

## ISS Projekt – Query by Example

Autor: Petr Mičulek (xmicul08)

Jazyk implementace: Python3.7, distribuce Anaconda

1) (c) – pro všechny nahrávky

Název (+.wav)	Délka stopy [h:m:s:ss]	Počet vzorků
sa1	00:00:03.46	55297
sa2	00:00:02.79	44616
si1337	00:00:05.22	83581
si1967	00:00:01.17	18728
si707	00:00:04.27	68272
sx167	00:00:04.17	66795
sx257	00:00:02.03	32438
sx3	00:00:01.60	25643
sx437	00:00:03.40	54372
sx77	00:00:03.42	54745

2)

Název (+.wav)	Délka stopy [h:m:s:ss]	Počet vzorků
q1	00:00:00.70	11269
q2	00:00:01.05	16733

3)

Pro výpočet jsem použil `scipy.signal.spectrogram` s následujícími parametry:

`nperseg= int(0.025 * fs)` (=400, délka segmentu)

`noverlap= int(0.015 * fs)` (=240, délka překryvu segmentů)

`nfft=512` (velikost pro rychlou Fourierovu transformaci)

Hodnota `nfft` byla vybrána dle nápovědy v zadání úkolu č. 4

(`nfft = 512` →  $512/2 + 1 = 257$  frekvencí)

4)

Na maticové řešení jsem nepřišel, hodnoty features získány skrz tři vnořené for cykly. Řádek nultý (z původních 257), obsahující nulovou frekvenci, jsem zahodil. Zbylých 256 řádků posčítáno do features o 16 řádcích.

5)

Využíval jsem “projíždění” query okolo nahrávky. Korelační koeficienty jsem počítal pouze dokud se query plně překrývala s nahrávkou (na základě znalosti, že hledané slovo bude v nahrávce celé). Posledních  $N_q$  hodnot ve scores tedy zůstává nulových, což vhodně ukazují grafy níže. Jako velikost kroku v rámci věty jsem použil posun o 5 rámců.

Skóre je počítáno jako suma Pearsonových korelačních koeficientů, vypočítaných skrz `scipy.stats.pearsonr`. Tato suma je pro normalizaci hodnot vydělena délkou query.

6)

Všechny grafy jsou níže.

7)

V hodnotách score očekáváme peak v místě přesného překryvu začátku hledaného slova a query. Pro hledání takovýchto peaků používám `scipy.signal.find_peaks`, ale nejspíše by stačilo práhování konstantní hodnotou. Jako parametry `find_peaks` jsem zvolil `height=0.5` (hodnota korelace) a `threshold=0.005` (hodnota, o kterou se musí peak lišit od svého okolí). Obě hodnoty byly zvoleny spíše empiricky a vzhledem k spokojenosti s výsledky jsem s nimi dále neexperimentoval. Tyto peaky jsou zobrazeny v grafech, společně se zaznačením rozsahu nalezeného slova.

8)

Věty

Název nahrávky	Obsah	Q1 nalezena	Q2 nalezena
sa1	She had your dark suit in greasy wash water all year.	ne	ne
sa2	Don't ask me to carry an oily rag like that.	ne	ne
si1337	This was chiefly because of the bluish white autofluorescence from the cells.	ne	ne
si1967	Twelve o'clock level.	ne	ne
si707	The avocado should have a give to it, as you hold it, when it is ripe.	ne	ne
sx257	We are open every Monday evening.	ne	ne
sx3	This was easy for us.	ne	ne
sx3	They used an aggressive policeman to flag thoughtless motorists.	ne	ne
sx77	Bagpipes and bongos are musical instruments.	<b>ano</b>	ne
sx167	Employee layoffs coincided with the company's reorganization.	ne	<b>ano</b>

Query = vyhledávaná slova

Název	Obsah	Počet nálezů
q1	instruments	1
q2	reorganization	1

Nálezy

Název query	Název nahrávky	Vzorky - počátek	Vzorky - konec	Správnost nálezu
q2	sx167	49090	65823	ano
q1	sx77	43473	54742	ano

správnost nálezu ověřena poslechem exportovaného segmentu

9)

Řešení bylo nad všechna očekávání úspěšná. Výsledky dokonce ukazovaly ještě ostřejší peak v grafu, dokud jsem velikost kroku měl nastavenou na 1. Z toho vyvozují, že jsem měl štěstí s volbou query a s konzistencí ve stylu řeči. Detektor by se určitě dal zlepšit metodami zmíněnými níže v protokolu, případně jsem přemýšlel o přístupu, kde bych zkoušel získat data více relativního charakteru (např. vývoj nejsilnějších frekvencí v čase – něco jako derivace?), aby by se dala lépe hledat podobnost např. v intonaci.

(hlas se zhoupne → funkce nejsilnějších frekvencí by měla obsahovat také obsahovat konvexní/konkávní úsek)









