

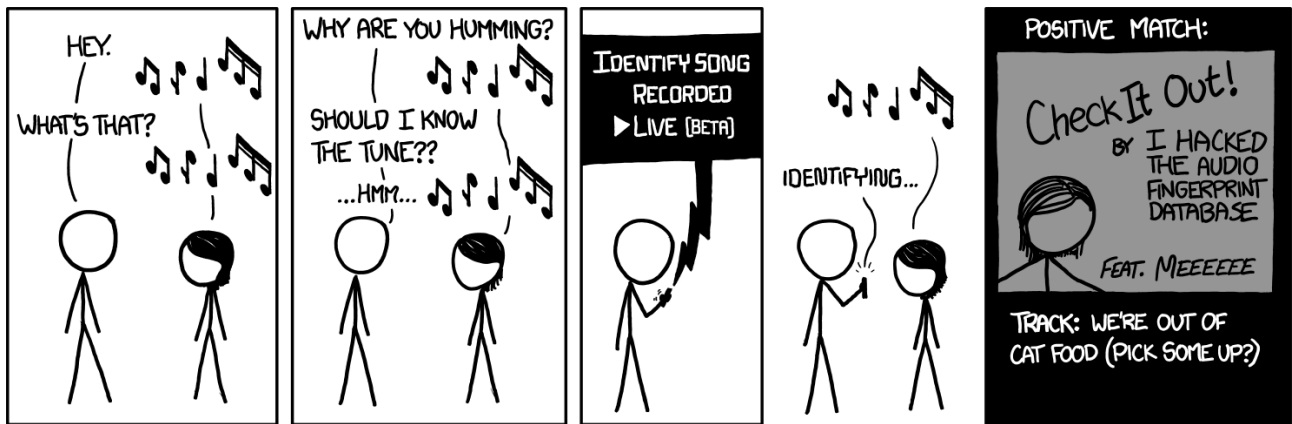


Concours national d'informatique

Épreuve écrite d'algorithmique
Louvain-la-Neuve

9 mars 2019

ALAKAZAM



1 Préambule

Bienvenue à **Prologin**. Ce sujet est l'épreuve écrite d'algorithmique et constitue la première des trois parties de votre épreuve régionale. Sa durée est de 3 heures. Par la suite, vous passerez un entretien (20 minutes) et une épreuve de programmation sur machine (4 heures).

Conseils

- Lisez bien tout le sujet avant de commencer.
- **Soignez la présentation** de votre copie.
- N'hésitez pas à poser des questions.
- Si vous avez fini en avance, relisez bien, ou préparez votre présentation pour l'entretien.
- N'oubliez pas de passer une bonne journée.

Remarques

- Le barème est donné à titre indicatif uniquement.
- Indiquez lisiblement vos nom et prénom, la ville où vous passez l'épreuve et la date en haut de votre copie.
- Tous les langages sont autorisés, veuillez néanmoins préciser celui que vous utilisez.
- Ce sont des humains qui lisent vos copies : laissez une marge, aérez votre code, ajoutez des commentaires (**seulement** lorsqu'ils sont nécessaires) et évitez au maximum les fautes d'orthographe, sinon ça va barder.
- Le barème récompense les algorithmes les plus efficaces : écrivez des fonctions qui trouvent la solution le plus rapidement possible.
- Si vous trouvez le sujet trop simple, relisez-le, réfléchissez bien, puis dites-le-nous, nous pouvons ajouter des questions plus difficiles.

2 Abracadabra

Félicitations! Vous venez de créer votre start-up. Ou plutôt, vous avez plein d'argent sur les bras, et vous aimeriez en profiter pour vous consacrer à votre grande passion : l'algorithmique.

La magie n'existe pas, c'est bien connu. Il s'agit juste d'un subterfuge pour séduire des investisseurs ou des clients et vendre des goodies (ex. la magie de Noël, la saga Harry Potter, etc.). Au fondement il y avait les algorithmes, et des gens en costume se sont mis à les breveter, et les vendre. Pourquoi pas vous ?

En l'occurrence, vous aimeriez bien construire une application qui permet d'identifier une musique à partir d'un extrait. Ça tombe bien, personne n'y avait pensé avant !

3 Hocus Pocus

Il faut bien commencer quelque part, et écrire du code le plus vite possible¹, parce que mine de rien je vous fais perdre du temps avec mes pavés de texte inutiles. C'est pourquoi vous allez vous échauffer sur une version simplifiée.

On suppose dans un premier temps que les musiques sont encodées comme chaînes de caractères composées de lettres de **a** à **g** (de *la* à *sol*) et d'espaces (les silences). Par exemple, "**facade effacee**" correspond à la musique « *fa la do la ré mi* (pause) *mi fa fa la do mi mi* » où l'on reconnaît bien sûr la deuxième Chaconne de Bach, en écoutant attentivement.

Question 1

(2 points)

Écrivez une fonction `correspond(motif, candidat)` qui renvoie `True` si la musique `motif` est incluse dans la musique `candidat`. Pour cette question vous ne pouvez pas utiliser le mot-clé `in` de Python, ou autre méthode rendant cette fonction plus triviale que nécessaire, et ce point gratuit.

4 Habeas Corpus

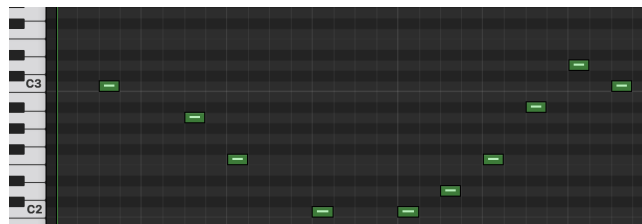


FIGURE 1 – Un sceptro... spectra... bref, un diagramme temps-fréquence, quoi.

Dans la vraie vie, les musiques sont plutôt représentées comme à la Figure 1. L'axe des abscisses représente le temps en millisecondes tandis que l'axe des ordonnées représente les fréquences en Hz des notes importantes de la musique. On appelle cela un spectrogramme, ou un diagramme temps-fréquence.

Question 2

(1 point)

Quelle structure de données choisir pour représenter un spectrogramme ? Lisez les questions qui suivent avant de formuler votre réponse.

Question 3

(2 points)

Réécrivez votre fonction `correspond(motif, candidat)` sur des spectrogrammes, de façon naïve.

Question 4

(2 points)

Quel est le nombre d'opérations de votre fonction s'il n'y a que 512 fréquences possibles, que les temps sont exprimés en millisecondes, qu'on a 300 points dans le spectrogramme toutes les 10 secondes, que l'extrait est de 10 secondes et que la musique dure 5 minutes ?

Ce nombre d'opérations risque de s'avérer impraticable pour des millions de musiques. Cherchons mieux.

1. Noon, ne faites pas ça !

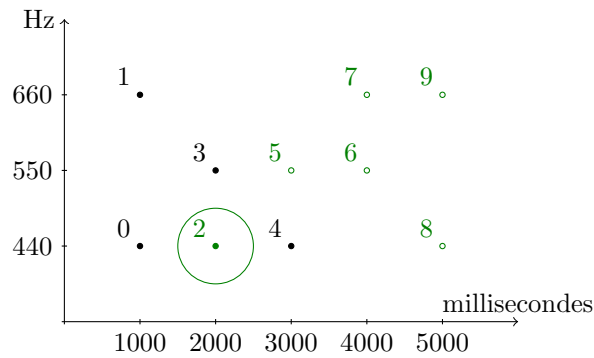


FIGURE 2 – Un exemple de zone, où `TAILLE_ZONE` vaut 5 et `DISTANCE_ANCRE` vaut 3.

5 Habemus Papam

Question 5

(1 point)

Écrivez une fonction qui numérote les points d'un spectrogramme à partir de 0, par fréquence croissante puis temps croissant. Vous avez le droit d'utiliser une fonction de tri `sort` pour répondre à cette question.

À présent on suppose que les points d'un spectrogramme sont donnés dans l'ordre de leurs numéros. Nous considérons des zones de `TAILLE_ZONE` points consécutifs ancrées en un point qui se trouve `DISTANCE_ANCRE` points avant le premier point de la zone. En d'autres termes, le point numéro i sera l'ancre de la zone qui s'étend du point $i + \text{DISTANCE_ANCRE}$ au point $i + \text{DISTANCE_ANCRE} + \text{TAILLE_ZONE}$.

Afin de créer une empreinte pour une musique, nous souhaitons identifier pour chaque point du spectrogramme les informations suivantes :

- la fréquence en Hz de l'ancre d'une zone dans laquelle se trouve le point ;
- la fréquence du point considéré ;
- la différence en temps en millisecondes entre l'ancre et le point.

On appelle un tel triplet une miette², et l'ensemble de toutes les miettes d'une musique forme une empreinte. Par exemple, pour l'image à la Figure 2, le point 2 est l'ancre de la zone [5, 9]. Donc la miette correspondant à la paire de points (2, 5) pour cette zone est (440, 500, 1000) et la miette correspondant à la paire (2, 8) pour cette zone est (440, 440, 3000). Un autre exemple de zone est [4, 8] ancrée par le point 1.

Question 6

(1 point)

Quelle est la taille en bits d'une miette³, sachant qu'il y a 512 fréquences possibles et qu'on suppose qu'il ne s'écoule jamais plus de 16 secondes entre un point et l'ancre d'une zone dans laquelle il se trouve ? Pourquoi est-ce pratique ?

Question 7

(3 points)

Écrire une fonction qui prend en argument un spectrogramme et renvoie son empreinte.

Question 8

(2 points)

Quel est le nombre d'opérations effectuées par votre fonction si `TAILLE_ZONE` est de 5, `DISTANCE_ANCRE` de 3 et qu'il y a 300 points dans un extrait de 10 secondes ?

Vous avez accès à une base de données `bdd` qui est une liste de spectrogrammes de musiques, identifiées par leur position dans la liste. Nous entrons dans le vif du sujet.

². Oui bon écoutez, j'étais pas inspiré ce matin.

³. « Quelle Est La Taille En Bits D'Une Miette ??? » *I need this on a T-shirt.* — Steven Moffat

Question 9

(2 points)

Écrivez une fonction qui détermine efficacement les identifiants des musiques de la base de données qui partagent au moins k miettes avec la cible.

Nous devons à présent identifier un score de pertinence des résultats.

Question 10

(2 points)

Est-ce que compter le nombre de miettes en commun suffit à identifier la musique dans la base de données? Si oui expliquez pourquoi, sinon donnez un contre-exemple.

Question 11

(3 points)

Définissez une fonction de pertinence pour ordonner les résultats de votre méthode et justifiez la pertinence de votre fonction de pertinence. Utilisez-la pour écrire une fonction qui renvoie les 10 résultats les plus pertinents de la base de données de musiques.

Question 12

(3 points)

Donnez une borne supérieure pour le nombre d'opérations effectuées par votre fonction au maximum. Pour simplifier, on supposera qu'il y a (au moins) 2 valeurs possibles pour la différence en temps en millisecondes des miettes, et que la répartition des points est uniforme. Comparez-la avec la réponse de la question 4.

Question 13

(4 points)

Que manque-t-il à votre algorithme pour fonctionner sur des musiques chantonnées par des humains? Donnez au moins trois pistes d'amélioration, et donnez des idées pour les résoudre.

6 Lorem Ipsum

Question bonus 14

(2 points)

Racontez-nous un souvenir malheureux de votre enfance.

Question bonus 15

(1 point)

Quelle est votre Chaconne préférée de Bach et pourquoi?

Question bonus 16

(0 point)

Continuez l'adage populaire :

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Le sujet comporte 4 pages (sans compter la page de garde), 16 questions, et 3 questions bonus. Les questions normales sont notées sur 28 points, plus les questions bonus et 1 point de présentation.