

О.Є. Поморцева, С.М. Кобзан

Харківський національний університет міського господарства імені О.М. Бекетова, Україна

ДОСЛІДЖЕННЯ ПРОБЛЕМИ РОЗПІЗНАННЯ МОВЛЕННЯ ДЛЯ ВИРІШЕННЯ СПЕЦІАЛЬНИХ ЗАВДАНЬ

Метою статті є проведення дослідження на тему машинного розшифрування інформації з аудіо або відеофайлів у текстову форму. У результаті досліджень виявлено сильні і слабкі сторони програм, які застосовуються для цих цілей. Зроблено висновок що для виконання розшифровки аудіо файлів та автоматичного переведення їх у текстовий формат перспективним напрямком є застосування сервісів із вбудованою функцією штучного інтелекту. Перспективним є застосування таких систем для спеціальних підрозділів Збройних сил України.

Ключові слова: транскрибація, тайм-код, розшифровка мови, геолокація, база даних, геоінформаційна система.

Постановка проблеми

В даний час розшифровка інформації з аудіо або відео у текстову форму (транскрибація) вкрай актуальна. Це може бути необхідним для людей з обмеженими фізичними можливостями або захворюваннями або для тих, хто воліє споживати контент у беззвучному режимі, отримуючи інформацію у письмовій формі. Особливу актуальність транскрибація має в даний час (в умовах проведення бойових дій) в Україні для вирішення спеціальних завдань щодо відстеження, пошуку та виявлення певного контенту, який передається різними засобами зв'язку в розмовах при обробці аудіофайлів.

Найкращу якість тексту дає ручне розшифрування, проведене професіоналом. Щоб провести транскрибацію потрібно мати хороший слух, грамотність, ерудицію, вміння працювати з різними програмними засобами. Тому, дуже перспективно використовувати сучасні програмні засоби, накопичувати та обробляти інформацію за допомогою сучасних баз даних та, можливо, із застосуванням геоінформаційних систем [1, 2] у випадках, коли необхідна прив'язка до місцевості з використанням атрибутивної інформації та шарової структури системи [3].

Аналіз останніх досліджень і публікацій

Найчастіше транскрибацію виконують для того, щоб у загальних рисах мати уявлення про те, про що йшлося в записі, отримати розшифровку лише певних моментів запису, перевести звук у текстовий вигляд із серйозними коригуваннями (прибрати слова-паразити, заїкання, безграмотно побудовані речення замінити більш коректними) [4].

Суть транскрибації полягає у обробці та автоматизації роботи з живою мовою. Головна перевага транскрибації мови за допомогою людей – це дотримання всіх орфографічних та пунктуаційних норм або розмітка додаткових метаданих. Автоматичні алгоритми досі допускають у цьому помилки, але їхня основна перевага – є час (або синхронність) транскрибації. Що стосується вирішення спеціальних завдань, де час – це самий вирішальний чинник [5].

Для того щоб провести класифікацію видів транскрибації, необхідно окремо виділити основні завдання розшифрування:

- розібрати все сказане спікерами.
- розбити промову на речення.
- прибрати слова-паразити, багатомовність та паузи.

Таким чином, можемо провести класифікацію видів транскрибації в залежності від сфери застосування та їх особливостей (табл. 1).

Таблиця 1

Види транскрибації та їх особливості

Вид транскрибації	Особенности
1	2
Надиктування	Автор читає підготовлений текст. Текст може бути використаний для статей, постів, ведення блогу або написання матеріалу в журнал. Просте розшифрування.

Продовження таблиці 1

1	2
Лекція	Студенти записують промову викладача на диктофон для подальшого розшифрування лекцій, у тексті яких зустрічаються складні терміни та незнайомі визначення. Відрізняється від диктування лише швидкістю промови оратора. Легкий вид розшифрування. Виникають труднощі у вигляді складної термінології.
Подкаст	Формат радіо, у якому рідко трапляються проблеми з промовою. Може бути монологом чи діалогом. Розшифровка потрібна для відображення субтитрів
Інтерв'ю	Формат, у якому бере участь дві або більше осіб. Розшифровка потрібна для субтитрів у відео, написання статті чи посту. Жорстка дословність не потрібна. Найпростіший вид транскрибації.
Сінхрон	Простий формат транскрибації. Відрізняється від інтерв'ю тим, що в ньому слід проставляти мітки (тайм-коди).
Вебінар	Мова одного чи кількох експертів. Розшифровка схожа на формат подкасту. Транскрибацію використовують для субтитрів, написання статті чи книги.
Фокус-група	Використовується в маркетингових дослідженнях та соціологічних опитуваннях. Складний формат розшифровки. Розшифровка потрібна для звітів, аналізу та досліджень. Складність розшифровки зростає пропорційно до тривалості, а також кількості учасників.
Конференція, круглий стіл	У конференції є ведучий та спікери. Розшифровка використовується для субтитрів, написання статей та постів. Складний тип транскрибації, особливо коли круглий стіл засідає без керівника і кожен мже перебивати іншого.
Телефоні розмови	Розшифровка використовується для аналізу дій клієнтів, збирання даних про них, виявлення протиправних настроїв. Використовується не лише для бізнесу, а й для вирішення специфічніших завдань (силовики можуть прослуховувати розмови потенційних порушників закону).
Субтитри	Використовуються у відео. Текст повинен точно повторювати мову того, хто говорить. Потрібно не просто перекладати мову в текстовий вигляд, але ще й проставляти у спеціальному форматі, тайм-код початку та кінця кожної репліки.
Монтажний лист	Застосовується в кіно та кліпах. У таблиці вказані кадри з репліками героїв та технічними коментарями. Потрібно не лише розписати промову персонажа, а й зафіксувати все, що зараз перебуває в кадрі.
Прослуховування	Цей тип транскрибації найскладніший. Запис йде на приховану камеру або захований диктофон – якість звуку не завжди задовільна.
Суд	Один із найскладніших типів транскрибації. Потрібно розбирати кожне слово, в обов'язковому порядку тих, хто говорить, потрібно ідентифікувати. Не допускається ні яке перекручування змісту.

Виклад основного матеріалу

Після проведеної у роботі класифікації видів транскрибації, можна визначити основні, найчастіше використовувані та ефективні програмні

продукти, які дозволяють якісно вирішувати завдання розпізнавання мови. Важливі особливості програм, за допомогою яких можна виконувати транскрибацію, представлені в таблиці 2.

Таблиця 2

Програмне забезпечення для транскрибації

Програма	Регістрація	Робота з файлами	Безкоштовно вне ПЗ	Багатокористувальницька	Обмеження по розміру файла	Онлайн сервіс	Якість перекладу
1	2	3	4	5	6	7	8
Google документ	–	–	+	+	ні	+	не стабільний
Google Translate	–	–	+	–	5000 символів	+	не стабільний

Продовження таблиці 2

1	2	3	4	5	6	7	8
Vocalmatic	+	+	–	–	ні	–	не стабільний
Speechlogger	+	–	– безкоштовна α -версія	–	ні	–	добре, штучний інтелект
Dictation	–	–	–	–	ні	–	не стабільний
Voicebot	–	–	–	–	ні	–	відмінне, штучний інтелект
Voco	–	+	–	–	ні	–	відмінне, самонавчається
InqScribe	–	–	–	–	ні	–	відмінне
Speechpad	–	+	–	–	ні	+ –	добре, ручний режим
RealSpeaker	–	–	–	–	ні	–	не стабільне
Express Scribe	–	+	–	–	ні	–	відмінне
VoiceIn Voice Typing	+	+	–	–	ні	+	добре
Transcribe	+	+	+, пробна версія тиждень	–	ні	+	відмінне, штучний інтелект
Transcribe by Wreally	+	+	–	–	ні	+	відмінне
Алгоритм YouTube	–	–	+	–	ні	+	не стабільний
Dragon Dictation	+	+	–, тестовий період	–	ні	+	відмінне, штучний інтелект
Speechnotes	–	–	–	–	ні	+	
Перекладач від «Яндекса»	–	–	+	–	10000 символів	+	не стабільний
Rev	–	+	–	–	ні	+	відмінне, штучний інтелект
Temi	+	+	–	–	ні	+	відмінне

Для підвищення точності машинного навчання, важливим є правильне анотування. Наприклад, якщо у записі більше одного спікера, важливо не лише точно перевести запис до тексту, а й визначити спікерів. Без цього нейромережа зможе переводити голос у текст, але не зможе відстежити зміну спікера. Це може бути критичним недоліком при вирішенні певних завдань щодо відстеження, пошуку та виявлення певного контенту в телефонних розмовах та аудіофайлах. При вирішенні подібного роду завдань однією з важливих характеристик буде підтримка кількох мов, тому що нерідко розмови ведуться з

використанням двох і більше мов або окремих слів з іншої мови (так званий суржик).

Для підвищення точності одержуваного за допомогою програми-транскрибатора тексту необхідні терабайти чітко анотованих даних [6]. Програми, що володіють штучним інтелектом окрім вилучення сутностей для розуміння сенсу мови, дозволяють розпізнавати і розуміти форму: комбінації звуків, літер, складів, які складаються в слова та речення. Тільки так машина зможе коректно та правильно декодувати людську мову (рис. 1).

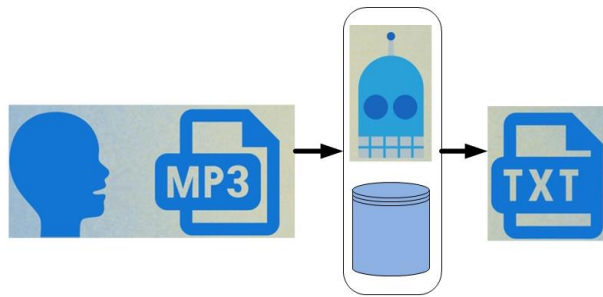


Рис. 1. Схема процесу транскрибації з використанням програм, що використовують штучний інтелект та бази даних

Також значний інтерес становить визначення місцезнаходження спікера – геолокація, можливо, навіть з визначенням конкретного об'єкта нерухомості [1, 2, 7]. Це може використовуватися для збору даних та подальшого аналізу настроїв громадськості та швидкого реагування з подальшою локалізацією протиправних дій.

Висновки

1. Авторами проведено дослідження існуючих сервісів за допомогою яких можна виконати транскрибацію, виділено позитивні та негативні сторони для кожного з сервісів.
2. Види траскрибації розглянуті та класифіковані в роботі залежно від сфери їх застосування.
3. Окремо виділені задачі для розшифрування та особливості розшифрування.
4. Перспективним напрямом транскрибації, особливо під час вирішення спеціальних завдань, є використання не просто готових сервісів, а застосування сервісів із вбудованою функцією штучного інтелекту, так званих систем, що самонавчаються.

Література

1. Кобзан С., Поморцева О. Дослідження ринку нерухомості за допомогою ГІС. Тенденції та перспективи розвитку //Збірник наукових праць ЛОГОС. – 2021. – том 3. С. 151-156.
2. Кобзан С. М. Формування ринку нерухомості: практичні аспекти та особливості оцінки. Монографія. 2019. – 267 с.
3. Петрушина В., Бойцова М., Кобзан С. Операції с нерухомістю. 2008. – 678 с.
4. Сайт SendPulse – Програми для транскрибації, которые помогут расшифровать речь в текст [Електронний ресурс]. –2022. Режим доступу: <https://sendpulse.com/ru/blog/transcription-software>.
5. Сайт - Перспективи забезпечення воєнної кампанії 2023 року: український погляд, В. Залужний, М. Забродський [Електронний ресурс]. 2022. Режим доступу: <https://www.ukrinform.ua/rubric-ato/3566162-ak-zabezpeciti-voennu-kampaniu-u-2023-roci-ukrainskij-poglad.html>

6. Толстохатко, В. А., Поморцева, О. Є., Патракеєв І. М. Бази даних: проектування та використання для обліку нерухомого майна: навч. посібник. Харк. нац. ун-т міськ. госп-ва ім. О.М. Бекетова, 2014. – 176 с.
7. Olena Pomortseva, Sergiy Kobzan, O. Voronkov, A. Yevdokimov. Geospatial Modeling of the Infrastructure Facility Optimal Location / Second International Conference on Sustainable Futures: Environmental, Technological, Social, and Economic Matters, which held at Kryvyi Rih National University, Kryvyi Rih, Ukraine, on May 19-21 2021. EasyChair . №5537.

References

1. Kobzan, S., Pomortseva, O. (2021). Real estate market research using GIS. Trends and prospects for development. *Collection of scientific works ЛОГОС 2021*, vol. 3. pp. 151 – 156. Oxford, United Kingdom.
2. S. M. Kobzan. (2019). Real estate market formation: practical aspects and features of evaluation. Monograph / Kyiv: Yurinkom Inter. - 267 p.
3. Petrushina V., Boitsova M., Kobzan S. (2008). Real estate transactions. Factor Publishing House. - 678 p.
4. Transcription software to help transcribe speech into text. Retrieved from: <https://sendpulse.com/ru/blog/transcription-software>.
5. Prospects for ensuring the military campaign of 2023: the Ukrainian view, V. Zaluzhnyi, M. Zabrodskyi. Retrieved from: <https://www.ukrinform.ua/rubric-ato/3566162-ak-zabezpeciti-voennu-kampaniu-u-2023-roci-ukrainskij-poglad.html>.
6. Tolstokhatko, V. A., Pomortseva, O. E., Patrakeev, I. M. (2014) Databases: design and use for real estate accounting: training. manual. Hark. national city university farm named after OHM. Beketov - 176 p.
7. Olena Pomortseva, Sergiy Kobzan, O. Voronkov, A. Yevdokimov (2021). Geospatial Modeling of the Infrastructure Facility Optimal Location / Second International Conference on Sustainable Futures: Environmental, Technological, Social, and Economic Matters, which held at Kryvyi Rih National University, Kryvyi Rih, Ukraine, on May 19-21. EasyChair . №5537.

Рецензент: доктор економічних наук, професор К.А. Мамонов, Харківський національний університет міського господарства імені О.М. Бекетова, Україна

Автор: ПОМОРЦЕВА Олена Євгенівна

кандидат технічних наук, доцент, доцент кафедри Земельного адміністрування та ГІС
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – elenapomor7@gmail.com
ID ORCID: <https://orcid.org/0000-0002-4746-0464>

Автор: КОБЗАН Сергій Маркович

кандидат технічних наук, доцент, доцент кафедри Земельного адміністрування та ГІС
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – s.kobzan@gmail.com
ID ORCID: <https://orcid.org/0000-0002-5257-8117>

RESEARCH OF THE PROBLEM OF SPEECH RECOGNITION FOR SOLUTION OF SPECIAL TASKS

O. Pomortseva, S. Kobzan

O.M. Beketov National University of Urban Economy in Kharkiv, Ukraine

In the article, the authors conducted a study of the actual problem of machine translation of information from audio or video files into text form (transcription). This is necessary for people with limited physical capabilities, or diseases or for those who need to process information in the form of a text file. The process of transcription is relevant at present (in the conditions of hostilities). Today in Ukraine, transcription is necessary to solve complex special tasks. Namely, solving the task of searching and identifying certain content that is transmitted by various means of communication in conversations in the form of audio files. Such tasks are currently quite relevant and quite time-consuming and take same time.

To solve this problem, the authors conducted a study and identified the strengths and weaknesses of the programs that are often used for these purposes. The types of transcription and the software currently used are presented in separate tables with all their features. Existing automatic language transcription algorithms still make significant errors, but their main advantage is time (or synchronicity). When it comes to solving special tasks, time is the most decisive factor.

Terabytes of clearly annotated data are needed to increase the accuracy of the text received by the transcriber program. Programs with artificial intelligence, in addition to extracting essences to understand the meaning of language, allow us to recognize and understand the form: combinations of sounds, letters, and syllables that are built into words and sentences. Only in this way will the machine be able to decode human speech correctly and correctly. An extremely important task is to determine the location of the speaker - geolocation, even with the determination of the specific location of the real estate object. This can be used for data collection and subsequent analysis of public sentiment and rapid response with subsequent localization of illegal activities. In the article, the authors concluded that for decoding audio files and automatically converting them into text format, a promising direction is the use of not just ready-made services, but the use of services with a built-in artificial intelligence function, so-called self-learning systems.

Keywords: *scription, time code, language decoding, geolocation, database, geographic information system.*