Mining Information and Discover Knowledge from Microblogs Data Imprint System

Introduction
imprint system, well designed system for querying, analyzing tweets microblogs which consider as stream of rich data that carries different types of information. it has three main operations: 1- locate hashtag, specify location of origin tweet, 2- show trend evolution: describe how trend spread along range of distance and specify these locations. 3- Cluster account by their interest: present the most topics user write about it.

Objective
Assist in mining information.
Discover knowledge from microblogs datasets to provide a fully description of how information start, show its evolution, specify its origin and determine its location for meet information seekers needs.

Motivation

Challenges
01 Number of user from a whole world, 339 million, 309 million user(1)
02 The size of data are big, 10 TB over 5 years, 3% processed (1)
03 The arrival rate of data is high, 5 million per day, 8.772 per second (1)
04 Number of tweet are uncountable.
05 Make data more benefit for 2030 vision.

System Overview

Imprint system query
Trend evolution
Locate hashtag
Cluster user
Python
Tools
Php
PostgreSQL

Web interface

QR code