вимагає файлів заголовків. Його код пишеться вбудовано.

Об'єктно-орієнтованість. Мова C# - це об'єктно-орієнтована мова програмування. Так само, Ooprs полегшує розробку та обслуговування, порівняно з мовою програмування, орієнтованою на процедури. Однак надто складно управляти, якщо код зростає із збільшенням розміру проекту. Крім того, програмування на C# підтримує інкапсуляцію даних, успадкування, поліморфізм, інтерфейси.

Сучасна мова програмування. Мова C# є однією з сучасних мов програмування, оскільки вона базується на сучасній тенденції. Однак він дуже

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дата		

простий, потужний для створення масштабованих, сумісних та надійних додатків.

Тип-захищена. Мова C # - це безпечний для коду код, який може отримати доступ лише до місця пам'яті та має дозвіл на виконання. Отже, це покращує безпеку програми. Мовою C # не можете виконувати небезпечні трансляції, такі як перетворення double у логічну форму. Його типи значень (примітивні типи) ініціалізуються нулями, а посилальні типи (об'єкти та класи) компілятором автоматично ініціалізуються до нуля.

Сумісність. Сумісність - це процес, який дозволяє програмам C # робити майже все, що може робити власний додаток C ++. Коротше кажучи, сумісність мови - це здатність коду взаємодіяти з кодом, який написаний за допомогою іншої мови програмування. Це може допомогти максимізувати повторне використання коду і, отже, підвищити ефективність процесу розробки. Мова C # забезпечує підтримку використання СОМ-об'єктів, незалежно від того, якою мовою їх було використано. Однак він також підтримує спеціальну функцію, яка дозволяє програмі викликати будь-який власний АРІ.

Масштабованість та постійна оновленість. Мова C# - це комп'ютеризована, масштабована та оновлювана мова програмування. Однак одна важлива річ полягає в тому, щоб оновити ваш .Net framework. вам доведеться вбити ваші старі файли та оновити їх новими.

Мова структурованого програмування. Мова C # - це структурована мова програмування. Однак структуровані мови програмування - це підмножина процедурного програмування, яка забезпечує логічну структуру програми, що пишеться, щоб зробити її більш ефективною та простішою для розуміння та модифікації. Іншими словами, для вирішення великих проблем програмування C # ділить проблему на менші модулі, які називаються функціями або процедурами, кожен з яких несе певну відповідальність, тому мова C # називається структурованою мовою програмування.

Багата бібліотека. Мова C # багата на бібліотеку. Так що він забезпечує безліч вбудованих функцій, які пришвидшують розробку.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		39

Орієнтована на компоненти. Мова С # є мовою програмування, орієнтованою на компоненти, і підтримує компонентно-орієнтоване програмування через концепції методів, властивостей, подій та атрибутів (або метаданих), дозволяючи самостійні та самоописуючі компоненти функціональних можливостей, які називаються збірками.

Розмір великого програмного забезпечення обчислюється сотнями модулів, десятками і сотнями тисяч машинних інструкцій. Існують програмні засоби, що містять до двох-трьох десятків структурних ієрархічних рівнів, побудованих з модулів.

На схемі показана структурна схема програми (рисунок 3.1):



Рисунок 3.1 – Структура програмного додатку перекладу текстів

Модуль запуску системи. Цей блок формується функціональними підсистемами для авторизації користувачів у системі, конфігурації системних параметрів (за замовчуванням), перевірки цілісності програмної системи,

перевірки наявності необхідних програмних та апаратних компонентів. в системі (бібліотеки програмного забезпечення, пристрої вводу-виводу тощо).

Модуль вводу текстового повідомлення – функції даного методі дозволяється організувати процес вводу текстової інформації як безпосередньо з клавіатури користувача, так і з текстових файлів які розміщені на жорстких носіях даних або в глобальній мережі. В даному модулі не передбачені функції аналізу тексту тому вважається, що користувач ввів саме ту інформацію яку він хоче перекласти на іншу мову.

Модуль попереднього аналізу – оскільки повідомлення можуть бути введені не на тій мові яка вказана як джерело, то передбачена деяка перевірка коректності вводу інформації, зокремо проводиться перевірка на наявність символів, що є не характерними для конкретної мови. Дана перевірка дозволяє зменшити кількість хибних спрацювань та підвищити рівень результатів перекладу. Проте даний підхід має один недолік, оскільки проводиться перевірка, а користувач має незначний досвід роботи з програмою, то він може незрозуміти підказок програмної системи, та видалити ті частини тексту, що насправді повинні бути написані іншою мовою, наприклад власні імена.

Модуль розбиття тексту – Функції даного програмного модуля призначені для виділення в загальному тексті окремих слів. Окрім того іншою не менш важливою частиною даного модуля є функції групування слів в окремі словосполучення на основі коефіцієнту сумісності пар слів. Формування словосполучень дозволяє більш точно передати суть тексту під час перекладу, оскільки надає можливість зробити переклад більш «людянішим» на основі вживання виразів які не можна перекласти дослівно.

Модуль формування запитів – функціонал даного модуля дозволяє сформувати запит для онлайн сервісу перекладу текстів на основі деяких шаблонів та API-функцій, що підтримуються відповідним сервісом. Це дозволяє організувати інтерфейс між програмним додатком та онлайн сервісом. Структура запитів надається сервісами і повідомлення які їм не підтримують не будуть оброблені в подальшому.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						41
Змн.	Арк.	№ докум.	Підпис	Дата		

Модуль взаємодії з сервісом перекладу – Це допоміжний модуль, що дозволяє проводити перевірки доступу до сервера, проводити процедури зміни параметрів доступу до сервісу, проходити ідентифікацію та виводить системні повідомлення у разі виявлення помилок або несправностей під час обміну інформацією. Також в даному блоці реалізована процедура перевірки коректності надісланих та отриманих повідомлень, а також забезпечує повторне відправлення запитів у випадку виявлення технічних проблем.

Модуль орфографічного аналізу тексту – даний модуль забезпечує обмін повідомленнями з онлайн сервісом перевірки орфографії. Оскільки перевірка орфографії це окрема складна наукова задача, а також враховуючи, що правила в різних мовах змінюються достатньо часто було вирішено скористатись зовнішніми сервісами для вирішення даної задачі.

Модуль виводу інформації на екран – даний модуль використовується для візуалізації отриманих від сервісу перекладів повідомлень у зручному для користувача форматі. Окрім того, дозволяє вільно копіювати, редагувати та переміщувати отримані текстові дані у інші програмні системи для подальшої роботи з ними, наприклад у ms word.

Для проведення попереднього моделювання розроблювального програмного додатку, використовуються ряд UML діаграм. Вибір UML пов'язаний з тим, що:

Ваша програмна система професійно розроблена і задокументована перед кодуванням. Ви будете точно знати, що отримуєте, заздалегідь.

Оскільки дизайн системи стоїть на першому місці, багаторазовий код легко помітити та закодувати з найвищою ефективністю. Ви матимете менші витрати на розробку.

Логічні «діри» можна помітити на кресленнях дизайну. Ваше програмне забезпечення буде поводитися так, як ви очікуєте. Сюрпризів стає менше.

Загальний дизайн системи визначатиме спосіб розробки програмного забезпечення. Правильні рішення приймаються до того, як в реальності стикнетесь з погано написаним кодексом. Знову ж кошторис загальних витрат буде меншим.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						42
Змн.	Арк.	№ докум.	Підпис	Дата		

Якщо розробникам потрібно спілкуватися із зовнішніми підрядниками або навіть із власними програмістами, це набагато ефективніше.

Легко налагоджувати будь-яку проблему. Більш обширна програма може зайняти до годин і годин пошуку, щоб знайти помилку в ній, а також може спричинити проблеми в системі пізніше. Отже, програма добре розроблена за допомогою UML, так що кожне із завдань містить свої власні коди, які програмісту легше налагоджувати.

В першому варіанті розглядатиметься розроблювальна система як сукупність акторів та прецедентів з якими вони можуть взаємодіяти під час роботи з програмою (рисунок 3.2):

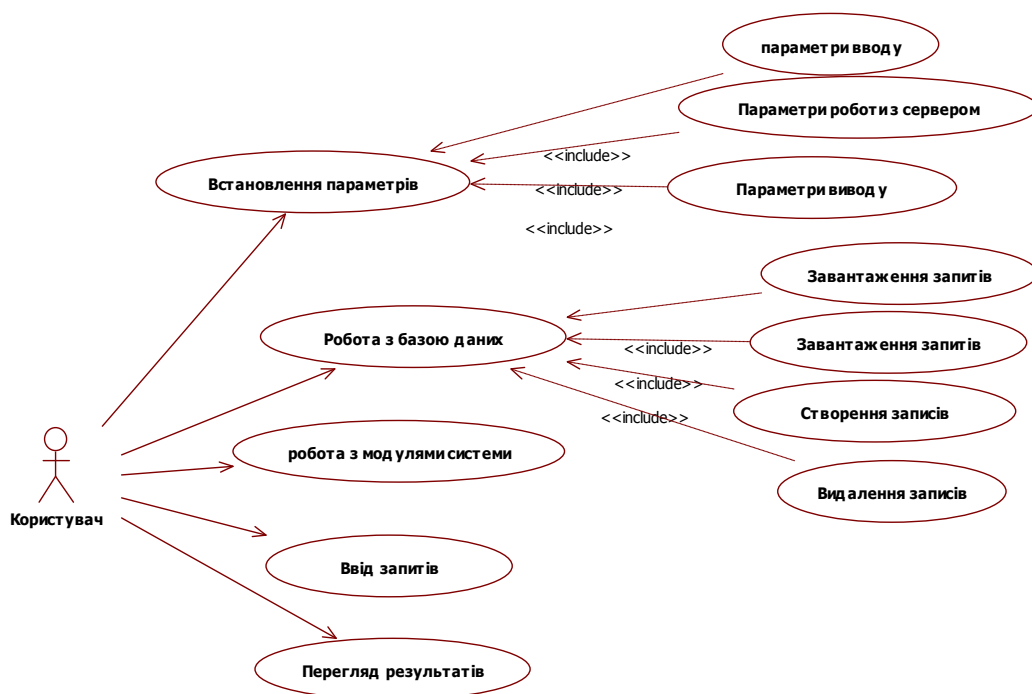


Рисунок 3.2 – Діаграма прецедентів програмного додатку перекладу текстів на основі використання онлайн сервісу перекладу

При моделюванні було тип акторів, що будуть взаємодіяти з програмною системою. Актори типу "Користувач" мають змогу створювати, редагувати, відправляти запити на серверну частину програмного засобу та переглядати отримані результати. Окрім функції обробки текстової інформації дана група користувачів матиме змогу встановлювати різні параметри роботи з програмною

системою, що значно підвищує рівень взаємодії користувачів з програмом та дозволяє їм отримувати більш точні та більш якісніші результати перекладу.

Інший важливий етап що був розглянутий під час моделювання програмного додатку – це процес обміну повідомленнями між різними групами користувачів та окремими елементами програмної системи та сервісу перекладу. Результати модулювання наведено на рисунку 3.3.

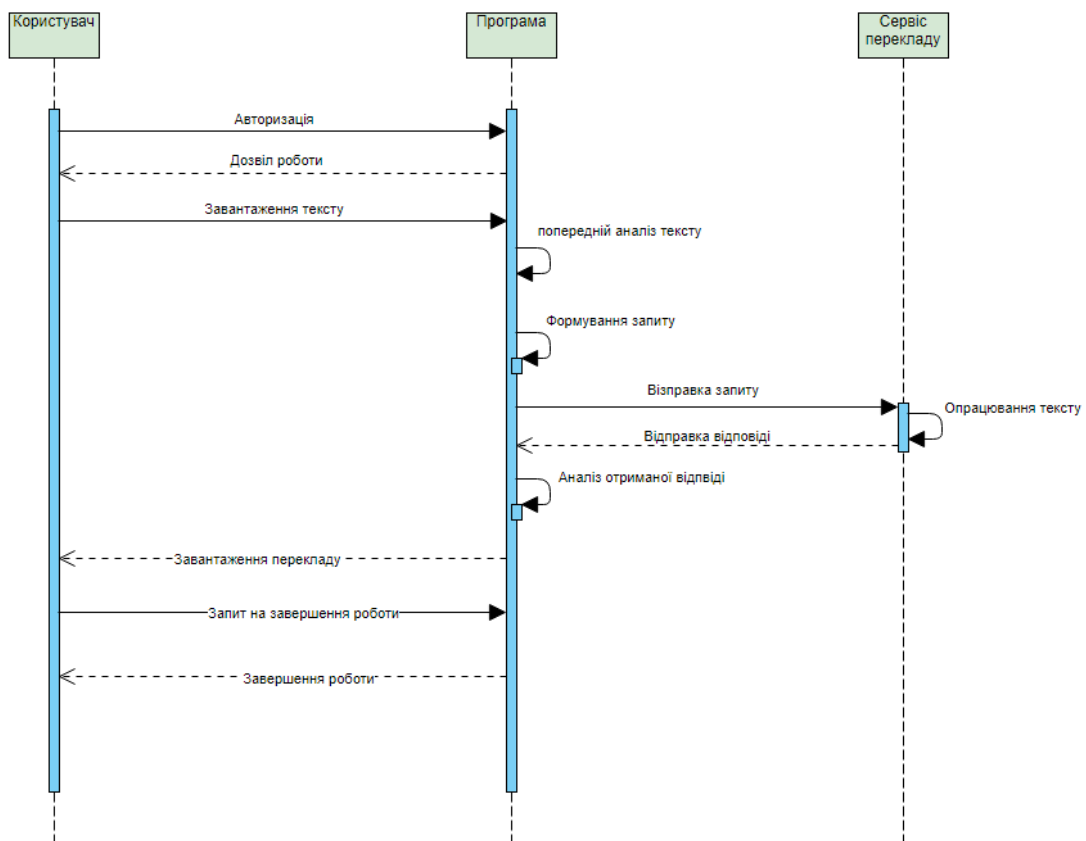


Рисунок 3.3 – Діаграма послідовності взаємодії користувачів з програмною системою

Моделювання послідовності кроків роботи з програмною системою проілюстрував, що в програмі:

- відсутні неоднозначні дії, що можуть сповільнити програму;
- для досягнення мети користувачу слід виконати мінімальну кількість дій;
- час на освоєння програми є мінімальний та не залежить від знань комп'ютерної грамотності.

Моделювання показала, що проектування програми в напрямку роботи в ній одного типу користувачів дало можливість не розпошувати та обмежувати



функції, що реалізовані в програмі, та значно підвищити цілісність та надійність розроблювального програмного додатку.

Для швидкого оволодіння практичними навичками роботи з системою розроблений простий та інтуїтивно зрозумілий інтерфейс (рисунок 3.4):

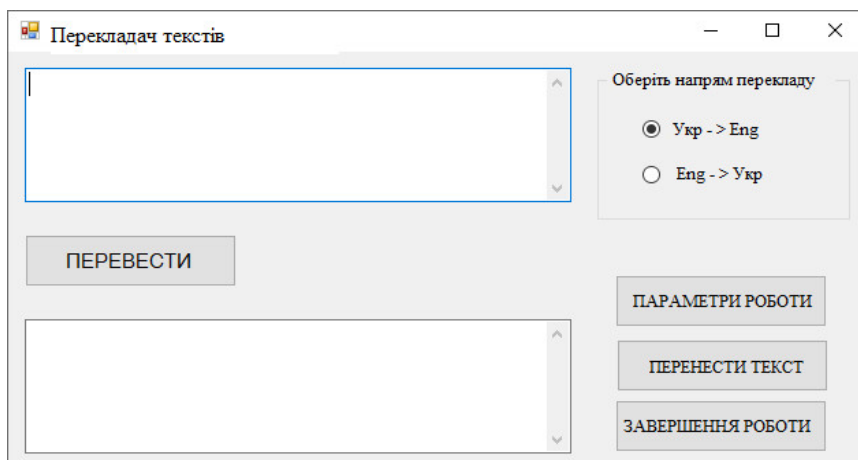


Рисунок 3.4 – Головне вікно розробленої системи

При проектуванні інтерфейсу були враховані результати аналізу програм-аналогів, а також принцип мінімалізму який рекомендує використання мінімальної кількості візуальних елементів на головному вікні для забезпечення швидкого опанування програмними додатками або пристроями новими користувачами навіть з мінімальним досвідом роботи з подібними системами до того.

Графічний інтерфейс містить два вікна для вводу текстових повідомлень та для виводу отриманої інформації. За допомогою них користувач отримує можливість для обміну запитами з сервісом онлайн перекладу. Окрім того вони займають 80% робочої області програми, що є достатньо зручним та зрозумілим рішенням.

Інша частина головного вікна забезпечує можливість користувача виконувати додаткові функції при роботі з програмною системою. Зокрема змінювати параметри роботи різних модулів системи, передавати отримані результати в інші програмні додатки тощо. Даний блок перемикачів доступний для всіх, проте під час активації процесу встановлення параметрів користувачеві

виводиться системне повідомлення про можливість виникнення критичних ситуацій при некоректному встановленні параметрів.

Дана структура та інтерфейс сприяє швидкому опануванню та простоті в роботі з програмною системою, а також дозволяє вільно нарощувати систему як в вертикальному напрямку (шляхом додавання нових модулів системи) так і в горизонтальному (шляхом додавання нових алгоритмів в відповідні функціональні модулі).

### 3.2 Підсистеми додатку онлайн перекладу текстів

API (скорочення від англійського Application Programming Interface) - це інтерфейс програмування додатків. Інтернет-сервіси, бібліотеки, операційна система і ін. Надають розробникам інтерфейси для взаємодії зі своїми наборами готових класів, методів і даних. В результаті цього ми можемо в своїх додатках використовувати функції інших інтернет-сервісів.

API надають багато сервісів, зокрема це соціальні мережі Facebook, Instagram і Twitter; також це можуть бути всілякі перекладачі, конвертери валют, інформаційні та новинні сервіси. Крім того, можна зробити API на власному web-ресурсі і використовувати його при розробці своїх додатків, або зробити його публічним - і тоді інші люди зможуть ним скористатися.

Створимо клас, який буде реалізовувати взаємодію з API перекладача на основі web-запитів. Клас містить єдиний метод `public string Translate (string s, string lang)`. Для роботи даного метода йому необхідно передати в якості параметрів наступну інформацію: рядок `s` - вихідний текст на мові оригіналу для перекладу, стрічка `lang` – це напрямок перекладу (наприклад, «Укр-Eng» - переклад буде проводитись з української на англійську мову). Даний метод після своєї роботи буде повертати стрічку з переведеним текстом.

Процес встановлення параметрів реалізований в класі шляхом визначення для кожного з відповідних полів класу деякого значення. Дані значення

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		46

встановлюються як по замовчуванню так і на основі параметрів, що буде корегувати користувач. Нижче наведено частин програмного коду з встановленням даних параметрів.

```
class Translator
{
    public string Translate (string s, string lang)
    {
        if (s.Length > 0)
        {
            WebRequest request = WebRequest.Create (
"https://translate.google.net/api/v1.5/tr.json/translate?"
            + "Key = API-KEY"
            + "& Text =" + s
            + "& Lang =" + lang);
```

При передачі тексту для перекладу відбуваються наступні дії. Формується запит на основі шаблону API-функції та проводиться спроба передати її на сервер. Якщо спроба була успішною, то на екран виводиться відповідне повідомлення у разі виникнення помилки повертається пуста стрічка та виводиться системне повідомлення про це. Для організації передачі запиту на переклад текстової інформації використовується програмний код:

```
WebResponse response = request.GetResponse ();
using (StreamReader stream = new StreamReader
(response.GetResponseStream ()))
{
    string line;
    if ((line = stream.ReadLine ()) != null)
    {
        Translation translation = JsonConvert.DeserializeObject
<Translation> (line);
        s = "";
        foreach (string str in translation.text)
        {s += str;
        }}
        return s;
    }
    else
        return "";
}
```

Для організації перевірки стану сервера, та для проведення тестових запитів було спроектовано клас Translation. Основними методами даного класу є code – призначений для зберігання системних налаштувань, lang – функції для

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

отримання або встановлення параметрів перекладу пари мов, а також `text` – що відповідає за збереження текстових масивів даних, що вводяться користувачем або отримуються від сервісу онлайн перекладів.

```
class Translation
{
    public string code {get; set; }
    public string lang {get; set; }
    public string [] text {get; set; }
}}
```

Для реалізації графічного інтерфейсу користувача було розроблено окремий клас, параметри якого та методи дозволяють реалізувати набір візуальних компонентів, призначених для вводу/виводу текстових повідомлень та встановлення параметрів роботи програми. Код обробника події натискання кнопки "Перевести" оператори вибору для завдання потрібного напрямку перекладу, що залежить від обраного положення перемикача, а також виклик методу `Translate` безпосередньо для виконання перекладу.

```
public partial class Form1: Form
{
    Translator yt;
    public Form1 ()
    {
        InitializeComponent ();
        yt = new Translator();
    }
    private void translateButton_Click (object sender, EventArgs e)
    {string lang;
    if (ukEnRadioButton.Checked == true)
    {lang = "ua-en";}
    else
    {lang = "en-ua";}
    outputTextBox.Text = yt.Translate (inputTextBox.Text, lang);
    }}}
```

Запит до API перекладача містить ряд обов'язкових параметрів – це: `key = <API-Key> & text = <текст для перекладу> & lang = <напрямок перекладу>`. Окрім того шаблон запиту містить декілька необов'язкових з кавих полів, що можуть бути опущеними під час формування запиту. Запит до API повертає відповідь у форматі структури JSON, що дозволяє її просто використовувати надолі.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						48
Змн.	Арк.	№ докум.	Підпис	Дата		

### 3.3 Тестування додатку онлайн перекладу текстів

Оцінки розробленого програмного додатку та порівняння його з програмами аналогами був проведений етап тестування. Тести проводились на робтчій станції з такими параметрами (таблиця 3.1):

Таблиця 3.1 – Параметри персонального комп'ютера для тестування програми

Параметр	Значення
Корпус	Zalman Z1 Black + Chieftec APS-550SB
HDD	1000GB
Відеокарта	Asus PH-GTX1060-3G
ОЗУ	4Gb
Процесор	AMD Ryzen 3 2200G BOX 120
материнська плата	Asus B350M-E 90
діагональ дисплея	19
роздільна здатність дисплея	1920 x 1080
тип матриці	IPS
частота оновлення	60 Гц
Інтерфейси	HDMI
відношення сторін	16: 9

Технічні параметри комп'ютера, що використовувався під час тестів задовільняють технічним умовам для проведення тестування та дозволяють провести оцінку запропонованим алгоритмам та реалізованому програмному забезпеченню. При цьому достовірність даних буде на відповідном рівні.

Під час тестування тестові тексти були розділені на декілька класів за рівнем складності. Рівні визначались як за кількістю переданої інформації так і за наявністю в тексті стійких словосполучень або синтаксичних або орфографічних помилок.

В першій групі для тестування були відібрані прості слова або сталі словосполучення. Кількість тексту була мінімальною (до 10 слів). Орфографічні помилки відсутні. В результаті отриманих перекладів можна зробити висновки, що текст перекладався в реальному часі з незначними затримками, що можна пояснити часом передачі інформації через мережу. Точність перекладу 100%. Результати перекладу наведено на рисунку 3.5.

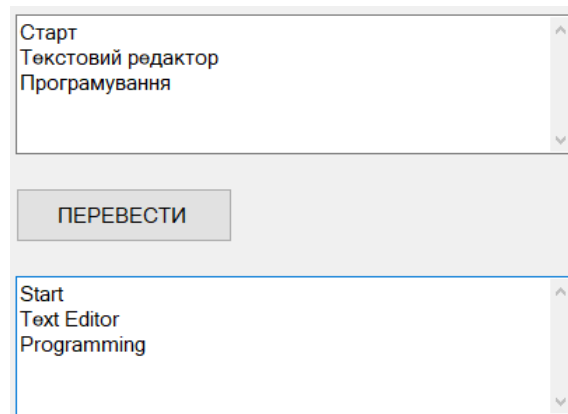


Рисунок 3.5 – Приклад перекладу простого тексту

Для проведення складнішого перекладу використовувались тексти, що містять велику кількість слів та можуть містити словосполучення які не є явно відображені в тексті (між ними можуть бути вставлені інші слова). Орфографічні помилки відсутні. Приклад такого перекладу наведено на рисунку 3.6.

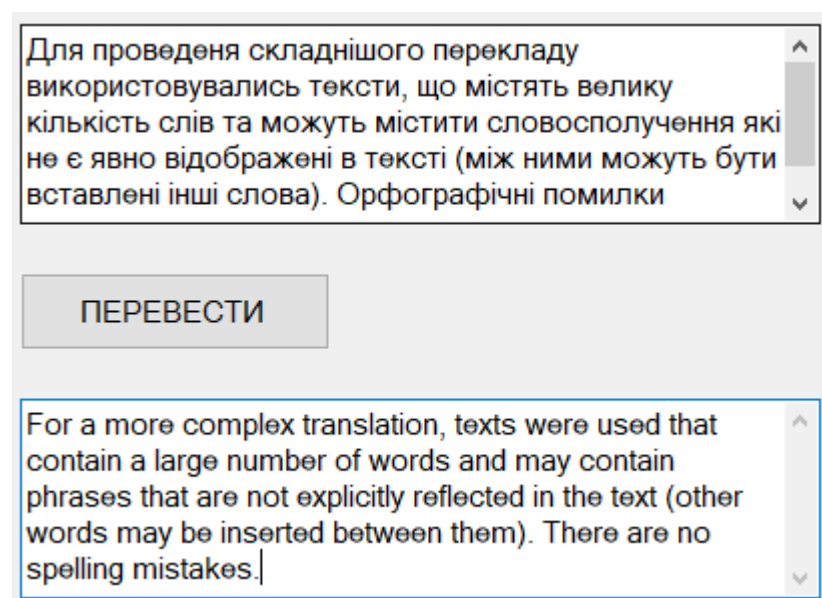


Рисунок 3.5 – Приклад перекладу простого тексту

В даному випадку текст переклався без помітних часових затримок та на високому рівні, що дозволяє підтвердити високий рівень розробленого програмного забезпечення.

Для проведення тестування останньої групи текстів використовувались тексти, що містять велику кількість слів та містять орфографічні помилки. В результаті отриманого перекладу можна зробити висновки, що отриманий результат не є якісним оскільки деякі слова були транслітеровані а не перекладуні окрім того під час опрацювання текстів з помилками деякі слова були перкладені неправильно, що в цілому спотворило суть тексту. Проте дана проблема відноситься до рівня якості вхідної інформації, а тому не може бути недоліком запропонованого алгоритму та розробленої програмної системи.

За результатами експериментів розроблений алгоритм перекладу на основі використання онлайн сервісів виконував поставлені задачі за визначений час і з максимальною точністю.

Перевагами розробленого алгоритму є:

- можливість роботи із текстами різної складності (переклад здійснюється не залежно від складності чи кількості вхідних слів та словосполучень);
- здатність працювати з різними мовами;
- вибір вхідних виробничих параметрів є мінімальним та прозорим, що дозволяє легко реалізувати запропонований алгоритм.

Недоліком запропонованого алгоритму є:

- додатковий час, витрачений на попередній аналіз вхідного тексту;
- відсутність автоматичного визначення мови оригінального тексту (залежить від вибору користувача);
- складність перекладу тексту який містить багато синтаксичних та орфографічних помилок.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						51
Змн.	Арк.	№ докум.	Підпис	Дата		

## 4 ТЕХНІКО-ЕКОНОМІЧНИЙ РОЗДІЛ

Метою техніко – економічного розділу кваліфікаційної роботи є здійснення економічних розрахунків, спрямованих на визначення економічної ефективності програмного додатку згідно затвердженого технічного завдання та на основі обраних методів, алгоритмів та технологій, а також прийняття рішення про його подальший розвиток і впровадження або ж недоцільність проведення відповідної розробки. Для проведення даного дослідження необхідно провести ряд розрахунків.

### 4.1 Розрахунок витрат на розробку програмного додатку

Витрати на розробку і впровадження програмного додатку ( $K$ ) на основі запропонованих алгоритмів та розробленої структури, що враховують результати аналізу програм-аналогів включають:

$$K = K_1 + K_2,$$

де  $K_1$  – витрати на розробку апаратного та програмного забезпечення грн.;  
 $K_2$  – витрати на відлагодження і дослідну експлуатацію програми рішення задачі на комп'ютері, грн.

Витрати на розробку апаратних та програмних засобів включають:

- витрати на оплату праці розробників ( $B_{оп}$ );
- витрати на відрахування у спеціальні державні фонди ( $B_{ф}$ );
- витрати на матеріали та комплектуючі ( $П_в$ );
- накладні витрати ( $H$ );
- інші витрати ( $I_в$ );
- витрати на використання комп'ютерної техніки ( $B_{КТ}$ ).

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		52



Розрахунок витрат на оплату праці.

Витрати на оплату праці включають заробітну плату (ЗП) всіх категорій працівників, безпосередньо зайнятих на всіх етапах проектування. Розмір ЗП обчислюється на основі трудоемності відповідних робіт у людино-днях та середньої ЗП відповідних категорій працівників.

У розробці проектного рішення задіяні наступні спеціалісти - розробники, а саме: керівник проекту; студент-дипломант; консультант техніко-економічного розділу (таблиця 4.1).

Таблиця 4.1 – Вихідні дані для розрахунку витрат на оплату праці

№п/п	Посада виконавців	Місячний оклад, грн.
1	Керівник ДП, ст. викладач	7449
2	Консультант техніко-економічного розділу, ст. викладач	7449
3	Студент	1400

Витрати на оплату праці розробників проекту визначаються за наступною формулою (4.1):

$$B_{OP} = \sum_{i=1}^N \sum_{j=1}^M n_{ij} \cdot t_{ij} \cdot C_{ij} , \quad (4.1)$$

де  $n_{ij}$  – чисельність розробників  $i$ -ої спеціальності  $j$ -го тарифного розряду;

$t_{ij}$  – затрачений час на розробку проекту співробітником  $i$ -ої спеціальності  $j$ -го тарифного розряду, год;

$C_{ij}$  – годинна ставка працівника  $i$ -ої спеціальності  $j$ -го тарифного розряду.

Середньо годинна ставка працівника може бути розрахована за такою формулою (4.2):

$$C_{ij} = \frac{C_{ij}^0(1+h)}{PЧ_i}, \quad (4.2)$$

де  $C_{ij}$  – основна місячна заробітна плата розробника  $i$ -ої спеціальності  $j$ -го тарифного розряду, грн.;

$h$  – коефіцієнт, що визначає розмір додаткової заробітної плати (при умові наявності доплат);

$PЧ_i$  - місячний фонд робочого часу працівника  $i$ -ої спеціальності  $j$ -го тарифного розряду, год. (приймаємо 168 год.).

Коефіцієнт  $h$ , який визначає розмір додаткової заробітної плати, для керівника та консультанта техніко-економічного розділу дорівнює 0,47.

Результати розрахунку записують до таблиці 4.2.

Таблиця 4.2 – Розрахунок витрат на оплату праці

№ п/п	Посада виконавців	Час розробки, год	Погодинна заробітна плата, грн/год.	Витрати на розробку, грн
1	Керівник ДП, старший викладач	16	64,3	1028,8
2	Консультант техніко-економічного розділу, доцент	2	59,9	119,8
3	Студент	144	8,33	1199,52
Разом				2348,12

Відрахування на соціальні заходи. Величну відрахувань у спеціальні державні фонди визначають у відсотковому співвідношенні від суми основної та додаткової заробітних плат. Згідно діючого нормативного законодавства сума відрахувань у спеціальні державні фонди складає 20,5% від суми заробітної плати:

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		54

$$B_{\Phi} = \frac{20,5}{100} \cdot 2348,12 = 481,36 \text{ грн.}$$

Розрахунок витрат на матеріали та комплектуючі.

Загальна сума витрат на матеріальні ресурси ( $B_M$ ) визначається за формулою (4.3):

$$B_M = \sum_{i=1}^n K_i \cdot C_i, \quad (4.3)$$

де  $K_i$  – витрата  $i$ -го типу матеріалу, натуральні одиниці вимірювання;

$C_i$  – ціна за одиницю  $i$ -го типу матеріалу, грн.;

$i$  – тип матеріального ресурсу;

$n$  – кількість типів матеріальних ресурсів.

Таблиця 4.3 – Зведені розрахунки матеріальних витрат

№ п/п	Найменування матеріальних ресурсів	Од. виміру	Факт. витрачено матеріалів	Ціна за одиницю, грн.	Сума, грн	Транспортні витрати (10% від суми)	Загальна сума, грн
	Допоміжна література	шт	1	1000	1000	100	1100
	Папір (формат А4)	уп	2	100	200	20	220
	Ручка кулькова	шт	2	10	20	2	22
	Олівець простий	шт	2	10	20	2	22
	Диски CD-R	шт	2	20	40	4	44
	Зошит, 96 арк	шт	1	50	50	5	55
	Тонер для принтера	уп	1	90	90	9	99
	Канцелярські маркери (синій, зелений)	шт	2	20	40	4	44
	Р а з о м						1606,00

Витрати на використання комп'ютерної техніки.

Витрати на використання комп'ютерної техніки ( $B_{KT}$ ) включають витрати на амортизацію комп'ютерної техніки, витрати на користування програмним забезпеченням, витрати на електроенергію, що споживається комп'ютером. За даними обчислювального центру ЗУНУ для комп'ютера типу IBM PC/ATX вартість години роботи становить 12 грн. Середній щоденний час роботи на комп'ютері – 2 години. Розрахунок витрат на використання комп'ютерної техніки приведений в таблиці 4.4.

Таблиця 4.4 – Розрахунок витрат на використання комп'ютерної техніки

№ п/п	Назва етапів робіт, при виконанні яких використовується комп'ютер	Час використання комп'ютера, год.	Витрати на використання комп'ютера грн.
1	Проведення досліджень та оформлення їх результатів у вигляді звіту	60	7200
2	Оформлення техніко-економічного розділу	8	96
3	Оформлення ДП	12	144
Разом		80	960

Накладні витрати.

Накладні витрати проектних організацій включають три групи видатків: витрати на управління, загальногосподарські витрати, невиробничі витрати. Вони розраховуються за встановленими відсотками до витрат на оплату праці. Середньостатистичний відсоток накладних витрат приймемо 150% від заробітної плати:

$$H = 1,5 \cdot 2348,12 = 3522,18 \text{ (грн).}$$

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		56

Інші витрати.

Інші витрати є витратами, які не враховані в попередніх статтях. Вони становлять 10% від заробітної плати:

$$I_B = 2348,12 \cdot 0,1 = 234,81 \text{ (грн)}.$$

Витрати на розробку програмного забезпечення складають:

$$K_1 = B_{OP} + B_{\Phi} + B_M + H + I_B + B_{KT},$$

$$K_1 = 2348,12 + 481,36 + 1606,00 + 3522,18 + 234,81 + 960,00 = 8849,47 \text{ (грн)}.$$

Витрати на відлагодження і дослідну експлуатацію програмного продукту визначаємо за формулою (4.4):

$$K_2 = S_{m.g.} \cdot t_{vid} \quad (4.4)$$

де  $S_{m.g.}$  – вартість однієї машино-години роботи ПК, грн./год;

$t_{vid}$  – комп'ютерний час, витрачений на відлагодження і дослідну експлуатацію створеного програмного продукту, год.

Загальна кількість днів роботи на комп'ютері дорівнює 30 днів. Середній щоденний час роботи на комп'ютері – 2 години. Вартість години роботи комп'ютера дорівнює 12 грн., тому  $K_2 = 12 \cdot 60 = 720$  грн.

#### 4.2 Визначення експлуатаційних витрат

Для оцінки економічної ефективності розроблювальної програмної системи слід порівняти її з аналогом, тобто існуючим програмним забезпеченням ідентичного функціонального призначення.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						57
Змн.	Арк.	№ докум.	Підпис	Дата		

Експлуатаційні одноразові витрати по програмному забезпеченню і аналогу включають вартість підготовки даних і вартість роботи комп'ютера (за час дії програми):

$$E_{\Pi} = E_{1\Pi} + E_{2\Pi},$$

де  $E_{\Pi}$  – одноразові експлуатаційні витрати на ПЗ (аналог), грн.;

$E_{1\Pi}$  – вартість підготовки даних для експлуатації ПЗ (аналог), грн.;

$E_{2\Pi}$  – вартість роботи комп'ютера для виконання проектного рішення (аналог), грн.

Річні експлуатаційні витрати  $B_{E\Pi}$  визначаються за формулою:

$$B_{E\Pi} = E_{\Pi} * N_{\Pi},$$

де  $N_{\Pi}$  – періодичність експлуатації ПЗ (аналог), раз/рік.

Вартість підготовки даних для роботи на комп'ютері визначається за формулою:

$$E_{1\Pi} = \sum_{l=1}^n n_i t_i c_i,$$

де  $i$  – категорії працівників, які приймають участь у підготовці відповідних даних ( $i=1,2,\dots,n$ );

$n_i$  – кількість працівників  $i$ -ої категорії, осіб.;

$t_i$  – трудомісткість роботи співробітників  $i$ -ої категорії по підготовці даних, год.;

$c_i$  – середнього годинна ставка працівника  $i$ -ої категорії з врахуванням додаткової заробітної плати, що знаходиться із співвідношення:

$$c_i = \frac{c_i^0 (1 + b)}{m},$$

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						58
Змн.	Арк.	№ докум.	Підпис	Дата		

де  $c_i^0$  – основна місячна заробітна плата працівника  $i$ -ої категорії, грн.;

$b$  – коефіцієнт, який враховує додаткову заробітну плату (прийmemo 0,57);

$m$  – кількість робочих годин у місяці, год.

Для роботи з даними як для проектного рішення так і аналогу потрібен один працівник, основна місячна заробітна плата якого складає:  $c = 6061$  грн.

Тоді:

$$c_1 = \frac{6061(1+0,57)}{22*8} = 56,64 \text{ грн/год}$$

Трудомісткість підготовки даних для проектного рішення складає 1 год., для аналога 1,5 год.

Таблиця 4.5 – Розрахунок витрат на підготовку даних та реалізацію проектного рішення на комп'ютері

№	Час роботи співробітників, год.	Середньогодинна заробітна плата, грн./год.	Витрати, грн.
	Проектне рішення		
1	1	56,64	56,64
	Аналог		
1	1,5	56,64	84,95

Витрати на експлуатацію комп'ютера визначається за формулою:

$$E_{2П} = t * S_{МГ}$$

Де  $t$  – витрати машинного часу для реалізації рішення (аналогу), год.;

$S_{МГ}$  – вартість однієї години роботи комп'ютера, грн./год.

$$E_{2П} = 1 * 12 = 12 \text{ грн.}; E_{2А} = 1,5 * 12 = 18 \text{ грн.}$$

$$E_{П} = 56,64 + 12 = 68,64 \text{ грн.}; E_{А} = 84,95 + 12 = 96,95 \text{ грн.}$$

$$B_{ЕП} = 68,64 * 252 = 17297,28 \text{ грн.}; B_{ЕА} = 96,95 * 252 = 24431,4 \text{ грн.}$$

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		59

Обчислення накладних витрат.

Накладні витрати пов'язані з обслуговуванням виробництва, утриманням апарату управління підприємства (фірми) та створення необхідних умов праці.

В залежності від організаційно-правової форми діяльності господарюючого суб'єкта, накладні витрати можуть становити 60–100 % від суми основної та додаткової заробітної плати працівників.

$$H_B = 0,7 * B_{OP} = 0,7 * (c_1 * 168), \quad (4.7)$$

де  $H_B$  – накладні витрати.

$$H_B = 0,7 * 9515,52 = 6660,86 \text{ грн.}$$

Складання кошторису витрат та визначення собівартості. Результати проведених розрахунків зведемо у таблицю 4.6.

Таблиця 4.6 – Кошторис витрат ( $B_{КС}$ )

№ п/п	Найменування витрат	Сума витрат, грн.
1	Витрати на оплату праці ( $B_{OP}$ )	2348,12
2	Відрахування у спеціальні державні фонди ( $B_{Ф}$ )	481,36
3	Витрати на матеріали та комплектуючі ( $B_M$ )	1606,00
4	Накладні витрати на розробку ( $H$ )	3522,18
5	Інші витрати ( $I_B$ )	234,81
6	Витрати на відлагодження і дослідну експлуатацію програмного продукту ( $K_2$ )	720
7	Накладні витрати експлуатацію ( $H_B$ )	6660,86
8	Річні експлуатаційні витрати ( $B_{EA}$ )	24431,4
Разом		40004,73



Розрахунок ціни проекту.

Договірна ціна ( $C_D$ ) для проектних рішень розраховується за формулою (4.8):

$$C_D = B_{КС} \cdot \left(1 + \frac{p}{100}\right), \quad (4.8)$$

де  $B_{КС}$  – кошторисна вартість, грн.;

$p$  – середній рівень рентабельності, % (приймаємо 20% за погодженням з керівником).

$$C_D = 40004,73 \cdot (1 + 0,2) = 48005,68 \text{ грн.}$$

#### 4.3 Визначення економічної ефективності та терміну окупності

Економічна ефективність ( $E_\phi$ ) полягає у відношенні результату виробництва до затрачених ресурсів:

$$E_\phi = \frac{\Pi}{B_{КС}}, \quad (4.9)$$

де  $\Pi = C_D - B_{КС}$  – прибуток, грн.;

$B_{КС}$  – кошторисна вартість, грн..

$$E_\phi = 8000,95 \text{ грн.} / 40004,73 \text{ грн.} = 0,2.$$

Поряд із економічною ефективністю розраховують термін окупності капітальних вкладень ( $T_p$ ):

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						61
Змн.	Арк.	№ докум.	Підпис	Дата		

$$T_P = \frac{1}{E_P} . \quad (4.10)$$

Тобто:  $T_P = 1/0,2 = 5р.$

Прийнятним вважається термін окупності близький до 7 років.

Розраховані економічні показники проекту занесемо до таблиці 4.7.

Таблиця 4.7 – Економічні показники розробки

№ п/п	Показник	Значення
1.	Собівартість, грн.	40004,73
2.	Плановий прибуток, грн.	8000,95
3.	Ціна, грн.	48005,68
4.	Економічна ефективність	0,2
5.	Термін окупності, рік	5

Враховуючи основні економічні показники з таблиці 4.7, можна зробити висновок, що при економічній ефективності 0,2 та терміні окупності – 5 роки проводити роботи по впровадженню даного програмного додатку є доцільним та економічно вигідним.

## ВИСНОВКИ

На основі аналізу сучасних алгоритмів аналізу та перекладу текстів, систем автоматичного перекладу текстових даних та проведених експериментів, можна зробити наступні висновки:

1) Проведено аналіз та класифікацію інформації, а також досліджено канали передачі даних, що дозволило визначити основні характеристики та принципи обміну інформацією в глобальних мережах;

2) Проаналізовано існуючі технології машинного перекладу текстових даних на різні мови світу, що дозволило виділити статистичний підхід як один з перспективних та швидких вирішень поставленої задачі.

3) Здійснено аналітичний огляд існуючих програмних систем для обробки та машинного перекладу текстової інформації, що надало можливість виділити основні структурні складові таких систем, а також їх переваги та недоліки.

4) Проведено дослідження алгоритмів статичного поділу текстової інформації на окремі складові, що дозволило розділити вхідне текстове повідомлення на окремі слова та словосполучення, а також формувати запити для онлайн сервісу перекладу на основі визначених шаблонів.

5) Розроблено алгоритм машинного перекладу текстової інформації за допомогою API-інтерфейса онлайн перекладача, що дозволило спроектувати та реалізувати програмний додаток перекладу текстів;

6) Розроблена структура програмної системи на основі модульного підходу забезпечує максимальну ефективність та простоту подальшого розвитку системи як в напрямку збільшення функціональних модулів, так і збільшення кількості алгоритмів, що реалізовані в них.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		63

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Сергійчук В.В., Жаборинська І.В. Алгоритм машинного перекладу текстових даних на основі онлайн-сервісів веб-браузерів. Збірник тез IV Науково-практична конференція молодих вчених і студентів «Інтелектуальні комп'ютерні системи та мережі», Тернопіль, 02 червня 2021 р. с. 28.

2. Методичні рекомендації до виконання кваліфікаційної роботи з освітнього ступеня “Бакалавр” спеціальності 123 «Комп'ютерна інженерія» галузі знань 12 Інформаційні технології / О.М. Березький, Л.О.Дубчак, Г.М. Мельник, Ю.М. Батько / Під ред. О.М. Березького. Тернопіль: ЗУНУ, 2020. 60с.

3. Методичні вказівки до виконання практичних робіт з дисципліни «Техніко-економічне обґрунтування розробки комп'ютерних систем»/ Н.Я. Савка, І.Р. Паздрій / Під ред. О.М. Березького. Тернопіль: ТНЕУ, 2019. 40 с.

4. Методичні вказівки до оформлення курсових проектів, звітів про проходження практики, випускних кваліфікаційних робіт для студентів спеціальності «Комп'ютерна інженерія» / І.В. Гураль, Л.О. Дубчак / Під ред. О.М. Березького. Тернопіль: ТНЕУ, 2019. 33 с.

5. Бартлетт Н. Программирование на Delphi Путеводитель. The Coriolis Group, Inc., 1996, Издательство НИПФ "Диасофт Лтд.", 2016. 116с.

6. Вебер Дж. Технология C++ в подлиннике. QUE Corporation, 2016, "ВНУ-Санкт-Петербург", 2017. 256с.

7. Волш А. И. Основы программирования на C++ для World Wide Web. IDG Books Worldwide, Inc., 1996, Издательство "Диалектика", 2016. 458с.

8. Марков А. С. «Базы данных. Введение в теорию и методологию. Финансы и статистика». 2016. С.24-35.

9. Абрамов С. А. Задачи по программированию. М.: Наука, 2018. 256с.

10. Березин Б.И., Начальный курс Delphi. М.: ДИАЛОГ-МИФИ, 2016. 331с.

11. Бондарев В.М. Основы программирования. Харьков: Фолио, Ростов н/Д: Феникс, 2017. 446с.

12. Вирт Н. Алгоритмы и структуры данных. М.: Мир, 2019. 345с.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
						64
Змн.	Арк.	№ докум.	Підпис	Дата		

- 13.Гладков В. П. Задачи по информатике на вступительном экзамене в вуз и их решения: Учебное пособие. Пермь: Перм. техн. ун-т, 2014. 516с.
- 14.Грогоно П. Программирование на языке Delphi. М.: Мир, 2012. 216с.
- 15.Дагене В.А. 100 задач по программированию. М.: Просвещение, 2013. 106с.
- 16.Джамса К. Библиотека программиста Java. Jamsa Press, 2016, ООО "Попурри", 2016. 656с.
17. Марков А. С. «Базы данных. Введение в теорию и методологию. Финансы и статистика». 2016. Р. 24-35.
- 18.Заварыкин В.М. Основы информатики и вычислительной техники. М.: Просвещение, 2019. 556с.
- 19.Касаткин В. Н. Информация. Алгоритмы. ЭВМ. М.: Просвещение, 2011. 219с.
- 20.Кен А. Язык программирования Delphi. Addison-Wesley Longman,U.S.A.,1996, Издательство "Питер-Пресс", 2017. 378с.
- 21.Керниган Б. Язык программирования Delphi. Пер. с англ. М.: Финансы и статистика, 2012. 391с.
- 22.Ляхович В.Ф. Руководство к решению задач по основам информатики и вычислительной техники. М.: Высшая школа, 2014. 127с.
- 23.Мейнджер Дж. Delphi Основы программирования. McGraw-Hill,Inc.,1996, Издательская группа ВНУ, Киев, 2017. 346с.
- 24.Миков А. И. Информатика. Введение в компьютерные науки. Пермь: Изд-во ПГУ, 2018. 442с.
- 25.Могилев А. В. Информатика: Учеб. пособие для студ. пед. Вузов. М.: Изд. центр «Академия», 2019. 629с.
- 26.Нотон П. JAVA:Справ.руководство. М.:БИНОМ:Восточ.Кн.Компания, 2016: Восточ.Кн.Компания. 447с.
- 27.Нотон П. Полный справочник по Java. McGraw-Hill,1997, Издательство "Диалектика",2017. 556с.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		65

28.Ренеган Э.Дж. 1001 адрес WEB для программистов :Новейший путеводитель программиста по ресурсам World Wide Web:Пер.. Минск:Попурри, 2017. 512с.ил.

29.Родли Дж. Создание Java-апплетов. The Coriolis Group,Inc.,1996, Издательство НИПФ "ДиаСофт Лтд.", 2016. 466с.

30.Секреты программирования для Internet на Java. Ventana Press, Ventana Communications Group, U.S.A., 2016, Издательство "Питер Пресс", 2017. 396с.

31.Семакина И. Г. Информатика. Задачник-практикум: В 2 т.. М.: Лаборатория Базовых Знаний, 2019. 476с.

32.Сокольский М.В. Все об Intranet и Internet. М.:Элиот, 2018. 254с.

33.Тассел Д. Стил, разработка, эффективность, отладка и испытание программ. М.: Мир, 2011. 56с.

34. Тюрин Ю.Н. Анализ данных на компьютере. М.: ИНФРА-М, Финансы и статистика, 2015. 384с.

35.Флэнэген Д. Java in a Nutshell. O'Reilly & Associates, Inc., 1997, Издательская группа BHV, Киев, 2018. 473с.

36.Чен М.С. Программирование на C++:1001 совет:Наиболее полное руководство по Java и Visual J++ :Пер.с англ. Минск:Попурри, 2017. 640с.ил.

37.Эферган М. C++: справочник. QUE Corporation, 2017, Издательство "Питер Ком", 1998. 256с.

38.G. Yang. «Human face detection in a complex background. Pattern Recognition », 27 (1): 2014. P.53-63.

39.Kotropoulos C. «Acoustics, Speech, and Signal Processing», 2017. ICASSP-97, 2017. IEEE International Conference on pp.2537-2540 v. 4

40.Leung TK. «Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching» 2015.p P.83-95.

41.Yow KC. Feature-based human face detection. Image and vision computing 15 (9), 2017. P.713-735.

42.Sinha, P. Perceiving and Recognizing threedimensional forms. PhD thesis, Massachusetts Institue of Technology, 2016. 278p.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		66

43.Lanitis, A «Image Anal. Classifying variable objects using a flexible shape model »Image Processing and its Applications, 2015., P.70-74.

44.Viola P. «Rapid Object Detection using a Boosted Cascade of Simple Features», proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2011., vol. 1, 518p.

45.Jones MJ. «Robust real-time face detection», International Journal of Computer Vision, vol. 57, no. 2, 2014., P.137-154.

46.Buchatskiy AN. «Selection of the Optimal Color Space for Reducing False Positives Rate in the Viola-Jones Method», Актуальні проблеми інфотелекомунікацій в науці та освіті, II Міжнародна науково-технічна та науково-методична конференція. Санкт-Петербург, 2013.

47.Ethan R. ORB: an efficient alternative to SIFT or SURF. Computer Vision (ICCV), IEEE International Conference on. IEEE, 2011. P. 2564–2571.

48.Stefan L. BRISK: Binary Robust Invariant Scalable Keypoints. Computer Vision (ICCV), 2011. P. 2548–2555.

49.Pablo F. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In British Machine Vision Conference (BMVC), 2013.

50.Martin A. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. Of the ACM24: 2001. P.381–395,.

51.Патин М.В. Сравнительный анализ методов поиска особых точек и дескрипторов при группировке изображений по схожему содержанию. Молодой ученый. 2016. №11. С. 214-221.

52.Гонсалес Р. Цифровая обработка изображений в среде MATLAB //М.: Техносфера. 2016. 616с.

					КР.КІ.07097/17.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		67