MI-VMW - semestrální projekt

téma: Klasifikace hudebního žánru

autoři: Jan Dufek, Jan Bouchner

Fakulta informačních technologií ČVUT

ZS 2013/2014

Popis řešeného projektu

Cílem projektu bylo vytvořit aplikaci, která by byla schopna klasifikovat vstupní nahrávku, tedy audio soubor, do jednoho z předem definovaných hudebních žánrů.

Vstup
Libovolná hudební nahrávka ve formátu *mp3* nebo *wav*.

Dotaz
Při dotazu jsou identifikovány nejpodobnější databázové skladby a na jejich základě je skladba na dotazu zařazena do daného žánru.

Výstup
Koláčový graf rozpoznaných hudebních žánrů a jejich procentuální odhad příslušnosti skladby do těchto žánru.

Aplikace

Stavba
Aplikace obsahuje následující části:
- Extrakce deskriptorů z audia
- Podobnostní míra pro porovnání dvojice audio souborů, tj. jejich deskriptorů
- Identifikace shody nahrávané skladby se žánrem
- Databázová komunikace
- Webové rozhraní

Použité technologie

Použité programovací jazyky:
- Java SE, Java EE (Spring MVC)
- JSP pro webové rozhraní

Databázové úložiště informací:
- Pro uložení informací o skladbách jsme zvolili databázi MySQL. Když jsme na začátku pracovali se standardem *MPEG7*, vzhledem k formátu popisu standardu, který je v XML, jsme přemýšleli o použití XML databáze, ale nakonec jsme kvůli změně okolností (viz dále) použili databázi relační.

Extrakce deskriptorů z audio souboru:
- K extrakci deskriptorů z audio souboru jsme se rozhodli použít aplikaci *jAudio* poté, co jsme první polovinu semestru experimentovali s knihovnou *MPEG-7 audio encoder*.
- O tomto experimentování viz dále v kapitole *Rešerše a implementace*.

Definice žánrů

Definice žánrů probíhala tak, že bylo zvoleno několik desítek skladeb daných žánrů, čímž byla vytvořena referenční množina skladeb v databázi. Klastrování skladeb do žánrů probíhalo manuálně, kdy byla při nahrávání extrahovaných informacích o skladbách do databáze (získané deskriptory) zasílána zároveň i informace o jaký hudební žánr se jedná. Jedna část databáze skladeb jsou tedy dechovky, jiná část databáze zase jazz atd.

Můžeme tedy mluvit o jakémsi supervised learning (pro vstupní data je určen správný výstup).

Informace byly extrahovány z celých skladeb, nikoliv jen z jejich vybraných částí stopy.

Zvolené žánry

Pro klasifikaci jsme si zvolili následující hudební žánry:

- dechovka
- ska
- jazz
- death metal
- hip-hop

Získaná data pro databázi

Kvůli potřebné různorodosti dat jsme se snažili mít od každého hudebního žánru alespoň 5 různých interpretů a jejich skladeb. Ze skladeb byly extrahovány jejich vlastnosti a ukládány do databáze.

Na konci extrakce trénovacích vlastností do databáze jsme měli k dispozici přes 800 skladeb, přes 108 000 vlastností (přičemž je ale 1 vlastnost v tomto případě uložena jako 1 prvek vektoru – pokud je tedy získaný deskriptor reprezentován vektorem o 10 prvcích, zabírá v databázi 10 řádek) a ukládali jsme informace o celkem 18 deskriptorech (z nichž jsme nakonec použili jen pár).

Rešerše a implementace

Extrakce deskriptorů

Aby bylo možné provádět porovnávání podobnosti skladeb a na základě toho klasifikovat skladbu do příslušného žánru, bylo potřeba z audio souborů získat informace. Konkrétně v problematice rozpoznávání řeči, hudby, specifických zvuků a podobně jde o to, že se musí nějakých způsobem zpracovávat audio signály. Tyto audio signály jsou reprezentovány různými

deskriptory (což mohou být např. vektory či skaláry) popisujícími audio signál z různých pohledů. Každý jednotlivý deskriptor popisuje nějakou vlastnost signálu, např.:

- výšky
- frekvence
- tóny (a jejich ostrost)
- harmonicita
- zvukové spektrum a harmonické stupnice
- hlasitost (síla) audio signálů
- tvar křivky audio signálu
- basy a další

Deskriptory sami o sobě toho o audiu jako o celku příliš neřeknou, ale síla je v jejich kombinaci, kdy vhodným zvolením sady vybraných deskriptorů lze získat silnou míru pro podobnost. Navíc jejich získání není v dnešní době až tolik obtížné, protože existuje hned několik nástrojů schopných deskriptory extrahovat.

My se zabývali dvěma, které jsou implementovány v Javě, a které jsou zároveň dost rozdílné, neboť jeden pracuje se standardem MPEG7 a druhý s alternativními deskriptory, které standard MPEG7 nepokrývá (takty, spektrální tok, rozeznávání zvuků od ruchů apod.).

Začátky se standardem MPEG7 a nástrojem MPEG7 audio encoder

URL: http://mpeg7audioenc.sourceforge.net/
Začali jsme s nástrojem *MPEG7 audio encoder*, neboť nám byl prvně na konzultačním cvičení doporučen jako lepší na základě výsledků z minulých let.

MPEG7 je standard, který popisuje multimédia komplexně (audio, video i obraz), nástroj MPEG7 audio encoder se zabývá ale jen audio signály, což se pro naše potřeby hodilo. Poté jsme experimentovali s extrakcí deskriptorů z audio souborů ve formátu mp3 a wav.

Volba deskriptorů

Audio deskriptory jsou děleny do dvou skupin:

- 17 *low-level* deskriptorů (popisují tóny, tvary křivek, tempo atd. – nízkoúrovňové informace o signálu)
- 5 *high-level* deskriptorů (obsahují již nějakou sémantickou informaci jako melodii, zvukové efekty atd.)

Na základě informací ve slajdech z 9. přednášky MI-VMW, kde jsou jednotlivé MPEG7 deskriptory popsány, a dalších informací na internetu (viz zdroje) jsme si vytipovali několik deskriptorů, které bychom mohli použít a pokusit se jejich kombinací získat vstup pro porovnávací funkci. Vybrali jsme si tyto:

- **AudioPower**
    - měří vývoj amplitudy signálu v čase, byl by to zřejmě náš nejobecnější deskriptor pro celkový přehled o signálu
- **AudioSpectralCentroid** (případně místo něj **SpectralCentroid**)
    - reprezentuje středový bod spektra (průměrnou hodnotu), což by nám pomohlo určit například to, zda dominují vysoké nebo nízké frekvence
    - SC je podobný, ale nemá vztah k harmonické struktuře spektra a je navržen tak, aby se pomocí něho dala rozpoznávat jasnost tónů, což by se dalo využít pro hudební nástroje (techno či hip-hop by tóny pravděpodobně vůbec nemělo nebo jen pár).
- **AudioSpectrumSpread**
    - pomohl by rozlišit, zda dominují čisté tóny (např. pro případy jazzu, ska, dechovky) nebo spíše ruchy (death metal, techno)
- **AudioHarmonicity**
    - podobně jako předchozí deskriptor. Pomohl by určit mezi zvuky v harmonickém/neharmonickém spektru (hudební nástroje versus syntetika v technu)
- **TemporalCentroid**
    - průměrné hodnoty jsou obecně velice dobré. Tento deskriptor by nám udal informaci o tom, kde (v čase) je zaměřena síla signálu (pro death metal by vycházel velice široký rozptyl, např. pro ska by vycházel v určitém okruhu kolem času, kdy se začínají hrát refrény)
- **HarmonicSpectralCentroid**
    - kvůli ostrosti tónu (třeba death metal by mě velice zajímavou amplitudu)
- Z high-level deskriptorů jsme chtěl vyzkoušet z kategorie Musical Instrument Timbre **HarmonicInstrumentTimbre** a z Melody pak **MelodySequence** pro rozeznání melodie, rytmu.

Extrakce deskriptorů

Po zprovoznění MPEG7 audio encoderu a prvním pokusu o extrakci deskriptorů z testovacího hudebního souboru formátu wav jsme dostali v XML výstupu deskriptory

- AudioPowerType
- AudioSpectrumCentroidType
- AudioSpectrumEnvelopeType
- AudioSpectrumSpreadType
- AudioWaveformType

A žádné další. Zkoumali jsme, zda se nejedná o nějakou defaultní hodnotu a my nemáme pomocí parametru určit, že chceme všechny. Ale právě defaultní hodnota je výpis všech (které se v audio souboru najdou). Zkusili jsme pár dalších souborů formátu wav a dostali jsme ještě deskriptory LogAttackTime a HarmonicSpectralVariation.

Z námi vybraných 6 deskriptorů vhodných pro tvorbu míry se nám tedy podařilo vyextrahovat 3, přičemž ty nejdůležitější (centroidy) jsme až na jeden vůbec nedostali. Pokusili jsme se poté na vstup posílat soubory formátu mp3. Zde jsme byli úspěšní ještě méně, získali jsme sice deskriptorů více (i centroidy), ale hodnoty byly na spoustě místech nulové (a to u 4 z 6 deskriptorů, které jsme chtěli používat). A vzhledem k tomu, že vhodná data do databáze snadněji získáme ve formátu mp3 než ve formátu wav, toto zjištění pro nás bylo klíčové v dalším rozhodování.

Co se týče high-level deskriptorů, ty software vůbec extrahovat neumí. Po další chvilce experimentování a jsme se tedy rozhodli, že místo tohoto projektu vyzkoušíme něco jiného i za cenu toho, že se vzdáme nastudovaných informací o MPEG7 a budeme muset vyzkoušet jiné deskriptory.

Zvolení nástroje jAudio jako vhodnějšího pro naše potřeby

URL: http://jaudio.sourceforge.net/

Volba deskriptorů

Na konci 9. přednášky Doc. Skopala se píše i o alternativních audio deskriptorech, konkrétně je zvýrazněno Mel Frequency Cepstral Coefficients (MFCCs) a definováno jako “excelentní vektor vlastností vhodných i pro rozpoznávání žánru”. Nezkoumali jsme podrobně výpočet MFCC parametrů, nezkoumali jsme male scale ani cepstral analýzu, snažili jsme se jen využít nástrojem vypočtené informace, pár jich vhodně zvolit a využít je pro porovnávání.

Po zjištění, že nástroj jAudio umí z audio souboru vyextrahovat spousty spektrálních parametrů (LPCC, všechny výše uvedené centroidy a další deskriptory a právě také MFCC), rozhodli jsme se jej vyzkoušet.

Extrakce deskriptorů

jAudio není framework nebo knihovna (oproti MPEG7 audio encoder). Jedná se o již hotovou aplikaci, která se dá použít spuštěním GUI nebo z příkazové řádky. My potřebovali autorem napsané třídy využívat v rámci naší aplikace, bylo tedy nutné stáhnout si zdrojové soubory, ty upravit (např. změnit modifikátory přístupu k atributům, v jednom případě udělat ze třídy třídu abstraktní) a kvůli extrakci informací napojit kód na naši aplikaci. jAudio ukládá pouze do XML souboru, my potřebovali crate objekty (přepravky) a následné uložení do databáze.

Navolili jsme si tedy, že chceme z audio souboru získat 12 vybraných MFCC parametrů a 6 obecnějších deskriptorů.

Příklad několika vybraných MFCC parametrů:

- MFCCStandardDeviation
- MFCCSpectralCentroid
- MFCCStrongestBeat

- MFCCBeatSum
- MFCCStrengthOfStrongest Beat
- MFCCOverallAverage
- MFCCDerivativeOfSpectralCentroid
- MFCCRunningMeanOfSpectralCentroid

Příklad několika vybraných obecných deskriptorů:
- SpectralCentroidOverallAverage
- StrongestBeatDeviation

Získané informace jsme poté ukládaly do relační databáze MySQL, ve které máme 4 tabulky:

- **data** - informace o vektorech a skalárech audio souboru
- **feature** - informace o uložené vlastnosti (název) a jejím typu (zda se jedná o vektor či skalár)
- **genre** - informace o žánru (název)
- **song** - uložené písně (audio soubory)

Podobnost a dotaz na nejbližší sousedy

Jak již bylo řečeno, pro potřeby vyhledávání jsme si vytvořili databázi. Pomocí nástroje pak vyextrahovali skaláry a vektory a uložili je do databáze s tím, že jsme přiřadili žánr. Po vložení těchto dat jsme konečně mohli začít s podobnostní funkcí.

Funkci jsme tvořili pomocí analýzy. Vzali jsme data z několika písní stejného žánru, porovnali a poté stejně poměřili i se skladbami z žánrů jiných. Stejným způsobem jsme udělali tuto činnost pro všechny žánry. Následně jsme vybrali deskriptory, které se u stejného žánru lišily málo, ale u jiného byl znát rozdíl. Zároveň jsme pomocí analýzy zjistili přibližné střední hodnoty, které jsme potřebovali pro možnost sjednocení výsledků různých deskriptorů. Tyto hodnoty byly ale záměrně mírně upravovány podle míry rozlišovací schopnosti. To znamená, že deskriptorům, které vykazovaly lepší výsledky jsme střední hodnotu snížili a naopak. Tím jsme jim zvýšili jejich váhu v konečném koeficientu.

Výsledkem analýzy bylo použití těchto deskriptorů:

- BeatSumStandardDeviation
- SpectralCentroidDeviation
- MFCCDerivativeOfSpectralCentroid
- MFCCRunningMeanOfSpectralCentroid
- MFCCStandardDeviation
- MFCCStandardDeviationOfSpectralCentroid

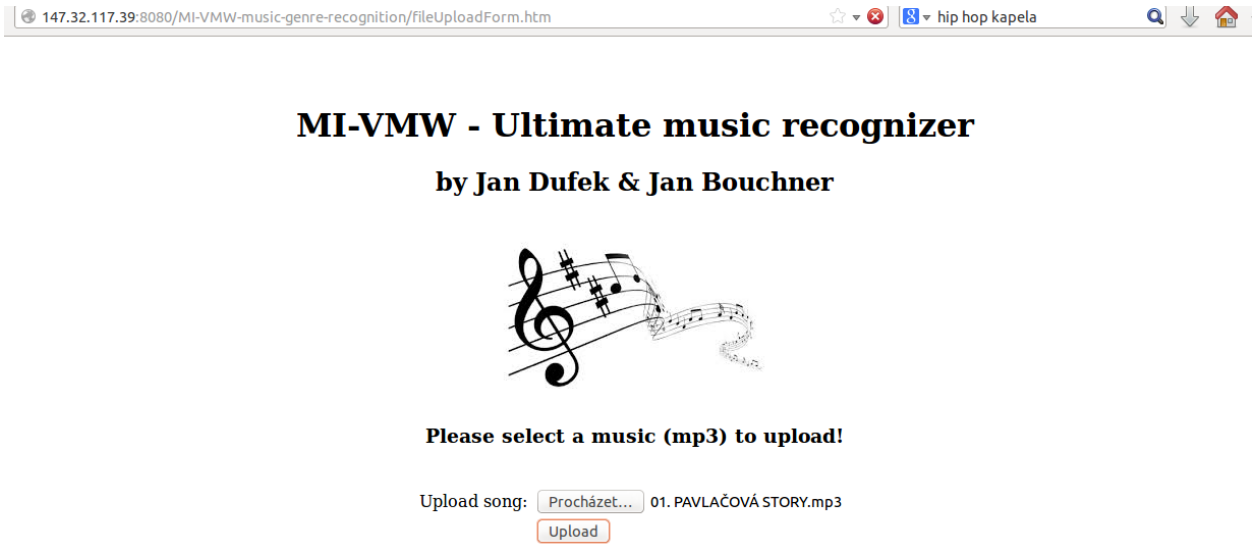
Samotná podobnostní funkce, která se používá při každém požadavku na zjištění žánru pracuje tak, že si vyextrahujeme vektory souboru a poté procházíme všechny žánry. U každého z nich si pro každou píseň uloženou v databázi spočteme vzdálenost mezi ní a načteným souborem. Tu získáme součtem normalizovaných euklidovských vzdáleností vybraných deskriptorů. Výsledek pak porovnáme se stanovenou konstantou, která taktéž vznikla po analýze naměřených hodnot. Pokud je vzdálenost nižší, tak se píseň považuje za podobnou. Po projetí všech písní tímto procesem vezmeme počet podobných a podělíme ho počtem všech písní daného žánru v databázi. Tím zjistíme procentuální zastoupení. Výsledek je pak váhou daného žánru.

Po zpracování všech žánrů pak jednoduše spočteme z kolika procent je nahraná píseň podobná jednotlivým kategoriím.

Ukázky aplikace

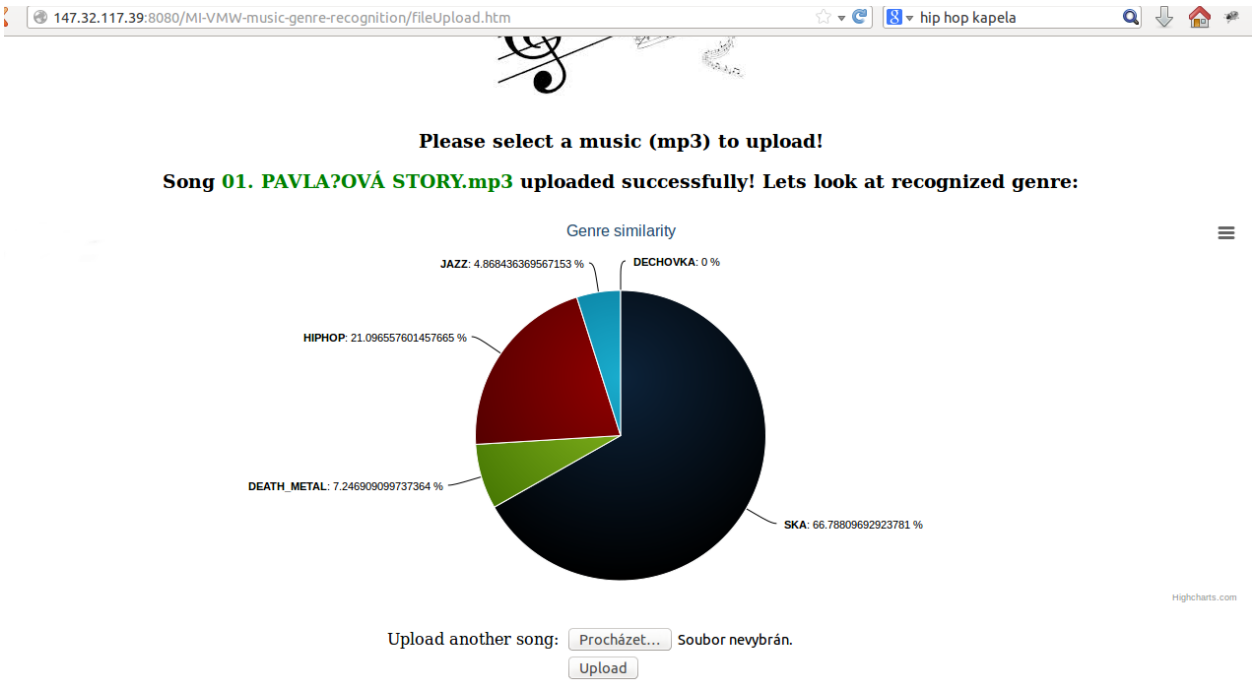
Vstup

Jednoduchý formulář pro nahrání audio souboru. Na vstupu je píseň Pavlačová story od české SKA kapely Tleskač.



Výstup

Na výstupu je znázorněna příslušnost skladby do daného žánru. V tomto případě bylo rozhodnuto, že píseň spadá do žánru SKA s příslušností cca 66%. Toto je ten **nejideálnější případ** výstupu :). Občas se aplikace zmýlí a o žánru nedokáže rozhodnout správně. Ve většině případů ale dokáže určit žánr alespoň na druhém či třetím místě (nebo alespoň na čtvrtém) s odchylkou několika procent od prvního žánru. Více viz kapitola o testování a diskuse na závěr.



Testování aplikace

Rozpoznávání žánrů

Provedeme si test na třech ne tak úplně náhodně vybraných písních. Ne tak úplně náhodně znamená to, že jsem si vybral 3 písně od 3 interpretů, kteří nejsou uloženi v databázi, a po poslechu písně mi přišly jako typičtí představitelé žánru, které by program mohl správně klasifikovat.

Test žánru ska

- **Tleskač - Krutej John**
  - hip-hop 34.2%
  - **ska 32.9%**
  - jazz 18.5%

    - dechovka 8.1%
    - death metal 6.3%
- **Sto zvířat - Nikdy nic nebylo**
    - **ska 54.7%**
    - hip-hop 45.3%
    - dechovka 0.0%
    - jazz 0.0%
    - death metal 0.0%
- **Fidel Castro - Skandál**
    - dechovka 28.0%
    - jazz 23.8%
    - **ska 21.2%**
    - death metal 17.9%
    - hip-hop 9.2%

Test žánru death metal

- **Napalm Death - Negative approach**
    - **death metal 57.5%**
    - hip-hop 23.9%
    - ska 10.8%
    - dechovka 7.7%
    - jazz 0%
- **Morbid Angel - Radikult**
    - ska 35.4%
    - **death metal 34.4%**
    - hip-hop 18.0%
    - jazz 6.5%
    - dechovka 5.8%
- **Deicide - Homepage for Satan**
    - **death metal 34.1%**
    - ska 28.9%
    - dechovka 14.6%
    - hip-hop 14.2%
    - jazz 8.2%

Test žánru jazz

- **Laco Deczi - Bengy**
    - dechovka 29.3%
    - **jazz 25.5%**
    - ska 19.7%
    - hip-hop 15.4%

    - death metal 10.1%
- **LOUIS ARMSTRONG _ ST. JAMES INFIRMARY**
    - dechovka 34.4%
    - **jazz 28.1%**
    - ska 16.9%
    - hip-hop 14.3%
    - death metal 6.3%
- **Caecilie Norby - The Dead Princess**
    - **jazz 30.9%**
    - death metal 27.8%
    - dechovka 26%
    - ska 11%
    - hip-hop 4.5%

Test žánru hip-hop

- **Moja Rec - What Up**
    - **hip-hop 66.1%**
    - ska 16.9%
    - dechovka 13%
    - jazz 4%
    - death metal 0%
- **Řezník - Ta holka v mým sklepě**
    - ska 21.0%
    - death metal 20.4%
    - **hip-hop 20.0%**
    - dechovka 19.5%
    - jazz 19.0%
- **RYTMUS - KRÁL**
    - ska 51.3%
    - **hip-hop 48.7%**
    - dechovka 0%
    - jazz 0%
    - death metal 0%

Test žánru dechovka

- **Eva a Vašek - Bílá orchidej**
    - jazz 32.0%
    - **dechovka 26.7%**
    - ska 22.6%
    - death metal 10.1%

- **Babouci - Meclovská polka**
    - **dechovka 33.5%**
    - jazz 29%
    - ska 17.4%
    - hip-hop 9.4%
    - death metal 10.7%
- **Moravanka - Před naším je zahrádka**
    - **dechovka 28.3%**
    - jazz 22.9%
    - ska 21.6%
    - hip-hop 14.1%
    - death metal 13.1%

Experimentování s parametry

Během testování samozřejmě občas docházelo ke špatným výsledkům. Při projevení trendu jsme zkoušeli experimentovat s parametry programu. Šlo především o váhy jednotlivých deskriptorů a konstantu rozdělující písně na podobné a nepodobné. Této činnosti by bylo ale potřeba věnovat řádově více času, abychom mohli globálně výrazněji zlepšovat výsledků

Závěr a diskuse

Bylo zkoušeno samozřejmě mnohem více písní, než ty, které jsou uvedené v kapitole o testování. Ta slouží jen jakýsi přehled a získání přibližné představy o měření. Během zkoušení jsme vypozorovali, že žánr *jazz* je velice podobným dvěma dalším žánrům (dechovka a ska), a vztah samozřejmě funguje i opačně, hlavně u dechovky. Jazz ve většině případů není jako žánr na prvním místě uhodnut a dostává se před něj již zmíněné žánry, někdy i oba současně. Všechny tři žánry jsou si v základu dost podobné, neboť v nich figurují dechové nástroje, zhruba stejný důraz na beaty a mají podobný rytmus a ne moc vysoké tóny. Poměrně dobré výsledky má hudební žánr *ska*, jehož rozpoznání ve vstupní písni je téměř pokaždé mezi prvními třemi žánry. Žánr *hip-hop* je na rozpoznání náchylný píseň od písně. Z pozorování vykazuje zřejmě nejvyšší extrémy - jednou je uhodnut na prvním místě s vysokou procentuální převahou, jindy je jeho procento velice nízké (nebyly výjimkou případy pro 5% spád do žánru u hip-hopové písně). *Death metal* je rozpoznáván poměrně dobře, ve spoustě případů se umisťoval na prvním místě, ale při zkoušení písní od kapely Sepultura se nám na první pozici dostával žánr dechovka (průměr kolem 40%) a žánr death metal byl na druhém místě se ztrátou několika procent (průměr kolem 35%). Nejlepší výsledky vykazují audio soubory, které žánrově spadají do žánru *dechovka* - do tohoto žánru jsou i poté v drtivé většině případů klasifikovány.

Aplikace by do budoucna mohla být rozšířena například o možnost “doučování”. Jinými slovy, program by se na výstupu ptal, zda daný žánr uhodnul. Pokud ne, uživatel vybere, který žánr to měl být a tato informace by byla uložena do databáze. Prostor pro zlepšení výsledků by byl i v

rozšíření databáze autory. Naše měla 100 - 200 skladeb na žánr od jednotek interpretů. To jistě není dostatečná velikost pro rozpoznávání.

Zdroje

1. https://edux.fit.cvut.cz/courses/MI-VMW/_media/lectures/lecture09.pdf
2. http://www.cs.bilkent.edu.tr/~bilmdg/bilaudio-7/MPEG7.html
3. https://sitepedago.telecom-paristech.fr/front/frontoffice.php?SP_ID=2238#SR1512
4. http://en.wikipedia.org/wiki/Mel-frequency_cepstrum
5. http://jaudio.sourceforge.net/
6. http://mpeg7audioenc.sourceforge.net/