

## Дипломна работа

Задание: Приложение за offline разпознаване на дублиращи се музикални файлове.

Мотивация: В ежедневието ни, доста често се озоваваме с дублиращи се музикални файлове от най-екзотичен формат, тип, bitrate и честотни ленти. Много често се налага да обновяваме остарели музикални файлове, които или са в лошо качество или са в някаква степен повредени. На пазара съществуват доста програми, които изпозват различна метаинформация, която се съхранява по файловете, но до колко надеждна може да е тя?

Наблюдава се, че доста хора прехвърлят аудио колекциите си от музикални CD-та на персоналните си компютри в дигитални формати като mp3, flac и т.н. Един от основните проблеми в този случай е, че върху музикалните CD-та, тази метаинформация не е налична. Генерираните файлове са с имена от типа Track\_1 или Unknown Author – Unknown Title 1, което не говори нищо за песента. Доста често при rip-ването (процеса на извличане на музиката от CD-тата в дигитален формат) , получаваме папки с дати, файлове с default имена и т.н.

Комбинирайки горните два случая, стигаме до извода, че не можем да разчитаме нито на метаинформацията (ако изобщо е налична), нито на заглавията на файловете за определяне на дублирани такива. Съществуват online услуги, предлагащи audio fingerprinting, чрез който се разпознава дадена песен (с някакъв процент на сигурност), като се прослуша част от нея и се изпрати за анализ към услугата, която от своя страна я разбива до съставни компоненти, генерира fingerprint и проверява в базата си данни. Ако има съвпадение, то бива предложено на потребителя. Ако не, записа се може да се идентифицира.

Плюсовете на тези системи са видими – могат единствено по audio fingerprinting да идентифицират името на песента, изпълнител и всякаква друга свързана информация.

Минусите при тях са нуждата от Internet свързаност, изложеност на информация в мрежата и възможност песента да не е регистрирана в базата данни. Последното е винаги като правило, когато говорим за авторски/собствени музикални изпълнения, които не са били достъпни от услугата и базата ѝ.

Идеята за приложение, което да намира и разпознава дублиращи се файлове е то да не се обвързва нито с дадена услуга, нито да изисква свързаност към Интернет. Основната му цел няма да бъде идентифициране на песента, а сравняване с другите песни от дадена колекция/колекции, с цел да групира дублиращите се и да определи коя е най-добре да бъде запазена с оглед на качество/размер/скорост на възпроизвеждане/файлов формат и т.н. Допълнително към това това приложение, може да се интегрира възможността да се прави и audio recognition към вече готови системи.

Идея: Приложението ще представлява Java базирана (не е задължително), десктоп програма, която ще прави следните стъпки:

- По зададена папка ще извлича списък с музикални файлове.
- Създава се локална база с аудио профила на музикалния файл, който ще се използва за намиране на дублиращи се файлове.
- Аудио профила ще бъде нормализиран, ще се използва и спектрален анализ на музикалния сигнал с цел отделяне на компонентите му за по – добро сравнение.
- За всеки файл, освен данните за път, име, аудио профил и спектър, ще се запазват и данни за големина, bitrate, формат и метаданни (ако са налични) – автор, заглавие и т.н..
- При сравняване на музикалните файлове ще се използват имената, аудио профила, спектъра и части от метаданните. Потребителя ще има възможност да задава приоритети на сравнение.
- Изхода ще бъде интерактивен списък с възможни действия върху групата файлове, като задаване на имена, промяна на метаданните, изтриване, запазване, преместване и др.

Възможностите на приложението мога да се разширят доста, с оглед на проведения аудио анализ. Съществува възможност да се добавят и разработят и допълнителни логически модули, които да определят музикален стил и жанр, чрез което да се постигне допълнително структуриране на дадена музикална колекция. С оглед на това, може да се предложи и възможност за препоръчване на и дори предложение за жанрове или стил за файлове, при които тази информация липсва.