

DISEÑO DE UNA ESTRUCTURA DE DATOS PARA MOLDEAR REDES SOCIALES

Investigadores USAL:

Director Mazza, Néstor (nestor.mazza@usal.edu.ar); Nicolet, Santiago

Resumen

Este proyecto persigue diseñar una estructura de datos que permita modelar una “red personal”, es decir, todos los contactos que tendría una persona, interpretando el concepto diferente que tienen las redes sociales, con sus variantes de presentación (perfiles) y el grado de influencia que cada contacto tiene sobre la persona centro o núcleo de la red, dada por el tipo de relación, mensajes entre ellos, menciones y cantidad de publicaciones.

Se busca establecer las reglas que identifican a cada contacto de la persona para realizar comparativas entre modelos y datos reales.

La implementación lógica utiliza un modelo de red y grafo, para relacionar las entidades y un modelo clave-valor en el almacenamiento de los datos dentro de cada entidad.

La implementación física utiliza un pequeño motor para inserciones de campos, basado en clave-valor. La estructura M, conjunto de mensajes a los que la persona objeto de estudio ha sido expuesta, se implementa en archivos de texto, donde las inserciones se realizan en forma ordenada por tiempo, siguiendo los lineamientos de las bases de datos NoSQL.

Los datos característicos de la prueba fueron tomados sobre la siguiente base: 4 clientes y 653 suscriptores entre contactos reales y simulados. Cada uno de los contactos generaba un mensaje aleatorio entre 50 y 350 caracteres y lo enviaba en un tiempo aleatorio entre cinco segundos y una hora. La duración de la prueba fue de 31 días (junio de 2016) y se obtuvo la siguiente información: mensajes generados a un promedio de 5,9 por segundo; una media de 161 caracteres por mensaje, un promedio de 493 579 mensajes por día; mensajes generados, por una cifra superior a 15,3 millones; una media de 90,49 MB generados por día, para un total de 2,65 GB de datos en archivo.

Para la prueba de esfuerzo y simulación de carga, se utilizaron dos de los clientes implementados tomando información real de las redes sociales y dos generadores de mensajes en un mes, para poder evaluar la respuesta en carga en un ambiente real.

Como conclusión podemos afirmar que a partir de la prueba de esfuerzo, la estabilidad tanto en la recepción de mensajes como en la escritura de los datos en la estructura fue satisfactoria.

Palabras clave: redes sociales; modelos; estructura; influencia

Abstract

This project aims to design a data structure that allows to model a “personal network”, this means all the contacts that a person would have, interpreting the different concept that all social networks have with their variants of presentation (profiles) and the degree of influence each contact has on

the center or core “person” of the network, given the type of relationship, messages between them, mentions, numbers of publications, etc.

It seeks to establish the rules that identify each contact of the “person” to allow comparisons between models and real data.

The logical implementation uses a network model and graph to relate the entities and a key-value model in the storage of data within each entity.

The material implementation uses a small field insertion engine based on key-value. The M structure (set of messages to which the person under study has been exposed) is implemented in text files where the insertions are made in an orderly manner by time following the guidelines of the NoSQL databases.

The characteristic data of the test were taken on: 4 clients and 653 subscribers between real and simulated contacts. Each of the contacts generated a random message ranging from 50 to 350 characters and sent it in a random time ranging from five seconds to an hour. The duration of the test was 31 days (June 2016) and the following information was obtained: messages generated at an average of 5.9 per second; an average of 161 characters per message, an average of 493,579 messages per day; messages generated, for a figure higher than 15.3 million; an average of 90.49 MB generated per day, for a total of 2.65 GB of data on file.

For the stress test and load simulation, the two implemented clients were used taking real information of the social networks and two generators of messages in a month, to be able to evaluate the answer load in a real environment

As conclusion we can affirm that from the stress test, the stability both in the reception of messages and in the writing of the data in the structure was acceptable.

Keywords: social networks; models; structure; influence