



# *Repositorios de Información ElasticSearch Informe*

Omar Teixeira González,  
UO281847

## **Tabla de contenidos:**

Tabla de contenidos:.....	2
Comentarios de los resultados del Ejercicio 2:.....	3
Descripción del tema elegido: .....	3
Revisión de la pertinencia de los tweets: .....	3
Documento 1 “exercise2GND5.tsv”: .....	3
Documento 2 “exercise2GND15.tsv”: .....	5
Documento 3 “exercise2JLH5.tsv”: .....	6
Documento 4 “exercise2JLH15.tsv”: .....	8
Resultados generales: .....	9
Comentarios de los ejercicios:.....	10
Ejercicio 1: .....	10
Ejercicio 2: .....	10
Ejercicio 3: .....	11

## Comentarios de los resultados del Ejercicio 2:

### Descripción del tema elegido:

He elegido el tema de “*talibán*”, que tiene relación con aquellos conflictos en el sur de Afganistán, debido a que un talibán representa a un seguidor de una de las interpretaciones ultraconservadoras del Islam, de esta manera, cualquier tema relacionado con conflictos bélicos en Pakistán o similares tendrá relevancia con respecto al tema elegido.

### Revisión de la pertinencia de los tweets:

Para determinar la relevancia de los tweets en cuestión se han buscado todos aquellos que tengan que ver necesariamente con el tema buscado o los directamente relacionados, como pueden ser guerras en zonas arábicas, intervenciones militares de EEUU, o actos terroristas realizados por los mismos.

Como se puede ver, lo elegido tiene alta relación con talibanes, por lo que los tweets que se consideren pertinentes respetarán este criterio que he seguido.

### Documento 1 “exercise2GND5.tsv”:

- 45627765 Wed Jun 24 00:32:17 +0000 2009 Taliban assassination could set back Pakistan's efforts - <http://bit.ly/lyPQL>

**Guarda relación con el tema.**

- 17993414 Wed Jun 24 00:32:22 +0000 2009 Pakistani team should stop thanking Allah for their victories that is unless they start blaming him for their losses too..

**Guarda relación con el tema.**

- 6820682 Wed Jun 24 00:36:06 +0000 2009 Someone trying to kidnap me to Pakistan or India

**Relación relativa con el tema (un 50/50), ya que puede tener que ver con talibanes debido a la connotación de crimen que puede llevar, no obstante (tras buscar el tweet en Twitter) puedo concluir que no tiene pinta de que guardase tanta relación.**

- 19224128 Wed Jun 24 00:38:28 +0000 2009 Assassination of Taliban factional leader exposes rifts among militants in Pakistan ( <http://bit.ly/XdWF8> )

**Guarda relación con el tema.**

- 37101901 Wed Jun 24 00:40:06 +0000 2009 Pakistani family's house in UK set on fire, 5 injured <http://www.despardes.com/flash.html#1>

**Relación relativa con el tema, ya que puede ser un acto cometido por talibanes, o puede ser un mero accidente en la zona.**

- 14502110 Wed Jun 24 01:01:31 +0000 2009 #Breaking Suspected U.S. Strike Kills at Least 60 in Pakistan via @NYT.com <http://digg.com/d1uq0A?t>

**Guarda relación con el tema.**

- 4047721 Wed Jun 24 01:16:02 +0000 2009 Pakistan has blocked France24 for reporting the Agosta kickback story <http://bit.ly/uOMA7> -- so it must be true!

**Guarda relación con el tema.**

- 16386704 Wed Jun 24 01:30:28 +0000 2009 Pakistani Taliban Amputate Hands of Man Convicted of Stealing.... <http://tinyurl.com/l8lhnc>

**Guarda relación con el tema.**

- 23154878 Wed Jun 24 01:35:38 +0000 2009 I hate how my mom calls everyone from the middle east "taliban" &gt;:|  
**Baja relación con el tema, a pesar de que tiene que ver con talibanes, no tiene que ver con lo que se buscaba, que eran los conflictos producidos y derivados por los talibanes.**
- 24501645 Wed Jun 24 01:43:58 +0000 2009 India Issues Arrest Warrants For 22 Pakistani Nationals Over Mumbai Terror Attacks: (RTTNews) - A court in India..  
<http://tinyurl.com/nrxt8f>  
**Guarda relación con el tema.**
- 24501645 Wed Jun 24 01:44:05 +0000 2009 Assassination in Pakistan exposes Taliban rifts: PESHAWAR, Pakistan (AP) - The assassination of the leader of a ..  
<http://tinyurl.com/mra9nb>  
**Guarda relación con el tema.**
- 16873773 Wed Jun 24 01:51:55 +0000 2009 @rachelm its not indo/pakistani curry, its a yellow east asian curry base, no curry leaves, mostly lemon grass and garlic  
**No guarda relación con el tema.**
- 28091955 Wed Jun 24 02:02:05 +0000 2009 Assassination in Pakistan exposes Taliban rifts (AP) <http://cli.gs/Gung8> #linktweet  
**Guarda relación con el tema.**
- 28091955 Wed Jun 24 02:09:20 +0000 2009 New groups takes credit for Pakistan blast <http://cli.gs/BBLn0> #linktweet  
**Guarda relación con el tema.**
- 38068031 Wed Jun 24 02:17:56 +0000 2009 New blog post: "Fissure" Premium Theme Available for Sale Tomorrow <http://bit.ly/MKtsF>  
**No guarda relación con el tema.**
- 36725782 Wed Jun 24 02:22:13 +0000 2009 BBC: Militant nexus: The ties on either side of Afghan-Pakistan border <http://kl.am/VZ5>  
**Guarda relación con el tema.**
- 22760040 Wed Jun 24 02:31:35 +0000 2009 Report: New York Times Bribed Taliban in Effort to Recover ... <http://bit.ly/QFMrr>  
**Guarda relación con el tema.**
- 49424638 Wed Jun 24 02:42:32 +0000 2009 RT We R Iranians .We r NOT Arabs or Afghans or Pakistanis. WE DO NOT KILL EACH OTHER. WE R UNITED SONS AND DAUGHTERS OF KOUROSH#iranelection  
**Guarda relación con el tema.**
- 27981558 Wed Jun 24 02:44:46 +0000 2009 Iran no embassy in U,S, Irainain diplomats work out of the Embassy of Pakistan call / e-mail at followin [12029564990request@daftar.org](mailto:12029564990request@daftar.org)  
**Guarda relación con el tema.**
- 16398222 Wed Jun 24 02:50:36 +0000 2009 MEMEORANDUM: Rival of Pakistan Taliban Leader Is Assassinated (New York Times): New York Times: Ri..  
<http://tinyurl.com/lu3jhc>  
**Guarda relación con el tema.**

## Documento 2 “exercise2GND15.tsv”:

- 45936227 Wed Jun 24 00:26:29 +0000 2009 Treasury Board pegs Afghan tab \$1.35B higher: The Treasury Board says that the cost of Canada’s military mission..  
<http://tinyurl.com/m2desn>  
**Guarda relación con el tema.**
- 45627765 Wed Jun 24 00:32:17 +0000 2009 Taliban assassination could set back Pakistan's efforts - <http://bit.ly/lyPQL>  
**Guarda relación con el tema.**
- 17993414 Wed Jun 24 00:32:22 +0000 2009 Pakistani team should stop thanking Allah for their victories that is unless they start blaming him for their losses too..  
**Guarda relación con el tema.**
- 6820682 Wed Jun 24 00:36:06 +0000 2009 Someone trying to kidnap me to Pakistan or India  
**Relación relativa con el tema (un 50/50), ya que puede tener que ver con talibanes debido a la connotación de crimen que puede llevar, no obstante (tras buscar el tweet en Twitter) puedo concluir que no tiene pinta de que guardase tanta relación.**
- 47221420 Wed Jun 24 00:36:49 +0000 2009 Hey #1up How many people can occupy an airplane/bomber? (#GameNight live &gt; <http://ustre.am/1Doi>)  
**No guarda relación con el tema.**
- 19224128 Wed Jun 24 00:38:28 +0000 2009 Assassination of Taliban factional leader exposes rifts among militants in Pakistan ( <http://bit.ly/XdWF8> )  
**Guarda relación con el tema.**
- 37101901 Wed Jun 24 00:40:06 +0000 2009 Pakistani family’s house in UK set on fire, 5 injured <http://www.despardes.com/flash.html#1>  
**Relación relativa con el tema, ya que puede ser un acto cometido por talibanes, o puede ser un mero accidente en la zona.**
- 14082590 Wed Jun 24 00:40:45 +0000 2009 @feednote\_test3 Militant nexus - The ties on either side of Afghan-Pakis - BBC News | News Front Page | World Edition <http://minu.ws/Z108s>  
**Guarda relación con el tema.**
- 19393033 Wed Jun 24 00:44:39 +0000 2009 Three German soldiers killed in Afghanistan: The soldiers were conducting a joint operation with Afghan forces w..  
<http://bit.ly/6xst7>  
**Guarda relación con el tema.**
- 19393033 Wed Jun 24 00:45:10 +0000 2009 Three German soldiers killed in Afghanistan: The soldiers were conducting a joint operation with Afghan forces w..  
<http://bit.ly/xKK7A>  
**Guarda relación con el tema.**
- 36216795 Wed Jun 24 00:49:41 +0000 2009 Infant mortality in Afghanistan is 15%! Why? Mothers still give birth at home w/o adequate care. See what HOPE is doing <http://trunc.it/jlyx>  
**No guarda relación con el tema.**
- 40549958 Wed Jun 24 00:49:43 +0000 2009 Handoff to Whitaker. Touchdown. Bombers couldn't stop a thing on that drive.  
**No guarda relación con el tema.**

- 15955099 Wed Jun 24 00:56:49 +0000 2009 Xbox Live: soadrocksever is currently Online. Playing Call of Duty 4. Operating an AC-130 gunship. (Xbox Live Nation)  
**No guarda relación con el tema.**
- 3322361 Wed Jun 24 00:59:27 +0000 2009 Should be lining a purse. Am working on an afghan instead...  
**No guarda relación con el tema.**
- 14502110 Wed Jun 24 01:01:31 +0000 2009 #Breaking Suspected U.S. Strike Kills at Least 60 in Pakistan via @NYT.com <http://digg.com/d1ug0A?t>  
**Guarda relación con el tema.**
- 14624359 Wed Jun 24 01:03:32 +0000 2009 [CNN.com]U.S. to restrict airstrikes in Afghanistan <http://bit.ly/UZjbX>  
**Guarda relación con el tema.**
- 18475875 Wed Jun 24 01:05:11 +0000 2009 Sudan's 'Coordinated' Genocide in Darfur Is Over, U.S. Envoy Says  
[www.flashqates.com/news.php](http://www.flashqates.com/news.php)  
**No guarda relación con el tema.**
- 4047721 Wed Jun 24 01:16:02 +0000 2009 Pakistan has blocked France24 for reporting the Agosta kickback story <http://bit.ly/uOMA7> -- so it must be true!  
**Guarda relación con el tema.**
- 40549958 Wed Jun 24 01:18:40 +0000 2009 Bombers lead 26 to 25  
**No guarda relación con el tema.**
- 15412003 Wed Jun 24 01:24:09 +0000 2009 new photo created <http://www.motivatedphotos.com/?id=25728> "Photo bomber" using tags photo bomber  
**No guarda relación con el tema.**

### Documento 3 “exercise2JLH5.tsv”:

- 45627765 Wed Jun 24 00:32:17 +0000 2009 Taliban assassination could set back Pakistan's efforts - <http://bit.ly/lyPQL>  
**Guarda relación con el tema.**
- 17993414 Wed Jun 24 00:32:22 +0000 2009 Pakistani team should stop thanking Allah for their victories that is unless they start blaming him for their losses too..  
**Guarda relación con el tema.**
- 6820682 Wed Jun 24 00:36:06 +0000 2009 Someone trying to kidnap me to Pakistan or India  
**Relación relativa con el tema (un 50/50), ya que puede tener que ver con talibanes debido a la connotación de crimen que puede llevar, no obstante (tras buscar el tweet en Twitter) puedo concluir que no tiene pinta de que guardase tanta relación.**
- 19224128 Wed Jun 24 00:38:28 +0000 2009 Assassination of Taliban factional leader exposes rifts among militants in Pakistan ( <http://bit.ly/XdWF8> )  
**Guarda relación con el tema.**
- 37101901 Wed Jun 24 00:40:06 +0000 2009 Pakistani family's house in UK set on fire, 5 injured <http://www.despardes.com/flash.html#1>  
**Relación relativa con el tema, ya que puede ser un acto cometido por talibanes, o puede ser un mero accidente en la zona.**

- 14502110 Wed Jun 24 01:01:31 +0000 2009 #Breaking Suspected U.S. Strike Kills at Least 60 in Pakistan via @NYT.com <http://digg.com/d1ug0A?t>  
**Guarda relación con el tema.**
- 4047721 Wed Jun 24 01:16:02 +0000 2009 Pakistan has blocked France24 for reporting the Agosta kickback story <http://bit.ly/uOMA7> -- so it must be true!  
**Guarda relación con el tema.**
- 16386704 Wed Jun 24 01:30:28 +0000 2009 Pakistani Taliban Amputate Hands of Man Convicted of Stealing.... <http://tinyurl.com/l8lhnc>  
**Guarda relación con el tema.**
- 23154878 Wed Jun 24 01:35:38 +0000 2009 I hate how my mom calls everyone from the middle east "taliban" &gt;:|  
**Baja relación con el tema, a pesar de que tiene que ver con talibanes, no tiene que ver con lo que se buscaba, que eran los conflictos producidos y derivados por los talibanes.**
- 24501645 Wed Jun 24 01:43:58 +0000 2009 India Issues Arrest Warrants For 22 Pakistani Nationals Over Mumbai Terror Attacks: (RTTNews) - A court in India.. <http://tinyurl.com/nrxt8f>  
**Guarda relación con el tema.**
- 24501645 Wed Jun 24 01:44:05 +0000 2009 Assassination in Pakistan exposes Taliban rifts: PESHAWAR, Pakistan (AP) - The assassination of the leader of a .. <http://tinyurl.com/mra9nb>  
**Guarda relación con el tema.**
- 16873773 Wed Jun 24 01:51:55 +0000 2009 @rachelm its not indo/pakistani curry, its a yellow east asian curry base, no curry leaves, mostly lemon grass and garlic  
**No guarda relación con el tema.**
- 28091955 Wed Jun 24 02:02:05 +0000 2009 Assassination in Pakistan exposes Taliban rifts (AP) <http://cli.gs/Gung8> #linktweet  
**Guarda relación con el tema.**
- 28091955 Wed Jun 24 02:09:20 +0000 2009 New groups takes credit for Pakistan blast <http://cli.gs/BBLn0> #linktweet  
**Guarda relación con el tema.**
- 38068031 Wed Jun 24 02:17:56 +0000 2009 New blog post: "Fissure" Premium Theme Available for Sale Tomorrow <http://bit.ly/MKtsF>  
**No guarda relación con el tema.**
- 36725782 Wed Jun 24 02:22:13 +0000 2009 BBC: Militant nexus: The ties on either side of Afghan-Pakistan border <http://kl.am/VZ5>  
**Guarda relación con el tema.**
- 22760040 Wed Jun 24 02:31:35 +0000 2009 Report: New York Times Bribed Taliban in Effort to Recover ... <http://bit.ly/QFMrr>  
**Guarda relación con el tema.**
- 49424638 Wed Jun 24 02:42:32 +0000 2009 RT We R Iranians .We r NOT Arabs or Afghans or Pakistanis. WE DO NOT KILL EACH OTHER. WE R UNITED SONS AND DAUGHTERS OF KOUROSH#iranelection  
**Guarda relación con el tema.**
- 27981558 Wed Jun 24 02:44:46 +0000 2009 Iran no embassy in U,S, Irainain diplomats work out of the Embassy of Pakistan call / e-mail at followin [request@daftar.org](mailto:request@daftar.org)  
**Guarda relación con el tema.**



- 16398222 Wed Jun 24 02:50:36 +0000 2009 MEMEORANDUM: Rival of Pakistan Taliban Leader Is Assassinated (New York Times): New York Times: Ri.. <http://tinyurl.com/lu3jhc>  
Guarda relación con el tema.

## Documento 4 “exercise2JLH15.tsv”:

- 45627765 Wed Jun 24 00:32:17 +0000 2009 Taliban assassination could set back Pakistan's efforts - <http://bit.ly/lyPQL>  
Guarda relación con el tema.
- 17993414 Wed Jun 24 00:32:22 +0000 2009 Pakistani team should stop thanking Allah for their victories that is unless they start blaming him for their losses too..  
Guarda relación con el tema.
- 6820682 Wed Jun 24 00:36:06 +0000 2009 Someone trying to kidnap me to Pakistan or India  
Relación relativa con el tema (un 50/50), ya que puede tener que ver con talibanes debido a la connotación de crimen que puede llevar, no obstante (tras buscar el tweet en Twitter) puedo concluir que no tiene pinta de que guardase tanta relación.
- 19224128 Wed Jun 24 00:38:28 +0000 2009 Assassination of Taliban factional leader exposes rifts among militants in Pakistan ( <http://bit.ly/XdWF8> )  
Guarda relación con el tema.
- 37101901 Wed Jun 24 00:40:06 +0000 2009 Pakistani family's house in UK set on fire, 5 injured <http://www.despardes.com/flash.html#1>  
Relación relativa con el tema, ya que puede ser un acto cometido por talibanes, o puede ser un mero accidente en la zona.
- 19393033 Wed Jun 24 00:44:39 +0000 2009 Three German soldiers killed in Afghanistan: The soldiers were conducting a joint operation with Afghan forces w..  
<http://bit.ly/6xst7>  
Guarda relación con el tema.
- 19393033 Wed Jun 24 00:45:10 +0000 2009 Three German soldiers killed in Afghanistan: The soldiers were conducting a joint operation with Afghan forces w..  
<http://bit.ly/xKK7A>  
Guarda relación con el tema.
- 36216795 Wed Jun 24 00:49:41 +0000 2009 Infant mortality in Afghanistan is 15%! Why? Mothers still give birth at home w/o adequate care. See what HOPE is doing <http://trunc.it/jlyx>  
No guarda relación con el tema.
- 15955099 Wed Jun 24 00:56:49 +0000 2009 Xbox Live: soadrocksever is currently Online. Playing Call of Duty 4. Operating an AC-130 gunship. (Xbox Live Nation)  
No guarda relación con el tema.
- 14502110 Wed Jun 24 01:01:31 +0000 2009 #Breaking Suspected U.S. Strike Kills at Least 60 in Pakistan via @NYT.com <http://digg.com/d1ug0A?t>  
Guarda relación con el tema.
- 14624359 Wed Jun 24 01:03:32 +0000 2009 [CNN.com]U.S. to restrict airstrikes in Afghanistan <http://bit.ly/UZjbX>  
Guarda relación con el tema.



- 4047721 Wed Jun 24 01:16:02 +0000 2009 Pakistan has blocked France24 for reporting the Agosta kickback story <http://bit.ly/uOMA7> -- so it must be true!  
**Guarda relación con el tema.**
- 17770648 Wed Jun 24 01:24:32 +0000 2009 in 55 minutes, POTUS didn't mention Iraq, Afghanistan, troops or war <http://tw6.us/hO> #tcot  
**Guarda relación con el tema.**
- 37782163 Wed Jun 24 01:29:44 +0000 2009 With ping-pong and puns, soldiers stay sane in Afghanistan <http://bit.ly/xirBU>  
**Guarda relación con el tema.**
- 16386704 Wed Jun 24 01:30:28 +0000 2009 Pakistani Taliban Amputate Hands of Man Convicted of Stealing.... <http://tinyurl.com/l8lhnc>  
**Guarda relación con el tema.**
- 23154878 Wed Jun 24 01:35:38 +0000 2009 I hate how my mom calls everyone from the middle east "taliban" &gt;|  
**Baja relación con el tema, a pesar de que tiene que ver con talibanes, no tiene que ver con lo que se buscaba, que eran los conflictos producidos y derivados por los talibanes.**
- 23053999 Wed Jun 24 01:39:17 +0000 2009 U.S. to restrict airstrikes in Afghanistan: The new U.S. commander in Afghanistan plans to issue a directive tha..  
<http://tinyurl.com/n9l449>  
**Guarda relación con el tema.**
- 24501645 Wed Jun 24 01:43:58 +0000 2009 India Issues Arrest Warrants For 22 Pakistani Nationals Over Mumbai Terror Attacks: (RTTNews) - A court in India..  
<http://tinyurl.com/nrxt8f>  
**Guarda relación con el tema.**
- 24501645 Wed Jun 24 01:44:05 +0000 2009 Assassination in Pakistan exposes Taliban rifts: PESHAWAR, Pakistan (AP) - The assassination of the leader of a..  
<http://tinyurl.com/mra9nb>  
**Guarda relación con el tema.**
- 16873773 Wed Jun 24 01:51:55 +0000 2009 @rachelm its not indo/pakistani curry, its a yellow east asian curry base, no curry leaves, mostly lemon grass and garlic  
**No guarda relación con el tema.**

## Resultados generales:

En base a los resultados obtenidos, se puede calcular la precisión obtenida con cada métrica en cada documento (con distinto número de términos a expandir), de esta forma:

- Métrica GND y Expansión 5: Si tomamos estrictamente los que guardan relación, la precisión sería del **75%**, porque hay 2 tweets que guardan una relación relativa, 1 que guarda baja relación y otros 2 que no guardan ninguna relación con el tema.
- Métrica JLH y Expansión 15: Si tomamos estrictamente los que guardan relación, la precisión sería del **50%**, porque hay 2 tweets que guardan una relación relativa y 8 que no guardan ninguna relación con el tema.
- Métrica GND y Expansión 5: Si tomamos estrictamente los que guardan relación, la precisión sería del **75%**, porque hay 2 tweets que guardan una relación relativa, 1 que guarda baja relación y otros 2 que no guardan ninguna relación con el tema.
- Métrica JLH y Expansión 15: Si tomamos estrictamente los que guardan relación, la precisión sería del **70%**, porque hay 2 tweets que guardan una relación relativa, 1 que guarda baja relación y otros 3 que no guardan ninguna relación con el tema.

En base a los resultados obtenidos, puedo concluir que la mejor métrica es JLH, debido a que en la expansión con menor número de términos (5) los resultados mantienen una precisión del 75% (como en GND), sin embargo, en la expansión con 15 términos JLH es la mejor métrica a utilizar, debido a que da una precisión del 70% frente al 50% de GND.

También se puede observar que, a medida que el número de términos aumenta, la precisión disminuye, esto debido a que al expandir la consulta con términos que pueden estar menos relacionados o ser menos generales con respecto al tema principal, generarán resultados que tienen poco o nada que ver con lo buscado, un ejemplo de esto puede ser lo que se ve en los documentos del videojuego *Call of Duty 4*, el cual, tiene lugar en Afganistán, pero no guarda relación con un conflicto real de talibanes.

## **Comentarios de los ejercicios:**

Previo a la realización de los ejercicios, he leído las diapositivas de teoría y repasado los conceptos dados en prácticas de manera a tener una mayor idea a la hora de realizar el trabajo en cuestión.

### **Ejercicio 1:**

Primero he creado el índice (fichero *exercise1Indexer.py*), basándome en el código del bulk-indexer-4, se ha ajustado a los requisitos pedidos, en este caso, shingles de 2 y 3 términos y el idioma en español.

Tras crear el índice se ha pasado a la funcionalidad principal del ejercicio, que correspondía a la generación de trending topics mediante la consulta dada en las diapositivas, lo que ha generado un json de los resultados obtenidos, el cual se ha recorrido y para recolectar los resultados.

Es precisamente en el for donde se realiza todo el proceso de búsqueda de id, sinónimos y tipos, para ello (y para hacer el código más entendible), se ha creado una clase TrendingTopic que guarde su id, texto, sinónimos y tipo, ya que, de esta forma se puede filtrar el diccionario (*fecha*, *lista de trending topics*) mostrando una fecha y hora y la lista de trending topics producidos en esa hora en concreto.

Dentro de esta clase TrendingTopic se encuentran 2 métodos principales, el primero, la búsqueda de los ids mediante Wikidata, y el segundo, la búsqueda de los sinónimos y tipos mediante el id conseguido.

Por último, una vez que el diccionario tiene todos los datos, se vuelca el contenido a un fichero, pudiendo ver y buscar en los resultados, tal y como se había pedido.

### **Ejercicio 2:**

Para el ejercicio 2 he comenzado de manera idéntica al anterior, creando el índice, el cual es básicamente el bulk-indexer-3, no obstante, he optado por duplicar el código de creación del índice para respetar la estructura del código y distinguir más fácilmente entre los índices dados en las prácticas con respecto a los creados en el trabajo.

Tras realizar el ejercicio erróneamente y preguntar a los profesores de la asignatura, he entendido mejor la mecánica y/o lo que se pedía, ya que, al comienzo del ejercicio tenía una duplicación de código considerable, ya que pensaba que tenía que hacerse todo en una misma ejecución, pero, actualmente, lo he modificado para que tenga un grado de parametrización, de

manera que se pueda elegir el tema a buscar, el número de términos con el que expandir la consulta y la métrica a utilizar.

Por ello, primero se pide al usuario que introduzca éstos parámetros y se pasan a una consulta que devuelve los términos más significativos, preparándolos para la búsqueda en la siguiente consulta.

La siguiente parte del ejercicio corresponde a la búsqueda de los términos expandidos, por lo que, se realiza la búsqueda con éstos términos y se vuelcan los resultados a un fichero, fichero en el cual aparecerá el id del usuario, la fecha y hora de la publicación del tweet y el texto del tweet en concreto.

Por último, tras haber generado todos los ficheros *.tsv* necesarios (en mi caso 4, con el tema de *talibán*, las métricas *gnd* y *jlh* y el número de términos a 5 y a 15), se ha procedido a la documentación y comparación solicitada, ya vista en la sección de [Comentarios de los resultados del Ejercicio 2](#).

### Ejercicio 3:

Para realizar este ejercicio, he leído la documentación dada para tener una idea ante como proceder, después he duplicado el código del ejercicio 2, ya que ambos se basan en lo mismo y he ido modificando las cosas pertinentes.

Lo primero que se ha modificado han sido los inputs, que, en este caso han sido el cuerpo de la consulta y la expansión de términos a realizar. Con esto, he cambiado prácticamente el resto del código exceptuando cosas puntuales como la exportación a un fichero (porque sigue siendo lo mismo).

Tras obtener la query a buscar (debe de introducirse en una única línea), se parsea el json introducido como string y se le pasa a la función encargada de buscar los documentos relacionados con la query introducida (de nuevo, el tema elegido ha sido *talibán*), se obtienen los resultados y se devuelven para poder utilizarlos en la siguiente parte del ejercicio.

La siguiente parte (y última si no se considera la exportación al fichero, porque es la misma que en el ejercicio 2) corresponde a la búsqueda de tweets relacionados con los documentos obtenidos, para lo que me he basado en los códigos de ejemplo de la documentación de *More like this*, y he devuelto un volcado de los mismos.