

4. Сагаровський А. А. Лексика Центральної Слобожанщини (Харківщини) як об'єкт словництва. *Сучасні проблеми мовознавства та літературознавства*. Ужгород, 2001. В. 4. С. 465-469.
5. Сагаровський А. А. Фрагмент діалектного словника Харківщини. *Волинь-Житомирщина*. Житомир, 2001. № 6. С. 249-254.
6. Сікорська З. С. Про особливості Словника діалектної лексики Луганщини. *Волинь-Житомирщина*. Житомир, 2005. № 14. С. 192-196.
7. Словник українських східнослобожанських говірок / К. Глуховцева, В. Леснова, І. Ніколаєнко, Т. Тернавська, В. Ужченко. Луганськ, 2002. 234 с.
8. Хобзей Н. В. Сучасне українське словництво. *Україна: культурна спадщина, національна свідомість, державність*. Львів, 2006. № 15. С. 819-826.

**Валентина Робейко, Микола Сажок, Руслан Селюх, Дмитро Федорин,  
Олександр Юхименко  
(м. Київ)**

### **МОНІТОРИНГ НОВИНИХ YOUTUBE-КАНАЛІВ ЗА ДОПОМОГОЮ СИСТЕМИ РОЗПІЗНАВАННЯ МОВЛЕННЯ**

Оперативний моніторинг телерадіопрограм є важливим суспільним завданням, що має на меті здійснення запису, зберігання та аналізу аудіовізуальних даних медіа-сервісів на всій території України. На сьогодні велика кількість світових телеканалів ведуть своє мовлення засобами веб-сервісу YouTube. Цей сервіс дає змогу користувачам Інтернету переглядати трансляції як наживо, так і окремі випуски телепередач або сюжетів у записі. Українські телеканали віднедавна активно опановують новий спосіб донесення медійної інформації до користувача. В недалекому минулому провести аудіовідеозапис трансляцій телеканалів можливо було через антену або з місцевих сервісів цифрового ТБ. Тепер же до переліку національних новинних каналів, що максимально використовують можливості YouTube, входять канали *Еспресо*, *5 канал*, *112*, *24*, *Zik* та ряд інших. Крім того окремі передачі новин, такі як *ТСН*, *Факти ICTV* тощо, створюють власні YouTube-канали, викладають туди записи своїх випусків та інколи транслюють їх наживо.

Перша з розроблених в Україні систем, що давала змогу проводити оперативний моніторинг аудіоінформації, забезпечувала надійне перетворення на текст записів спонтанного мовлення, в яких наявні лише незначні шуми і завади — [2]. Сюжети, знайдені за ключовими словами, користувач переглядав синхронно з розпізнаним текстом із метою аналізу або редагування.

На сьогодні рівень розробок у галузі розпізнавання мовлення

забезпечує 70–90 і більше відсотків послівної надійності розпізнаного тексту — залежно від наявності в аудіозаписі шумів, завад, а також одночасного мовлення кількох дикторів, мовлення кількома мовами [1].

Система перетворення мовлення на текст, що реалізована авторами для оперативного моніторингу звукової медійної інформації на основі низки інструментальних засобів, зокрема [3], поділяється на компоненту навчання розпізнаванню та власне розпізнавач. У процесі навчання розпізнаванню оцінюються параметри моделей на всьому рівні ієрархії мовленнєвих образів з використанням відповідного корпусу на основі мовлення та/або тексту. Розпізнавач, отримуючи мовленнєвий сигнал, спочатку здійснює перехід від амплітудно-часового подання сигналу в простір первинних ознак. Таким чином отримані акустичні вектори аналізуються на предмет наявності в них мовлення, та обчислюються додаткові ознаки, що моделюють індивідуальні особливості мовлення дикторів. Потім обчислюються та запам'ятовуються значення критерію для всіх перспективних гіпотез еталонного сигналу з суттєвим обмеженням лексичного контексту. І вже на основі значень параметрів перспективних гіпотез послідовностей слів формується остаточна відповідь розпізнавання з урахуванням впливу більш широкого лексичного контексту. Отже, розпізнавач для кожного мовленнєвого сегменту надає одну або більше послідовностей слів із оцінками початку слова, його тривалості та мірою довіри.

База даних і знань системи перетворення мовлення на текст формується з моделей, що використовуються в різних блоках розпізнавача. Параметри моделей оцінюються за навчальною вибіркою, що формується з корпусів як мовленнєвих, так і текстових, які відповідають предметній області. При цьому враховуються мови, тематика, лексика, прогнозований розподіл мовців і стилів мовлення, акустичне середовище, властивості каналів, через які передається звукова інформація тощо. Так для оцінки параметрів акустичних моделей фонем використано близько 500 годин анотованих записів телерадіоєфіру для української та російської мов. Цей же матеріал — для визначення голосової активності та моделювання індивідуальних особливостей мовця. Для побудови моделі врахування обмежень на слідування слів — лінгвістичної моделі — сформовано текстовий корпус з використанням понад 2 ГБ текстів, отриманих із новинних сайтів, зокрема [5].

Подальший аналіз оброблення розпізнаного тексту здійснюється як людиною, так і автоматичними засобами. Наприклад, експерт виправляє помилки в отриманому тексті для формування звіту. При автоматичному обробленні текстів виникає необхідність роботи з числами тощо. Тому для системи розроблено блок пост-процесингу відповіді розпізнавання, метою якого є полегшити сприйняття людиною отриманого тексту та підготувати результат розпізнавання для подальшого автоматичного оброблення:

регістром літер позначаються мови; кольором виділяються слова, щодо яких система “сумнівається”; здійснюється прогноз щодо знаків пунктуації на основі [4]; позначаються моменти вірогідного переходу черги говорити до іншої особи [6]; відновлюються числові вирази за послідовностями слів.

Запропонована схема перетворення мовлення на текст дала змогу отримати результат розпізнавання аудіоданих з YouTube-каналів у зручному вигляді як для користувача-людини, так і для подальшого автоматичного оброблення з метою оперативного моніторингу. Розроблено моделі для української, російської мов та їх одночасного використання, а також для англійської та німецької мов.

У ближчих планах: здійснювати пошук сегментів мовлення конкретної особи; тональний аналіз; розширення кількості мов; адаптація системи для диктування тексту, автоматичного формування субтитрів і реалізації усного діалогу між людиною та технічними системами.

### Література

1. Вінцюк Т. К., Сажок М. М., Селюх Р. А., Федорин Д. Я., Юхименко О. А., Робейко В. В. Автоматичне розпізнавання, розуміння та синтез мовленнєвих сигналів в Україні. *Керуючі системи та комп'ютери*. 2018. № 6. С. 7–24.
2. Сажок Н. Н., Робейко В. В., Федорин Д. Я., Селюх Р. А. Система преобразования телерадиовещания в текст для украинского языка. *Управляющие системы и машины*. 2015. № 6. С. 66–73.
3. Povey D., Ghoshal A., Boulianne G. et. al. The Kaldi Speech Recognition Toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.
4. Tilk O., Alumäe T. Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration. *Interspeech*. 2016. S. 3047-3051.
5. <http://lang.org.ua/uk/corpora/>
6. Zewoudie A. W., Luque J., Hernando J. The use of long-term features for GMM- and i-vector-based speaker diarization systems. *EURASIP Journal on Audio, Speech, and Music Processing (2018)* 2018. 14.