



Mathématiques discrètes

Projet : Algorithme de Shannon-Fano

Consignes Le but du projet est de présenter une application dans laquelle les mathématiques discrètes jouent un rôle fondamental.

Le rendu final du projet consistera en un article destiné au grand public au format pdf de 800-1000 mots plus une annexe numérique, qui pourra contenir par exemple une démonstration interactive, une vidéo explicative et/ou des graphiques générés par du code écrit par vous-même ; cette annexe sera rendue sous la forme d'un lien vers un dépôt en ligne. La forme exacte et la technologie utilisée pour l'annexe peut varier et est donc laissée au libre choix des étudiants. L'article et son annexe seront jugés non seulement sur le contenu mais aussi sur la clarté de la présentation, la qualité de rédaction, et la créativité.

Contenu Le sujet détaille quelques points à développer mais ceux-ci sont seulement proposés comme point de départ de votre travail. Vous êtes encouragés à développer d'autres pistes en lien avec les mathématiques discrètes. De même, la bibliographie conseillée est un point de départ. Vous pouvez vous appuyer sur d'autres sources sur lesquelles vous porterez un œil critique et que vous prendrez soin de citer correctement.

Charte de bonne conduite Lisez attentivement la charte de bonne conduite. Portez une attention particulière à citer toutes vos sources, y compris les exemples et les images que vous utiliserez. L'utilisation d'outils d'IA tels que ChatGPT est formellement interdite. L'équipe pédagogique sera très attentive à tous ces aspects lors de la correction.

Calendrier Consultez la page Moodle du cours pour les dates des principales étapes du projet.

Bref descriptif du sujet

La notion de codage est une notion clé de l'informatique. Pour coder un ensemble de n caractères en bits, une solution naïve consiste à prendre k bits avec $2^k > n$ puis choisir n codes parmi les 2^k codes possibles. Cependant, ce codage a deux limites :

- il ne prend pas en compte la fréquence d'apparition des caractères et peut donc ne pas être optimal
- certains codes peuvent avoir les mêmes p premiers bits : il faut attendre d'avoir lu les k bits pour pouvoir déchiffrer le caractère (avec nécessité de les stocker).

Le but de ce projet est de découvrir un algorithme permettant la construction d'un codage préfixe presque optimal étant donné un texte : l'algorithme de Shannon-Fano.

Bibliographie conseillée

- <http://mescal.imag.fr/membres/jean-marc.vincent/index.html/ALG05/Codage.pdf>
- https://fr.wikipedia.org/wiki/Codage_de_Shannon-Fano
- <https://perso.limsi.fr/anne/coursAlgo/Huffman.pdf>
- https://fr.wikipedia.org/wiki/Codage_de_Huffman

Pistes de développement

1. Introduire la notion de codage, de code préfixe.
2. Expliquer l'algorithme de Shannon-Fano.
3. Écrire un programme qui étant donné un texte renvoie l'arbre de Shannon-Fano associé et le texte codé suivant cet arbre.
4. Calculer la compression maximale obtenue en utilisant ce codage par rapport à un codage naïf.
5. (optionnel) Comparer cet algorithme avec celui de Huffman.