

ОБГРУНТУВАННЯ ВИБОРУ АЛГОРИТМУ МАШИННОГО ПЕРЕКЛАДУ ПРИ РОЗРОБЦІ ДОДАТКУ ДВОХСТОРОННЬОЇ КОМУНІКАЦІЇ

Вінницький національний технічний університет

Анотація

У даній роботі обгрунтовано використання системи GNMT від Google при розробці додатку двосторонньої комунікації, яка працює на основі нейронного навчання та забезпечує швидкий та точний переклад з використанням понад 100 мов, на підставі порівняльного аналізу алгоритмів машинного перекладу.

Ключові слова: алгоритм, мобільний додаток, аналіз, алгоритм машинний перекладу

Abstract

In this article substantiates the use of Google's GNMT system in the development of a two-way communication application that works on the basis of neural learning and provides fast and accurate translation using more than 100 languages, based on a comparative analysis of machine translation algorithms.

Keywords: algorithm, mobile application, analysis, machine translation algorithm

Подорожувати країнами є дуже популярним і цікавим заняттям. Сьогодні поїхати до іншої країни стало простіше, ніж вивчати нову мову. Саме тому не рідкість ситуація, коли людина приїжджає в іншу країну без знань іноземної мови, що досить ускладнює подорож [1]. Через дану проблему стає актуальною розробка мобільного додатку, що зможе у форматі голосового вводу автоматично перекладати повідомлення мовою, яку розуміє співрозмовник. Ключовим фактором стає простота, зручність використання та його доступність на різних платформах. Завдяки цьому додаток можна буде використовувати без обмежень в будь-який момент часу.

Основні етапи роботи програми: отримання вхідних даних, їх обробка та реалізація перекладу за використання алгоритмів машинного перекладу, обробка і відтворення кінцевого результату. Під час розробки, а саме на етапі роботи з вхідними даними, головною метою є реалізація перекладу високої якості і при найменших при цьому затратах у часі, забезпечивши при цьому велику кількість мов для вибору. Для цього уже створено багато готових рішень, які у собі використовують комбінації самих ефективних алгоритмів машинного перекладу. Одними з самих популярних для перекладу є наступні сервіси: PROMPT.One, Google Translate, Microsoft Translator [2].

У PROMT використовується гібридний машинний переклад, який поєднує статистичний машинний переклад і переклад на основі нейронних мереж. У першому випадку навчання відбувається на основі значних за обсягом корпусів паралельних текстів – і якість перекладу сильно залежить від якості і кількості даних, результат відрізняється більшою гладкістю, ніж переклади, виконані на основі аналітичного підходу. Нейронний підхід відрізняється високою семантичною якістю перекладу,

близькому до перекладу, виконаному людиною, можливістю перекладу контенту, який відсутній в навчальних даних і можливістю навчання системи на переклад спеціальної термінології [3].

Перевагами є точність перекладу завдяки поєднанню алгоритму статичного перекладу і на основі нейронних мереж.

Недоліками PROMT є: можливість перекладу лише для 20 мов, що є відносно малим показником, невисока швидкість перекладу через комбінування машинних алгоритмів, відсутність підтримки перекладу складних фраз, термінів фразеологізмів [4].

Система GNMT (Google's Neural Machine Translation) всередині Google Перекладача використовує велику штучну нейронну мережу, придатну для глибинного навчання. Система GNMT є само навчаючою, що забезпечує якісний переклад. Кінцевий результат формується з максимально можливою відповідністю граматиці людської мови.

Архітектура системного навчання GNMT випробувана на більш, ніж ста мовах, що підтримуються Google Перекладачем. Завдяки потужній внутрішній структурі, система з часом вчиться створювати більш якісні і природні переклади. GNMT здатна перекладати речення в цілому, а не частинами. GNMT здатна виконувати інтерлінгвістичний машинний переклад, кодуючи семантику речення замість того, щоб запам'ятовувати переклади окремих фраз [5].

Перевагами є самий широкий вибір мов — більше 100, понад 2250 вузьких напрямків, високий рівень та швидкість перекладу.

Недоліком є відсутність підтримки перекладу складних фраз, термінів фразеологізмів [6].

Для моделей перекладу Microsoft Translator характерно таке:

1) Нейронна мережа намагається перекладати як працює мозок людини. На високому рівні, нейронна мережа здійснює переклад в два етапи. На першому етапі моделюється слово, що треба перекласти на основі контексту цього слова в межах речення. По-друге, нейронна мережа перекладає модель цього слова, в межах контексту речення, в іншу мову.

2) SMT на базі синтаксису полягає в перекладі синтаксичних одиниць більше ніж переклад слів або рядків. Microsoft використовував синтаксичний SMT щоб перекласти тексти пов'язані з комп'ютерними текстами з англійської на інші мови. Постійні дослідження в цій сфері привели до великих поліпшень у відображенні слів та у перестановці слів.

3) SMT на основі фраз, машина вчиться знаходити відповідності перекладу у корпусі паралельних текстів співставлення речення до речення між двома мовами, без потреби у лінгвістичному поясненні оброблюваних даних.

Перевагами є точність перекладу середнього рівня.

Недоліками є відносно мала кількість мов — близько 60, менша швидкість через комбінування алгоритмів машинного перекладу та неможливість перекладу складних фраз, термінів фразеологізмів [7].

Порівняльна характеристика сервісів з використанням алгоритмів машинного перекладу наведена в таблиці 1.

Таблиця 1 – Порівняльна характеристика сервісів з використанням алгоритмів машинного перекладу

	Точність	Швидкість	Розмір мовної бази

PROMT	Висока	Середня	20+
Система GNMT	Висока	Висока	100+
Microsoft Translator	Середня	Середня	60+

Отже, для реалізації машинного перекладу на етапі роботи з вхідними даними програми доцільно використати систему GNMT від Google, яка працює на основі нейронного навчання та забезпечує швидкий та точний переклад з використанням понад 100 мов.

Список використаних джерел:

1. Карманные полиглоты [Електронний ресурс] – Режим доступу: <https://www.itworld.ru/tech/practice/120118.html>
2. OFFLINE TRANSLATION – BEST TRANSLATION APPS FOR ANDROID, IOS AND DESKTOP [Електронний ресурс] – Режим доступу: <https://ehlion.com/magazine/offline-translation-for-android-ios-and-desktop/>
3. PROMT - [Електронний ресурс] – Режим доступу: <https://ru.wikipedia.org/wiki/%D0%9F%D0%A0%D0%9E%D0%9C%D0%A2>
4. Обзор онлайн-переводчиков [Електронний ресурс] – Режим доступу: https://www.primavista.ru/rus/articles/online_translators_part3
5. Нейронный машинный переклад Google [Електронний ресурс] – Режим доступу: https://uk.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B8%D0%B9_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%B8%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%BA%D0%BB%D0%B0%D0%B4_Google
6. Машинный перевод и не только: гид по онлайн-переводчикам [Електронний ресурс] – Режим доступу: <https://vc.ru/ml/132172-mashinnyu-perevod-i-ne-tolkogid-po-onlayn-perevodchikam>
7. Google Trans-late vs Microsoft Trans-la-tor: Which Trans-la-tion App Should You Use [Електронний ресурс] – Режим доступу: <https://www.guidingtech.com/google-translate-vs-microsoft-translator-comparison/>

Савчук Тамара Олександрівна — PhD, професор кафедри комп'ютерних наук, професор, Вінницький національний технічний університет, м. Вінниця, e-mail: savchtam@gmail.com.

Гриценюк Олесь Васильович— студент групи 2-КН-176 кафедри комп'ютерних наук факультету ВНТУ, Вінницький національний технічний університет, м. Вінниця, e-mail: ruboken@vntu.edu.ua

Savchuk Tamara Oleksandrivna. — PhD, Professor of the Computer Sciences Chair, professor, Vinnytsia National Technical University, Vinnytsia, e-mail: savchtam@gmail.com.

Oles V. Hrytseniuk — student of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia, e-mail: ruboken@vntu.edu.ua