

Approximation de la codéviance

Rémi Hutin Rémy Sun

ENS Rennes

Département Informatique et Télécommunications

25 novembre 2015

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations



- Langage utilisé : Python 3

- Intersection de nos connaissances
- Simplicité
- Structure de liste/tableau
- Nombreux modules
- Typage

- Intersection de nos connaissances
- Simplicité
- Structure de liste/tableau
- Nombreux modules
- Typage

- Intersection de nos connaissances
- Simplicité
- Structure de liste/tableau
- Nombreux modules
- Typage

- Intersection de nos connaissances
- Simplicité
- Structure de liste/tableau
- Nombreux modules
- Typage

- Intersection de nos connaissances
- Simplicité
- Structure de liste/tableau
- Nombreux modules
- Typage

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

- Données brutes, volumineuses
- Comment les exploiter ?

- Données brutes, volumineuses
- Comment les exploiter ?

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

- Première idée : simple parcours
 - Problème : code très long et peu modulaire
- Amélioration : expressions rationnelles
 - Module Python : `re`
 - Code moins efficace mais bien plus élégant

- Première idée : simple parcours
 - Problème : code très long et peu modulaire
- Amélioration : expressions rationnelles
 - Module Python : `re`
 - Code moins efficace mais bien plus élégant

- Première idée : simple parcours
 - Problème : code très long et peu modulaire
- Amélioration : expressions rationnelles
 - Module Python : `re`
 - Code moins efficace mais bien plus élégant

- Première idée : simple parcours
 - Problème : code très long et peu modulaire
- Amélioration : expressions rationnelles
 - Module Python : `re`
 - Code moins efficace mais bien plus élégant

- Première idée : simple parcours
 - Problème : code très long et peu modulaire
- Amélioration : expressions rationnelles
 - Module Python : `re`
 - Code moins efficace mais bien plus élégant

```
(?P < host > .*)\s\[ (?P < D > [0-9]*) : (?P < H > [0-9]*) :  
(?P < M > [0-9]*) : (?P < S > [0-9]*) \] \s" (?P < request > .*)  
" \s (?P < code > [0-9]*) \s (?P < size > [0-9]*)
```

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

- Objectif : fonction de hachage 2-universelle
- Contribution : création d'une classe Hash

- Objectif : fonction de hachage 2-universelle
- Contribution : création d'une classe Hash

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

X	Y	cod(X,Y)
epahttp	sdschttp	-185
sdschttp	calgary	177646
epahttp	calgary	-254

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

- nombre d'essais : 20
- $k = 16$
- $d = 0.0001$

			CM_sketch(X,Y,k,d)	
X	Y	cod(X,Y)	Moyenne	Ecart-type
epahttp	sdschttp	-185	-248344	116404
sdschttp	calgary	177646	12311366	55424
epahttp	calgary	-254	-369948	114896

- nombre d'essais : 100
- $k = 64$
- $d = 0.0001$

			CM_sketch(X,Y,k,d)	
X	Y	cod(X,Y)	Moyenne	Ecart-type
epahttp	sdschttp	-185	-34199	17683
sdschttp	calgary	177646	3513236	670
epahttp	calgary	-254	-45931	114896

- 1 Analyse
 - Choix du langage
 - Données fournies
- 2 Problèmes préliminaires
 - Parsage des fichiers
 - Fonction de hachage
- 3 Résultats
 - Analyse préliminaire
 - Analyse sur le CM-sketch
 - Observations

- Augmentation de k : résultats plus proches de la réalité
- Diminution de d : écart type plus faible

- Augmentation de k : résultats plus proches de la réalité
- Diminution de d : écart type plus faible

