プログラムの動かし方

以下の順番で表す。

コンパイルファイル

実行文

→実行結果

以下の手順でプログラムを動かす。

1. データベース作成

データベースの生成プログラムは2つある。

1. zipf分布のデータベース作成

zipf\_generation\_database\_queri.cpp

./zipf\_generation\_database\_queri ウィンドウサイズ databaseのdata数 多重度上限 要素の種類数 傾きalpha

→ zipf\_database.txt を生成(zipf分布のデータベース,queriも出力可能)

(2)実データベース作成

元データ : connect.dat, mashroom.dat(数千列)

gen\_to\_database3.cpp

./gen\_to\_database3 元データのファイル名 SWサイズ databaseのdata数 多重度上限 元データの列数(connect = 10000, mash = 4000)

→ database.txt を生成(元データのデータベース)

1. Min-hashの割り当て値生成

gmihash2.cpp

./gmihash2 ハッシュ関数の数　要素の種類数 多重度上限 シード値

→Minhash.txtを生成

1. 実行時間の計測：実行時間が表示される
2. SWMHの実行時間の計測

SWMH.cpp

./SWMH Database.txt ウインドウサイズ　Minhash.txt 多重度上限

→Minlistの長さと実行時間を表示

1. ヒストグラム参照上限値固定実験の実行時間の計測

SWMH\_fix.cpp

./SWMH\_fix database.txt　SWの数　Minhash.txt マルチセットの上限 ヒストグラムの上限値

→Minlistの長さと実行時間を表示

1. バッチSWMHの実行時間の計測

BSWMH

./BSWMH database.txt　SWの数　Minhash.txt マルチセットの上限 更新の間隔

→Minlistの長さと実行時間を表示