

Examen Python – M1 TAL

Lundi 19 novembre 2019

Vous travaillerez sur les fichiers `les_etablissements_hospitaliers_franciliens.csv` (https://data.iledefrance.fr/explore/dataset/les_etablissements_hospitaliers_franciliens/) et `lecun_pos.txt`, version annotée en POS d'une présentation de l'intelligence artificielle de Yann LeCun au collège de France (<https://www.college-de-france.fr/site/yann-lecun/Recherches-sur-l-intelligence-artificielle.htm>)

1. Sur `les_etablissements_hospitaliers_franciliens.csv`
 1. Combien y-a-t'il d'établissements ? (2 points)
 2. Compter les établissements par département (2 points)
 3. Compter les établissements de type 'Centres Hospitaliers Spécialisés Lutte Maladies Mentales' par département (4 points)
 4. Combien d'établissements ont ouvert au XXe siècle ? Au XXIe siècle ? (4 points)
2. Sur `lecun_pos.txt`
 1. Lister les 20 mots les plus fréquents du texte (3 points)
 2. Lister les mots (mot : étiquette(s)) qui se terminent par « ment » (3 points)
 3. Lister les mots (mot : étiquette(s)) du texte qui peuvent recevoir plus d'une étiquette POS (4 points)