

Інформаційна технологія пошуку аудіозаписів на основі kd-дерев

Грійо Тукало О. Ф.¹, Ткаченко О. М.²

¹Асист. кафедри обчислювальної техніки, Вінницький національний технічний університет
Хмельницьке шосе 95, м. Вінниця, Україна, xxmargox@gmail.com

²Доц., к.т.н., доцент кафедри обчислювальної техніки, Вінницький національний технічний університет
Хмельницьке шосе 95, м. Вінниця, Україна, alextk1960@gmail.com

Анотація — Розроблено інформаційну технологію пошуку аудіозапису за аудіофрагментом. Запропоновано модель корпусу аудіозаписів, яка містить масив файлів, базу параметрів та метаданих і динамічне kd-дерево, що дозволило зменшити тривалість обраного для пошуку аудіофрагмента. Запропоновано нові оцінки міри близькості, які враховують специфіку застосування kd-дерев для пошуку найближчого вектора та метод комбінованого пошуку. Реалізація даного методу дозволила досягти суттєвого зменшення часу пошуку із збереженням його повноти та точності.

Ключові слова: пошук аудіозапису за аудіофрагментом, наближений пошук на основі kd-дерева, кластерний аналіз, параметризація.

Information technology for search of the audio recordings based on a kd-tree

Griyo Tukalo O. F.¹, Tkachenko O.M.²

¹Asist., Department of Computer Facilities, Vinnytsia National Technical University
95 Khmelnytske shose, Vinnytsia, Ukraine, xxmargox@gmail.com

²Assoc., Ph.D., Department of Computer Facilities, Vinnytsia National Technical University
95 Khmelnytske shose, Vinnytsia, Ukraine, alextk1960@gmail.com

Abstract — The information technology for search of the audio recording by an audio fragment is developed. The model of audio corpus, containing an array of audiofiles, a database of parameters and metadata, dynamic kd-tree, is proposed, it has reduced the audio fragment duration, selected for the 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 and the method of the combined search. The implementation of this method allows to achieve a significant reduction in the search time preserving its recall and precision.

Keywords: automatic search of an audio recording by the specified audio fragment, approximate search using kd-tree, cluster analysis, parameterization.

I. ВСТУП

Суть пошуку на основі аудіоконтенту полягає в тому, щоб автоматично визначати аудіозаписи, що містять частини, подібні до заданого аудіофрагменту. Більшість існуючих систем поєднує в собі такі етапи обробки аудіозаписів: аналіз вхідного аудіосигналу, виділення певних параметрів (ознак), вибір міри їх порівняння та визначення аудіозапису або деякої категорії аудіозаписів, найближчих до аудіофрагменту за обраною мірою.

Незважаючи на великі успіхи, сучасним інформаційно-пошуковим системам притаманні певні недоліки: висока вартість закордонного програмного забезпечення (ПЗ), закритість коду та неуніверсальність існуючого ПЗ, зокрема неможливість їхнього безпосереднього застосування для розв'язання таких задач, як моніторинг трансляції на радіостанціях

(телебаченні) для визначення кількості повторень в ефірі рекламного продукту, музичного твору тощо.

Все це зумовлює актуальність задачі автоматичного пошуку аудіозаписів [1]. У даній роботі розглядається один з можливих підходів до створення інформаційної технології пошуку (ІТП) аудіозапису за його фрагментом [2-7]. У зв'язку з великими розмірами мультимедійних архівів (музики, фільмів, теле- та радіопередач, різного роду стенограм, записів аудіореєстраторів тощо) для побудови реальної масштабованої системи пошуку аудіозаписів необхідним є застосування кластерного аналізу для побудови швидкого пошукового індексу в архіві та методів швидкого пошуку для забезпечення низької складності обчислень в процесі пошуку.

II. ОПИС ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ

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.

Під час пошуку виконується порівняння заданого аудіофрагмента та аудіозаписів в корпусі шляхом обчислення відстані між їх параметричними поданнями. Після визначення найближчого аудіозапису на основі його індексу I_q можна отримати метадані та відтворити аудіозапис в повному обсязі. Загальну схему пошуку аудіозапису наведено на рисунку 1. Далі буде описано реалізацію кожного з цих блоків.

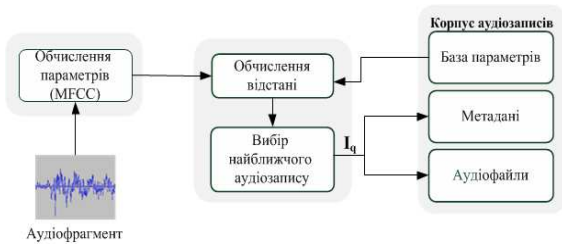


Рисунок 1 – Загальна схема пошуку аудіозапису

Як зазначено вище для виконання пошуку кожен аудіозапис було подано у вигляді послідовності параметрів. Як параметри було обрано мел-частотні кепстральні коефіцієнти (MFCC), що являють собою вектори параметрів розмірністю $d=13$. Параметризація дозволила компактно описати аудіозапис (зменшивши кількість інформації в десятки разів) і, тим самим, прискорити процес пошуку.

Попередньо було сформовано корпус, що являє собою сукупність аудіозаписів, їх метаданих та базу параметрів. Під час формування корпусу аудіозаписів для зменшення обсягів пам'яті, необхідної для зберігання бази параметрів, було застосовано кластерний аналіз, а саме вдосконалений метод кластеризації k-середніх [2]. В результаті аудіозаписи в корпусі було подано центроїдами кластерів, обчислених для параметрів MFCC.

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. З огляду на це та на основі проведеного аналізу методів швидкого пошуку запропоновано виконувати пошук аудіофрагмента в корпусі аудіозаписів з застосуванням kd-дерев, тобто впорядкувати центроїди (d -вимірні вектори) бази параметрів корпусу у вигляді ієрархічної структури – k -вимірного дерева пошуку [4, 6].

Для підвищення швидкості пошуку запропоновано об'єднати центроїди усіх аудіозаписів корпусу в межах одного файлу, щоб впорядкувати їх на основі kd-дерева, що зображено на рисунку 2, оскільки, коли кожен аудіозапис в

корпусі $\tilde{Y}_q, q=1, Q$ є окремим файлом, зі зростанням Q кількість операцій зростатиме лінійно $N_{op} = O(c \cdot Q \cdot \log m)$, $c = const$, що є неприйнятним для великих архівів аудіозаписів, тому що час пошуку стане занадто великим.

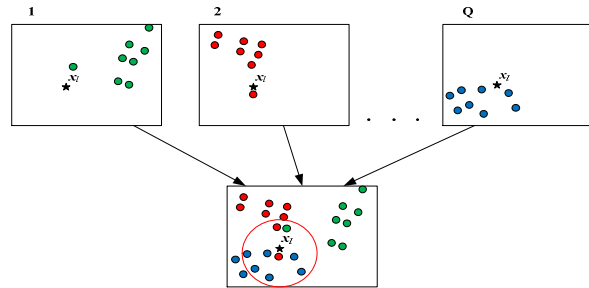


Рисунок 2 – Об'єднання файлів з центроїдами MFCC-параметрів

Проведені дослідження показали, що в результаті виконання пошуку в базі параметрів, впорядкованій на основі kd-дерева, в списку знайдених для невідомого аудіофрагмента X найближчих центроїдів відсоток центроїдів власного аудіозапису є більшим відносно частки центроїдів решти аудіозаписів корпусу, однак центроїд власного аудіозапису в корпусі не завжди знаходиться на першій позиції в списку найближчих. У зв'язку з цим зроблено висновок, що доцільно виконувати пошук не одного найближчого центроїда, оскільки це призводитиме до помилкових результатів, а деякої множини $C, |C|=k$ центроїдів.

Запропоновано та досліджено декілька варіантів оцінювання міри близькості невідомого аудіофрагмента $X = \{x\}, |X|=n, x = (x_1, x_2, \dots, x_d)$ з аудіозаписами $\tilde{Y}_q = \{\tilde{y}_q\}, |\tilde{Y}_q|=m, \tilde{y}_q = (\tilde{y}_{q1}, \tilde{y}_{q2}, \dots, \tilde{y}_{qd})$, які базуються на обчисленні відстані від фрагмента до заданої кількості k найближчих центроїдів та кількості потраплянь в список найближчих [5, 8].

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

$$WH(X, \tilde{Y}) = \max_q \sum_{l=1}^n \frac{\text{Weighted Hits}_{lq}}{n}, \text{Weighted Hits}_{lq} = \sum_{p=1}^k (D_{lk}^2 - D_{lp}^2) \cdot h_p.$$

Визначено головні чинники впливу на час пошуку:

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

На основі аналізу основних чинників впливу на продуктивність пошуку запропоновано метод комбінованого пошуку аудіозапису за заданим

аудиофрагментом, що дозволив значно скоротити час пошуку при збереженні повноти та точності результатів.

Суть методу формулюється таким чином: спочатку виконується швидкий пошук (з коефіцієнтом розповсюдження оберненої фази пошуку $0 < r < 1$) в корпусі зниженої розмірності, упорядкованому на основі kd-дерева, після чого серед кількох найбільш подібних до заданого аудиофрагмента аудіозаписів здійснюється повний пошук, під час якого виконується порівняння параметрів MFCC (а не центроїдів) кожного з цих аудіозаписів з параметрами невідомого аудиофрагмента, причому для виконання результативного пошуку достатньо лише декількох фреймів. Складність обчислень першого та другого етапів комбінованого пошуку — це сумірні значення. Визначено, що компроміс між обчислювальною складністю та повнотою і точністю пошуку досягається при $d = 6$ та $r = \frac{1}{5}$.

Для проведення експериментальних досліджень було розроблено програмне забезпечення для реалізації інформаційної технології. На рисунку 3 показано компоненти розробленої інформаційної технології.

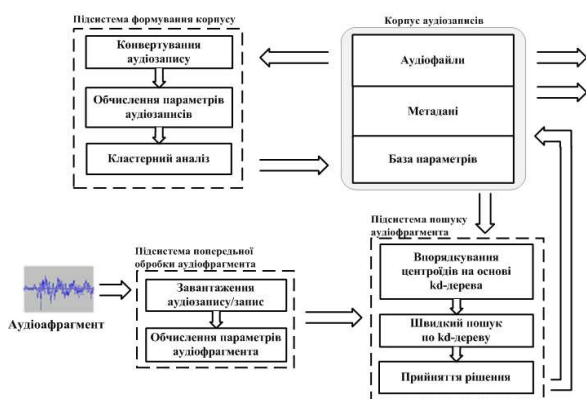


Рисунок 3 – Компоненти інформаційної технології пошуку аудіозаписів

Тестова версія корпусу містила 1000 аудіозаписів, однак методи, що реалізовані в програмному продукті, забезпечують можливості розширення архіву аудіозаписів. Для отримання параметрів MFCC аудіозаписи було перетворено в формат wav (mono). Кожен аудіозапис в корпусі було подано $m=1000$ центроїдів (в середньому на кластер припадає близько 20 векторів). 3 аудіозаписів корпусу випадковим чином було обрано 100 аудиофрагментів (тривалістю 1, 5 та 15 с) для пошуку. Результати досліджень показали, що застосування методу комбінованого пошуку порівняно з точним пошуком на основі kd-дерева дозволяє знизити складність обчислень приблизно в 28 разів та правильно визначити 99% аудіозаписів корпусу. Середній час пошуку (Windows7, Intel Core i5-4200M CPU @ 2.5GHz, 6GB RAM) становить менше 1 с.

Висновки

Запропоновано модель корпусу аудіозаписів, яка містить масив файлів, базу параметрів та метаданих і динамічне kd-дерево, що дозволило зменшити тривалість обраного для пошуку аудиофрагмента та реалізувати інформаційну технологію пошуку аудіозаписів з різною тривалістю звучання. Мінімальна тривалість аудиофрагмента, яка забезпечує повноту результатів пошуку аудіозапису 99,4%, має складати 1 с. Розроблено інформаційну технологію пошуку в архіві аудіозаписів, яка містить алгоритми та програми для реалізації запропонованих методів.

- [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. – 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. Дзись, 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-дерева / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 3. – С. 57 – 66. – ISSN 1999-9941.
- [5] Ткаченко А.Н., Грійо Тукало О.Ф. Метод быстрого поиска фрагмента аудиозаписи с усовершенствованной оценкой меры близости / А. Н. Ткаченко, О. Ф. Грійо Тукало // Международный научный журнал «Управляющие системы и машины». – 2015. – №6. – С.51 – 58. – Киев. – ISSN 0130-5395.
- [6] Ткаченко О. М. Пошук найближчого вектора у кодових книгах на основі бінарного дерева / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія: I Міжнар. наук.-практ. конф., 19–21 травня 2010 р. : тези доповідей. – Вінниця, 2010. – С. 160–161. – 2 с. – ISBN 978-966-641-356-0.
- [7] 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.
- [8] 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.