

|  |
| --- |
|  |
| ElasticSearch  Evaluación |
| 19 diciembre  Repositorios de la Información  Creado por: Eduardo Blanco Bielsa |

Ejercicio 2

Métrica GND

Se obtuvieron los siguientes resultados:

Tweet 1: Collision - No Injuries : south bound I5 just south of west bound SR60 - Fecha y autor: {'20327214', 'Fri Jun 26 19:36:14 +0000 2009'}

Tweet 2: Collision - No Injuries : south bound I5 just north of DITMAN AV - Fecha y autor: {'Fri Jun 26 00:01:11 +0000 2009', '20327350'}

Tweet 3: Hazard - Vehicle : east bound I10 TO north bound I5 connector - Fecha y autor: {'20327214', 'Fri Jun 26 20:21:12 +0000 2009'}

Tweet 4: Collision - Ambulance Responding : south bound I5 AT PENROSE ST - Fecha y autor: {'Wed Jun 24 18:16:06 +0000 2009', '20326441'}

Tweet 5: Collision - No Injuries : north bound I5 just south of north bound SR14 - Fecha y autor: {'Fri Jun 26 18:46:15 +0000 2009', '20327613'}

Tweet 6: Collision - No Injuries : south bound I5 AT PICO CANYON road - Fecha y autor: {'20327613', 'Sat Jun 27 00:01:18 +0000 2009'}

Tweet 7: [UPDATE] Collision - No Injuries : north bound I5 AT VISTA DEL LAGO road - Fecha y autor: {'Wed Jun 24 06:26:12 +0000 2009', '20327613'}

Tweet 8: [UPDATE] Collision - Ambulance Responding : south bound I5 just north of E 4TH ST - Fecha y autor: {'20327214', 'Wed Jun 24 20:01:10 +0000 2009'}

Tweet 9: Collision - Non Injury: West bound US50 to north bound I5 connector, South Sacramento http://bit.ly/O83h3 - Fecha y autor: {'Fri Jun 26 00:20:08 +0000 2009', '8867602'}

Tweet 10: Hazard : north bound SR110 AT north bound I5 - Fecha y autor: {'Fri Jun 26 14:06:09 +0000 2009', '20327214'}

Tweet 11: Animal on Road : north bound I5 AT north bound SR110 - Fecha y autor: {'20327214', 'Fri Jun 26 18:36:11 +0000 2009'}

Tweet 12: Hazard - Vehicle : north bound I5 AT north bound I710 - Fecha y autor: {'Fri Jun 26 23:01:32 +0000 2009', '20327350'}

Tweet 13: Hazard : south bound I5 TO north bound SR110 connector - Fecha y autor: {'20327214', 'Fri Jun 26 23:16:15 +0000 2009'}

Tweet 14: Hit & Run - No Injuries: South bound I5 at GARDEN (DOWNTOWN) HWY, North Sacramento http://bit.ly/eVsYD - Fecha y autor: {'Wed Jun 24 05:40:08 +0000 2009', '8867602'}

Tweet 15: Collision - No Details : south bound I5 just north of south bound I605 - Fecha y autor: {'Fri Jun 26 19:56:14 +0000 2009', '20327715'}

Tweet 16: Hazard : north bound I5 just north of ARTESIA BLVD - Fecha y autor: {'20327715', 'Wed Jun 24 16:56:02 +0000 2009'}

Tweet 17: Hazard : south bound I5 AT N SAN FERNANDO BLVD - Fecha y autor: {'Thu Jun 25 20:46:06 +0000 2009', '20326441'}

Tweet 18: Hazard : south bound I5 just south of NORWALK BLVD - Fecha y autor: {'Fri Jun 26 00:46:08 +0000 2009', '20327715'}

Tweet 19: Hazard : south bound I5 AT VISTA DEL LAGO road - Fecha y autor: {'20327613', 'Fri Jun 26 13:26:11 +0000 2009'}

Tweet 20: Hazard - Vehicle : south bound I5 ON MAGIC MOUNTAIN parkway off ramp - Fecha y autor: {'20327613', 'Wed Jun 24 23:36:01 +0000 2009'}

Métrica JLH

Si comparamos los 20 resultados de las dos métricas, vemos que son bastante semejantes, pero sin embargo, la métrica jlh es capaz de encontrar más repeticiones del mismo tuit.

Gnd

Jlh

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza baja

Si nos fijamos en este tuit, vemos como jlh encuentra más fechas y autores del mismo tuit:

Tweet 20: Hazard - Vehicle : south bound I5 ON MAGIC MOUNTAIN parkway off ramp - Fecha y autor: {'20327613', 'Wed Jun 24 23:36:01 +0000 2009'}

Ejercicio 3

La forma de producir una consulta “More Like This” de forma que se pueda emular la expansión de términos es mediante el uso de document id. De esta forma, al programa le pasamos el \_id del documento que queramos y utilizamos la siguiente consulta:

Texto

Descripción generada automáticamente

Diagrama, Dibujo de ingeniería

Descripción generada automáticamente