

УДК 621.39

О. М. Ткаченко, О. Ф. Грійо Тукало

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПОШУКУ АУДІОЗАПИСІВ НА ОСНОВІ АУДІОКОНТЕНТУ

Вінницький національний технічний університет, м. Вінниця

Анотація. Розроблено інформаційну технологію пошуку аудіозапису за аудіофрагментом в архіві аудіозаписів, в основу якої покладено пошук на основі аудіоконтенту. Запропоновано модель корпусу аудіозаписів, яка містить масив аудіофайлів, базу параметрів і метаданих та kd-дерево, що дозволило зменшити тривалість обраного для пошуку аудіофрагмента. Запропоновано нові оцінки міри близькості аудіозаписів, які враховують специфіку застосування kd-дерев для пошуку найближчого вектора. Запропоновано методи пошуку аудіозаписів, реалізація яких дозволила досягти суттєвого зменшення часу пошуку із збереженням його повноти та точності.

Ключові слова: автоматичний пошук аудіозапису за аудіофрагментом, наближений пошук найближчого вектора на основі kd-дерева, архів аудіозаписів, кластерний аналіз, мел-частотні кепстральні коефіцієнти.

Аннотация. Разработана информационная технология поиска аудиозаписи по аудиофрагменту в архиве аудиозаписей, в основу которой положен поиск на основе аудиоконтента. Предложена модель корпуса аудиозаписей, которая содержит массив аудиофайлов, базу параметров, метаданных и kd-дерево, что позволило уменьшить продолжительность выбранного для поиска аудиофрагмента. Предложены новые оценки степени близости аудиозаписей, которые учитывают специфику применения kd-деревьев для поиска ближайшего вектора. Предложены методы поиска аудиозаписей, реализация которых позволила достичь существенного уменьшения времени поиска с сохранением его полноты и точности.

Ключевые слова: автоматический поиск аудиозаписи по аудиофрагменту, приближенный поиск ближайшего вектора на основе kd-дерева, архив аудиозаписей, кластерный анализ, мел-частотные кепстральные коэффициенты.

Abstract. The information technology for search of the audio recording by an audio fragment in the audio archive is developed, which is based on the audio content search. The model of audio corpus, containing an array of audiofiles, a database of parameters and metadata, dynamic kd-tree, is proposed, which has reduced the duration of the selected audio fragment to search. There are suggested the new estimations of the measure proximity that take into account the specifics of the kd-tree use to find the nearest vector. Proposed search methods of audio allowed to achieve a significant reduction in the search time preserving its recall and precision.

Keywords: an automatic audio search by audio fragment, an approximate kd-tree search of the nearest vector, archive of audio recordings, cluster analysis, mel-frequency cepstral coefficients.

Вступ

Розвиток Інтернету, цифрових ЗМІ, мультимедійних технологій та нових способів обробки звукової інформації викликав великий інтерес і увагу до шляхів, якими інформаційні технології можуть бути застосовані до такого контенту. З можливістю доступу до великих архівів мультимедіа, практично в будь-якому місці і в будь-який час, необхідно було запропонувати нові способи навігації та взаємодії з цими архівами аудіозаписів [1].

Актуальність

Актуальність задачі автоматичного пошуку аудіозаписів зумовлено високою вартістю закордонного програмного забезпечення (ПЗ), закритістю коду або неуніверсальністю існуючого ПЗ.

Мета та задачі досліджень

Метою роботи є підвищення повноти та точності результатів і швидкості пошуку аудіозапису за заданим аудіофрагментом в архіві аудіозаписів за рахунок розроблення нової інформаційної технології.

У роботі розглядається один з можливих підходів до створення інформаційної технології пошуку (ІТП) аудіозапису за його фрагментом [2-7]. В основу запропонованої ІТП в архівах аудіозаписів покладено пошук на основі аудіоконтенту, тобто оцінювання подібності аудіозаписів виконується виключно на основі інформації, яку можна отримати з аудіофайлу.

Основними етапами (відповідно задачами) в процесі виконання автоматичного пошуку аудіозаписів є:

1. обчислення параметрів для опису аудіосигналу;
2. зменшення розмірності параметрів;
3. впорядкування параметрів;
4. безпосередньо виконання пошуку за обраною мірою оцінювання подібності аудіозаписів.

Розв'язання задачі параметризації аудіозапису

Параметрами для опису аудіозаписів було обрано мел-частотні кепстральні коефіцієнти (MFCC – Mel Frequency Cepstral Coefficients):

$$MFCC_i = \sum_{j=0}^{N_{FB}-1} \ln(s_j) \cos\left(\frac{\pi i(j+0,5)}{N_{FB}}\right), \quad i = 1, \dots, N_{MFCC}.$$

де N_{FB} – кількість трикутних фільтрів, що перекриваються і рівномірно розміщуються на шкалі мел-частот, s_j – спектральна енергія, N_{MFCC} – кількість коефіцієнтів (розмірність) вектора параметрів MFCC.

Основні характеристики, що застосовуються для обчислення MFCC-параметрів такі: частота дискретизації аудіофайлу – 44,1 кГц, довжина фрейму, для якого обчислюється вектор MFCC, – 20 мс, розмір перекриття фреймів – 50% (10 мс), кількість трикутних фільтрів – 40, кількість коефіцієнтів (розмірність вектора) MFCC – 13.

Узагальнена схема пошуку аудіозапису за аудіофрагментом

Пошук невідомого аудіозапису виконується шляхом порівняння його параметричного подання з базою параметрів аудіозаписів, що містяться в корпусі, критерієм порівняння є відстань. Загальну схему пошуку аудіозапису наведено на рисунку 1.

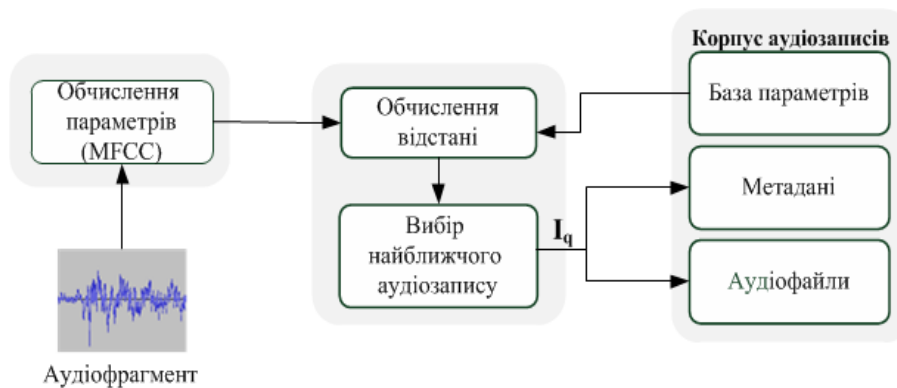


Рисунок 1 – Загальна схема пошуку аудіозапису

Подання аудіозапису за допомогою параметрів MFCC (векторів) дозволило компактно описати аудіозапис і тим самим прискорити процес пошуку в десятки разів. Подальше зменшення обсягів пам'яті, необхідної для зберігання бази параметрів, було досягнуто завдяки застосуванню методів кластерного аналізу для формування корпусу аудіозаписів. Кластеризацію було реалізовано за допомогою вдосконаленого методу k -середніх, детально описаного в [2].

Впорядкування бази параметрів на основі kd-дерев

У великих архівах аудіозаписів виконання пошуку шляхом повного перебору, тобто порівняння обраного аудіофрагмента з усіма центроїдами аудіозаписів стає неможливим, оскільки час пошуку (кількість обчислень відстаней) пропорційний $O(dN)$, де N – кількість центроїдів-векторів в корпусі, d – розмірність векторів. Тому пошук в корпусі аудіозаписів запропоновано виконувати з застосуванням kd-дерев, тобто впорядкувати центроїди (d -вимірні вектори) бази параметрів корпусу у вигляді ієрархічної структури – k -вимірного дерева пошуку [4, 7].

Оцінювання міри близькості аудіозаписів

Запропоновано декілька варіантів оцінювання міри близькості невідомого аудіофрагмента з аудіозаписами, які базуються на обчисленні відстані від фрагмента до заданої кількості k найближчих центроїдів та кількості потраплянь в список найближчих [5, 6].

- Оцінка міри близькості за приведеною відстанню (Distance):

$$WH(\mathbf{X}, \tilde{\mathbf{Y}}) = \min_q \sum_{l=1}^n \frac{\text{Weighted Hits}_q}{n}, \text{Distance}_q = \begin{cases} \min(D_p^2), \forall q \in \mathbf{I} \\ D_p^2, p = k, \forall q \notin \mathbf{I} \end{cases}$$

- Оцінка міри близькості за кількістю потраплянь h в список k найближчих центроїдів (Hits):

$$H(\mathbf{X}, \tilde{\mathbf{Y}}) = \max_q \sum_{l=1}^n \frac{\text{Hits}_q}{n}, \text{Hits}_q = \sum_{p=1}^k h_p, h_p = \begin{cases} 1, \forall q \in \mathbf{I} \\ 0, \forall q \notin \mathbf{I} \end{cases}$$

- Оцінка міри близькості за зваженою кількістю потраплянь в список k найближчих центроїдів (Weighted Hits):

$$WH(\mathbf{X}, \tilde{\mathbf{Y}}) = \max_q \sum_{l=1}^n \frac{\text{Weighted Hits}_q}{n}, \text{Weighted Hits}_q = \sum_{p=1}^k (D_k^2 - D_p^2) \cdot h_p.$$

На рисунку 2 показано результати порівняння заданого для пошуку аудіофрагмента з аудіозаписами в корпусі (власний аудіозапис під № 3) за запропонованими оцінками міри близькості, та зведена гістограма, на якій наведено всі варіанти оцінювання міри близькості у відносних одиницях для досягнення єдиного масштабу.

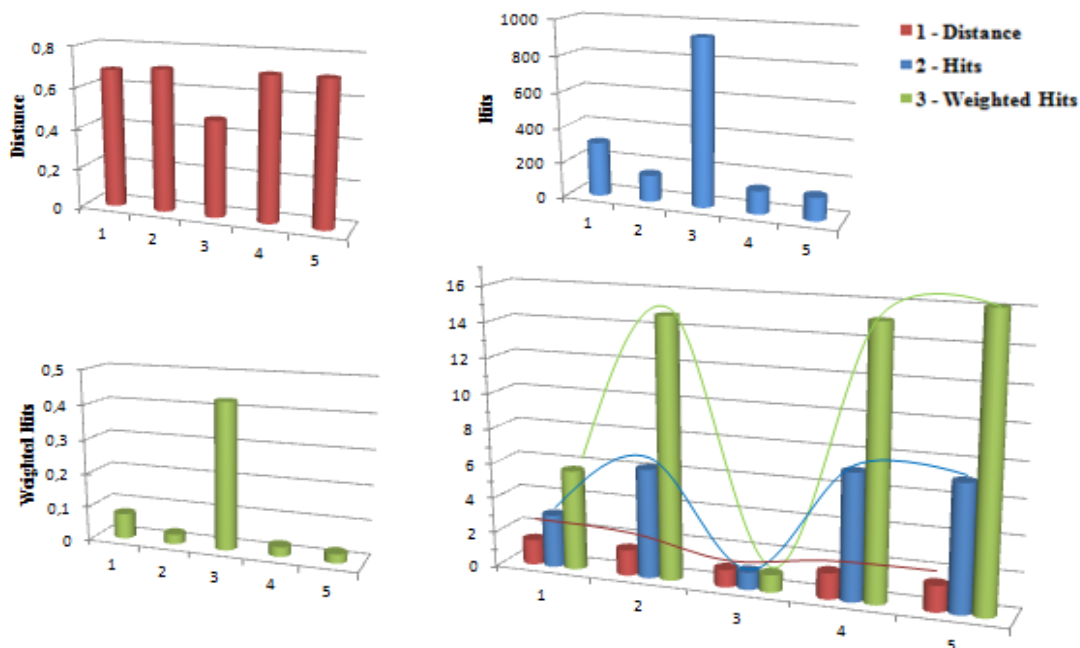


Рисунок 2 – Порівняння оцінок міри близькості невідомого аудіофрагмента з аудіозаписами корпусу

Відповідно до отриманих результатів найвищий рівень розрізнення власного та чужих аудіозаписів в корпусі забезпечує оцінювання міри близькості за зваженою кількістю потраплянь аудіозапису у список k найближчих (Weighted Hits), для якої оцінка відстані до власного аудіозапису при $k=20$ в середньому в 8 разів перевищує аналогічну оцінку відстані до інших аудіозаписів.

Підсумки аналізу основних чинників впливу на продуктивність пошуку

Аналіз основних чинників впливу на продуктивність пошуку показав, що на час пошуку суттєво впливають:

- розмірність векторів параметрів d ;
- коефіцієнт розповсюдження оберненої фази пошуку r .

Визначено, що компроміс між обчислювальною складністю методу та повнотою і точністю пошуку досягається при $d=6$ та $r=1/5$: сукупне зниження розмірності простору параметрів та виконання наближеного пошуку ($d=6, r=1/5$) дозволяє досягти зменшення складності обчислень в 388 разів при повній пошуку 90%.

Висновки

1. У роботі розглянуто один з можливих підходів до створення інформаційної технології пошуку (ІТП) аудіозапису за його фрагментом. В основу запропонованої ІТП в архівах аудіозаписів покладено пошук на основі аудіоконтенту, тобто оцінювання подібності аудіозаписів виконується виключно на основі інформації, яку можна отримати з аудіофайлу.

2. Запропоновано модель корпусу аудіозаписів, яка містить масив аудіофайлів, базу кластерів параметрів та метаданих і kd-дерево, що дозволило зменшити тривалість обраного для пошуку аудіофрагмента та реалізувати інформаційну технологію пошуку аудіозаписів з різною тривалістю звучання.

3. Запропоновано методи пошуку аудіозаписів, які дозволили досягти суттєвого зменшення часу пошуку із збереженням його повноти та точності.

Список літератури

1. Muller M. Signal processing for music analysis / M. Muller, D. Ellis, A. Klapuri, and G. Richard // IEEE Journal of Selected Topics in Signal Processing, V. 5, No. 6, oct. – 2011 – P. 1088 – 1110. – ISSN 1932-4553

2. Ткаченко О. М. Метод кластеризації на основі послідовного запуску k-середніх з удосконаленим вибором кандидата на нову позицію вставки [Електронний ресурс] / О. М.Ткаченко, О. Ф. Грійо Тукало, О. В. Дзісь, С. М. Лаховець // Наукові праці ВНТУ. – 2012. – № 2, В. 2. – . Режим доступу до журн.: <http://praci.vntu.edu.ua/index.php/praci/article/view/323/321>. – ISSN 2307-5376.

3. Ткаченко О. М. Підхід до оцінювання тривалості фрагмента для пошуку музичного твору за заданим шаблоном / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 1.– С. 31–40. – ISSN 1999-9941.

4. Ткаченко О. М. Метод підвищення швидкості пошуку фрагменту аудіозапису із застосуванням kd-дерев / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 3.– С. 57–66. – ISSN 1999-9941.

5. Ткаченко О. М. Ідентифікація музичного твору за фрагментом / О. М. Ткаченко, О. Ф. Грійо Тукало, Ю. Л. Далекий // Інформаційні технології та комп'ютерна інженерія: IV Міжнар. наук.-практ. конф., 28–30 травня 2014 р. : тези доповідей. – Вінниця, 2014. – С. 256–257. – 2 с. – ISBN 978-966-641-465-9.

6. Ткаченко О. М. Ідентифікація фрагмента музичного твору на основі приведеної власної відстані / О. М. Ткаченко, О. Ф. Грійо Тукало // Оброблення сигналів і зображень та розпізнавання образів: XII Всеукр. міжнар. конф., 3–7 листопада 2014 р. – Київ, 2014. – С. 23–26. – 4 с. – ISBN 978-966-479-069-4.

7. Грійо Тукало О. Ф. Інформаційна технологія пошуку аудіозаписів на основі kd-дерев [Текст] / О. Ф. Грійо Тукало, О. М. Ткаченко // Методи та засоби кодування, захисту й ущільнення інформації : V Міжнар. наук.-практ. конф, 19-21 квітня 2016 р. : тези доповідей – Вінниця : ВНТУ, 2016. – С.120-122.

Відомості про авторів

Ткаченко Олександр Миколайович – к.т.н., доцент кафедри обчислювальної техніки, Вінницький національний технічний університет, Хмельницьке шосе, 95, м. Вінниця.

Грійо Тукало Оксана Франсисківна – к.т.н., старший викладач кафедри обчислювальної техніки, Вінницький національний технічний університет, Хмельницьке шосе, 95.