ElasticSearch

Evaluación

---

22 DICIEMBRE

Repositorios de la Información
Creado por: Eduardo Blanco Bielsa

Ejercicio 2

Se ha escogido el topic "derailment and collision", un accidente que sucedió en la línea del metro de Wagshington y el el que muriero cerca de 9 personas y 70 resultaron heridas.

Vamos a aplicar los siguientes juicios de relevancia:

1 si el tweets está fuertemente asociado

2 si el tweets es relevante pero no demasiado

3 a los tweets menos relevantes

4 a los no relevantes

Se considerarán de más a menos relevantes en función de la cantidad de información que den sobre el accidente.

Métrica GND

Se obtuvieron los siguientes resultados:

User: 20327214, tweet: Collision - No Injuries : south bound I5 just south of west bound SR60, Date: Fri Jun 26 19:36:14 +0000 2009   **2**

User: 20327350, tweet: Collision - No Injuries : south bound I5 just north of DITMAN AV, Date: Fri Jun 26 00:01:11 +0000 2009 **2**

User: 20327214, tweet: Hazard - Vehicle : east bound I10 TO north bound I5 connector, Date: Fri Jun 26 20:21:12 +0000 2009 **2**

User: 20326441, tweet: Collision - Ambulance Responding : south bound I5 AT PENROSE ST, Date: Wed Jun 24 18:16:06 +0000 2009   **2**

User: 20327613, tweet: Collision - No Injuries : north bound I5 just south of north bound SR14, Date: Fri Jun 26 18:46:15 +0000 2009   **2**

User: 20327613, tweet: Collision - No Injuries : south bound I5 AT PICO CANYON road, Date: Sat Jun 27 00:01:18 +0000 2009   **2**

User: 20327613, tweet: [UPDATE] Collision - No Injuries : north bound I5 AT VISTA DEL LAGO road, Date: Wed Jun 24 06:26:12 +0000 2009   **2**

User: 20327214, tweet: [UPDATE] Collision - Ambulance Responding : south bound I5 just north of E 4TH ST, Date: Wed Jun 24 20:01:10 +0000 2009   **2**

User: 8867602, tweet: Collision - Non Injury: West bound US50 to north bound I5 connector, South Sacramento http://bit.ly/O83h3, Date: Fri Jun 26 00:20:08 +0000 2009   **1**

User: 20327214, tweet: Hazard : north bound SR110 AT north bound I5, Date: Fri Jun 26 14:06:09 +0000 2009 **2**

User: 20327214, tweet: Animal on Road : north bound I5 AT north bound SR110, Date: Fri Jun 26 18:36:11 +0000 2009 **2**

User: 20327350, tweet: Hazard - Vehicle : north bound I5 AT north bound I710, Date: Fri Jun 26 23:01:32 +0000 2009 **2**

User: 20327214, tweet: Hazard : south bound I5 TO north bound SR110 connector, Date: Fri Jun 26 23:16:15 +0000 2009 **2**

User: 8867602, tweet: Hit & Run - No Injuries: South bound I5 at GARDEN (DOWNTOWN) HWY, North Sacramento http://bit.ly/eVsYD, Date: Wed Jun 24 05:40:08 +0000 2009 **1**

User: 20327715, tweet: Collision - No Details : south bound I5 just north of south bound I605, Date: Fri Jun 26 19:56:14 +0000 2009 2

User: 20327715, tweet: Hazard : north bound I5 just north of ARTESIA BLVD, Date: Wed Jun 24 16:56:02 +0000 2009 3

User: 20326441, tweet: Hazard : south bound I5 AT N SAN FERNANDO BLVD, Date: Thu Jun 25 20:46:06 +0000 2009 3

User: 20327715, tweet: Hazard : south bound I5 just south of NORWALK BLVD, Date: Fri Jun 26 00:46:08 +0000 2009 3

User: 20327613, tweet: Hazard : south bound I5 AT VISTA DEL LAGO road, Date: Fri Jun 26 13:26:11 +0000 2009 2

User: 20327613, tweet: Hazard - Vehicle : south bound I5 ON MAGIC MOUNTAIN parkway off ramp, Date: Wed Jun 24 23:36:01 +0000 2009 2

Métrica Mutual Information

User: 20328183, tweet: Collision - Ambulance Responding : north bound SR170 just south of ROSCOE BLVD, Date: Thu Jun 25 00:01:11 +0000 2009 2

User: 20328099, tweet: Collision - Ambulance Responding : south bound I405 just north of W ROSECRANS AV, Date: Sat Jun 27 00:11:12 +0000 2009 2

User: 20327828, tweet: Collision - Ambulance Responding : south bound I710 AT east bound SR91, Date: Fri Jun 26 16:56:11 +0000 2009 2

User: 20328099, tweet: [UPDATE] Collision - Ambulance Responding : south bound I405 just south of east bound SR90, Date: Wed Jun 24 02:01:08 +0000 2009 2

User: 20327350, tweet: Collision - No Injuries : south bound I5 just north of DITMAN AV, Date: Fri Jun 26 00:01:11 +0000 2009 2

User: 20327828, tweet: Collision - No Injuries : north bound I405 just south of HAWTHORNE BLVD, Date: Wed Jun 24 17:21:09 +0000 2009 2

User: 20327828, tweet: Collision - No Details : south bound I405 just north of WESTERN AV, Date: Fri Jun 26 02:31:09 +0000 2009 3

User: 20327715, tweet: [UPDATE] Collision - No Injuries : south bound I605 just north of ALONDRA BLVD, Date: Thu Jun 25 14:06:13 +0000 2009 2

User: 20327828, tweet: Hit and Run - No Injuries : south bound I405 just north of AVALON BLVD, Date: Sat Jun 27 01:31:13 +0000 2009 2

User: 20327350, tweet: Collision - No Details : south bound I710 just north of ATLANTIC BLVD, Date: Fri Jun 26 07:41:07 +0000 2009 3

User: 20327715, tweet: Collision - No Details : north bound SR57 just south of east bound SR60, Date: Fri Jun 26 09:46:07 +0000 2009 3

User: 20327214, tweet: [UPDATE] Collision - Ambulance Responding : south bound SR110 just north of 4 LEVEL, Date: Wed Jun 24 06:51:04 +0000 2009 2

User: 20327613, tweet: Collision - No Injuries : north bound I5 just south of north bound SR14, Date: Fri Jun 26 18:46:15 +0000 2009 2

User: 20327214, tweet: Collision - No Injuries : south bound I110 AT W FLORENCE AV, Date: Wed Jun 24 19:41:05 +0000 2009 2

User: 20327214, tweet: [UPDATE] Collision - Ambulance Responding : south bound I5 just north of E 4TH ST, Date: Wed Jun 24 20:01:10 +0000 2009 2

User: 20327715, tweet: [UPDATE] Collision - Ambulance Responding : north bound I605 just south of ROSE HILLS road, Date: Sat Jun 27 00:56:19 +0000 2009 **2**

User: 20327350, tweet: [UPDATE] Collision - Ambulance Responding : south bound I710 just north of E IMPERIAL highway, Date: Sat Jun 27 02:21:14 +0000 2009 **2**

User: 20327350, tweet: Collision - No Details : south bound I710 AT FLORENCE AV, Date: Fri Jun 26 03:51:11 +0000 2009 **2**

User: 20327828, tweet: Hit and Run - No Injuries : south bound I405 AT ARTESIA BLVD, Date: Fri Jun 26 23:06:12 +0000 2009 **2**

User: 20328099, tweet: Collision - No Details : south bound I405 AT WILSHIRE BLVD, Date: Fri Jun 26 20:06:10 +0000 2009 **2**

Diferencias entre Gnd y Mutual Information

Gnd

Mutual Information

Por tanto, *según el criterio aplicado, tiene mayor precisión gnd, pues proporciona mayor información sobre el accidente.*

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:

```
query = {
    "query": {
        "more_like_this": {
            "fields": ["id_str", "text"],  # Los campos a usar por la query MLT
            "like": [
                {
                    "_id": doc_id
                }
            ],
            "min_term_freq": 1,  # Mínimo número de veces que el término aparece en el documento
            "min_doc_freq": 1,  # Mínimo número de documentos en los que el término aparece
            "max_query_terms": 25  # Máximo número de términos a incluir en la query MLT
        }
    }
}
```