Hacia un método de análisis del lenguaje y contenido emocional en la gestación y explosión del 15M en Twitter

En este trabajo se presentan los algoritmos y resultados del análisis cualitativo (contenido emocional) y cuantitativo (cohesión y temperatura) del lenguaje usado durante la gestación y explosión del 15M en la red de Twitter. Se aproxima la investigación comenzando por el estado del arte, para después pasar a describir las métricas habituales de éxito en este tipo de análisis. Con estas métricas como guía, se estudian las aproximaciones habituales al análisis cualitativo (manual y automático), así como su problemática asociada. A partir de este planteamiento, se razonarán las soluciones adoptadas y se hará una exposición completa de la implementación. También se introducirán las innovaciones planteadas en el análisis cuantitativo, como son la temperatura del lenguaje y su cohesión. Como se verá, el análisis cuantitativo arranca del uso de las entradas de la Wikipedia como fuente de identificación de conceptos y entidades. Finalmente, se exponen las relaciones entre lenguaje y viralidad de los mensajes, así como las conclusiones y trabajos futuros.

Óscar Marín Miró

Ingeniero de datos. Colectivo Outliers+Grupo de Investigación DatAnalysis15M

Introducción

¿Qué relación hay entre el contenido emocional de un texto y su capacidad de viralización? ¿Qué sucede con el lenguaje en las redes sociales en los momentos en los que estalla una revolución?

El presente artículo resume la aproximación, dificultades y resultados del análisis algorítmico de las emociones implicadas en los *tweets* relacionados con el 15M (desde su gestación hasta su estabilización) y su relación con la viralidad; así como la cohesión del vocabulario empleado en ellos. Ambos análisis tratan de responder las preguntas del párrafo anterior.

El *dataset* usado para ambos análisis se puede encontrar *online*[1]. Se trata del conjunto de 1.123.225 *tweets* con los *hashtags* #nolesvotes, #democraciarealya, #spanishrevolution, #acampadasol, #15m, #yeswecamp, #tomalacalle entre el 31 de Marzo de 2011 y el 8 de Julio de 2011.

Estado del arte en el análisis de texto

Las técnicas fundamentales implicadas en el análisis son: (a) Análisis emocional/de sentimiento y (b) Reconocimiento de conceptos/entidades

Respecto al análisis de sentimiento[2] (*sentiment analysis*); es muy difícil hablar de una 'figura de mérito' establecida o aceptada por la comunidad como representante de la calidad actual de la tecnología (Septiembre 2013).

Por un lado, existen varios tipos de análisis:

1. Análisis a nivel de documento (*Document-level Sentiment Analysis*): se tiene en cuenta la subjetividad del texto completo para derivar un único resultado.
2. Análisis a nivel de frase *(Sentence-level Sentiment Analysis*): en este caso, generamos un análisis de sentimiento diferente para cada frase.
3. Análisis a nivel de entidad (*Entity-level Sentiment Analysis*): en este caso, si encontramos dos entidades en una misma frase, para cada una de ellas haremos un análisis diferente. Por ejemplo: En la frase 'Pepsi es mejor que Coca-cola', un análisis a nivel de entidad nos arrojaría dos métricas de sentimiento: una para la entidad 'Pepsi' y otra para la entidad 'Coca-cola', idealmente sentimiento positivo para la primera y negativo para la segunda.

Por otro lado, nos enfrentamos con un problema dependiente de idioma, de tal manera que las figuras de éxito son específicas para cada idioma.

Finalmente, también nos encontramos con el problema del *dataset* de partida[1]: no es lo mismo resolver este problema sobre un conjunto de tuits (limitación de 140 caracteres), que sobre un texto editorial.

En resumen, es imposible hablar de una métrica de éxito que nos marque el estado del arte en la medición de sentimiento, puesto que hay estudios con datos de Twitter, a nivel de frase en inglés, también los hay con datos de *reviews* de películas, a nivel de documento, en inglés; y también hay estudios a nivel de entidad con datos de opiniones en español.

1 https://github.com/datanalysis15m/datasets/tree/master/oscarmarin (22/09/2013)
2 http://en.wikipedia.org/wiki/Sentiment_analysis (22/09/2013)

Hasta donde llega el conocimiento del autor, no existe un estudio que hable de un marco unificado a pesar de todas estas diferentes variables y que nos informe de los límites de la tecnología en cada uno de los casos.

En el ámbito del reconocimiento de entidades[3], nos encontramos con un problema parecido: es un problema tan dependiente del contexto lingüístico, que no se conoce un estudio que hable de un marco unificado a pesar de las diferencias de lenguaje y sobre todo del contexto (semánticas: política, deportes, ciencia, etc… o bien 'de medio': Twitter, Facebook, blogs, etc..)

Podría pensarse que la medida de 'inter-agreement'[4] de evaluadores humanos se puede tomar como techo, pero nuevamente, en ambos ámbitos, no existe un número claro y concluyente. Basta con lanzar a Google la consulta[5] para darse cuenta de lo espinoso del asunto.

El corpus como guía y la matriz de confusión como métrica de éxito

Aunque armados con algunas nociones sobre lo que se puede llegar conseguir en la medición de sentimiento/emociones y en la detección de entidades/conceptos, no disponemos de unas cifras claras.

Para avanzar en la calidad, se hace necesario, como siempre, contar con un corpus de referencia etiquetado manualmente[6]. En el caso que nos ocupa, se han elegido, de todos los tuits con contenido emocional detectado por el algoritmo en su fase inicial, 1.000 al azar, debido a las limitaciones de disposición de documentalistas en el trabajo. Por otro lado, se han elegido otros 1.000 al azar, independientemente de su etiqueta emocional (presente o no presente). De esta manera, contamos con 1.000 tuits (los primeros) para depurar la medición, y otros 1.000 (corpus de *test*) para comprobar la eficacia de las mediciones.

¿Cómo calculamos la *'bondad'* del algoritmo? Seguimos el camino clásico de evaluar la calidad del etiquetado automático: La matriz de confusión[7], junto con sus parámetros de *Recall, Precision*[8] y *F1*[9]

El problema del análisis

Con los dos instrumentos mencionados en el epígrafe anterior (corpus y matriz de confusión), podemos pasar a prototipar la solución, tomando como medida de su calidad la figura F1 sobre el corpus de test.

Antes de introducir la solución, se hace necesario dedicar unas líneas a las diferentes aproximaciones para resolver el problema de mediciones subjetivas en textos:

1. El problema del análisis manual: este tipo de análisis, obviamente, brilla por su calidad. Sin embargo, el principal problema es el tiempo empleado en el etiquetado manual de conjuntos de textos moderadamente grandes, como es el caso que nos ocupa, con más de 1 millón de documentos

3 http://en.wikipedia.org/wiki/Named-entity_recognition (22/09/2013)
4 Grado de acuerdo en la evaluación de un corpus de opiniones por diferentes evaluadores
5 https://www.google.es/search?q=interagreement+sentiment+analysis
6 http://en.wikipedia.org/wiki/Text_corpus (22/09/2013)
7 http://en.wikipedia.org/wiki/Confusion_matrix (22/09/2013)
8 http://en.wikipedia.org/wiki/Precision_and_recall (22/09/2013)
9 http://en.wikipedia.org/wiki/F1_score (22/09/2013)

2. El problema del análisis automático (lingüística): este tipo de análisis, necesita de etapas de procesado del lenguaje natural, que, independientemente de la técnica concreta usada, trata de:
    a. Normalizar el texto (pasar las formas derivadas a raíces), por ejemplo: pasar de 'árboles' a 'árbol'.
    b. Etiquetar sintácticamente las palabras. Por ejemplo, 'árboles' sería etiquetado como sustantivo, en caso de una identificación sintáctica correcta.
    c. Utilizar gramáticas que analicen las estructuras y dependencias latentes en el texto. Por ejemplo, 'el árbol refrescaba el patio', se podría identificar como un sintagma nominal ('el árbol'), seguido de otro verbal ('refrescaba') y de otro nominal ('el patio'). A partir de esta identificación de sintagmas, sería posible, mediante gramáticas, llegar a la tripleta SUJETO-VERBO-OBJETO ('el árbol', 'refrescaba','el patio'); o en su forma normalizada ('el árbol', 'refrescar', 'el patio), lo cual alivia enormemente tareas como la minería de opiniones/emociones o el reconocimiento de entidades.

Estas tres técnicas, combinadas, confieren al análisis de texto automático de una gran potencia. No obstante, cabe señalar las siguientes observaciones:

1. El proceso, en general, es computacionalmente costoso.
2. Errores introducidos en las etapas previas, afectan a las posteriores, de tal manera que un error en la identificación sintáctica se propaga en la identificación de estructuras gramaticales.
3. Los algoritmos que hay detrás de estos procesos, suelen estar entrenados en un contexto determinado y necesitan frases gramaticalmente muy correctas y completas para funcionar. Cuando nos vamos a un ámbito como Twitter, la combinación de (a) contexto diferente al del entrenamiento y (b) errores gramaticales más poco contexto (140 caracteres); hacen que la calidad del proceso completo caiga de manera estrepitosa.

1. El problema del análisis automático (aprendizaje máquina): el análisis automático basado en *Machine Learning*[10], descansa en la estadística. Los algoritmos necesitan como entrada un conjunto discreto de valores, denominados *features,* y la salida esperada (en este caso la emoción asociada). A partir de inferencia estadísticas, cuando se le presenta al sistema unas *features* extraídas de un texto 'no visto' previamente, el sistema devuelve la salida 'inferida'. Estos sistemas llevan un largo recorrido caminado y han resistido el paso del tiempo, sin embargo, en el tratamiento de texto, el problema radica en encontrar las *features* adecuadas. Para que se entienda la dificultad, se exponen dos escenarios (extremos):

- a. Las *features* son las palabras del texto: En este caso, al no entrenarse el sistema con secuencias de palabras, sino con palabras sueltas (Bag of words[11]), el sistema no distingue adecuadamente la diferencia entre 'odiar' y 'no odiar', con lo cual el rendimiento es pobre a todas luces.
- b. Las *features* son unidades gramaticales. En este caso, debería esperarse un incremento en la calidad, puesto que 'no odio' y 'odio' serían reconocidos como sintagmas diferentes. Sin embargo, al necesitarse de una etapa previa de tratamiento lingüístico, caeríamos en las mismas dificultades que en el caso 2)

Soluciones planteadas al problema del análisis

[10] http://en.wikipedia.org/wiki/Machine_learning (22/09/2013)
[11] http://en.wikipedia.org/wiki/Bag-of-words_model (22/09/2013)

1. **Expansión de raíces vs normalización de desinencias**: puesto que no podemos utilizar el procesado del lenguaje en el texto de entrada, por el ruido en el mensaje y la falta de contexto, nos topamos con la dificultad de que en los diccionarios empleados (con expresiones para cada una de las emociones) tenemos que escribir todas las desinencias de los verbos (por ejemplo). Para aligerar este problema, se ha optado por atacar la solución desde el ángulo contrario: Las listas de expresiones vienen en forma raíz (acompañadas de su función sintáctica) , y éstas se expanden con una morfología. Por tanto, si en el diccionario de 'indignación' aparece la palabra 'indignado' junto con su etiqueta de adjetivo, el sistema expandirá esta entrada en el diccionario de partida como 'indignado, indignada, indignados, indignadas', reduciendo dramáticamente (piénsese en los verbos) el tiempo de escritura y depuración del diccionario.
2. **Gramáticas ultra-ligeras**: las gramáticas tradicionales en este tipo de sistemas están basadas en una etapa anterior de identificación sintáctica. Como vimos, esto no es práctico debido al ruido introducido, con lo cual se usan gramáticas muy ligeras, basadas en *token*; para solventar exclusivamente la detección de dos circunstancias: (a) la aparición de negadores ('estar triste' vs 'no estar triste') y (b) la aparición de conjunciones o separadores de frase.
3. **Diccionario de excepciones**: al carecer el análisis de etiquetado sintáctico, se nos hace muy difícil distinguir entre 'buenos días' (sentimiento neutro) y 'los frutos secos son buenos para la salud cardiovascular' (sentimiento positivo). Es por esto, que se ha implementado un diccionario de excepciones, de tal manera que si se encuentra la palabra 'buenos', pero seguido de 'días', el etiquetado positivo de 'buenos' no se tenga en cuenta
4. **Algoritmo de detección de solapamientos**: basado en Aho-Corasick[12]. Nos permite detectar expresiones que contienen a otras, y por tanto deberían detectarse. Un caso ejemplo puede ser 'pena' (negativo, emoción 'tristeza') vs 'merecer la pena' (positivo)

[12] http://en.wikipedia.org/wiki/Aho%E2%80%93Corasick_string_matching_algorithm (22/09/2013)

Implementación: medición emocional

La implementación final se puede resumir en la siguiente figura:

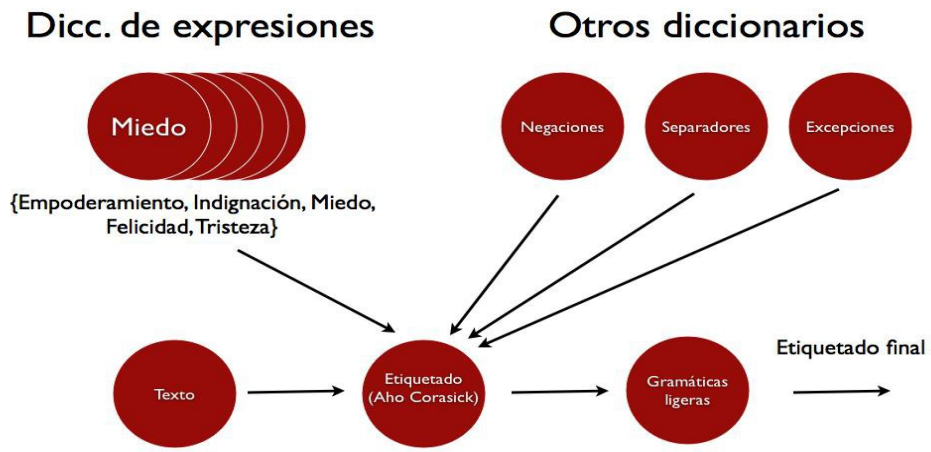


Figura 1: Esquema de la implementación de la detección de emociones en un texto

La secuencia lógica es la siguiente:

1. Se preparan diccionarios con expresiones relativas a cada una de las cinco emociones (Empoderamiento, Indignación, Miedo, Felicidad y Tristeza), en forma raíz, junto con su función sintáctica.
2. Ídem para los negadores, separadores, y excepciones
3. Se pasa un etiquetador sobre el texto, donde se reconocen las entradas de cada uno de estos diccionarios
4. Una gramática ligera combina adecuadamente las etiquetas de los diccionarios, ofreciendo el resultado final: La expresión ‘No estoy para nada triste hoy’, se detecta como ‘negación’ de tristeza.

Implementación: detección de entidades y conceptos

Para la detección de entidades y conceptos en los textos, se ha usado un extractor de entidades basado en las entradas de la Wikipedia en castellano. El proceso se reproduce en la siguiente figura:

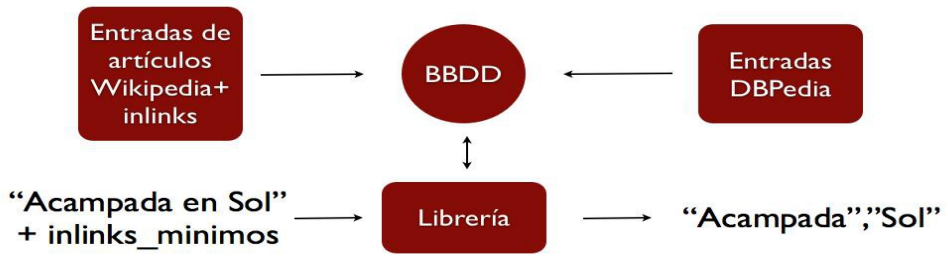


Figura 2: Esquema de la implementación de la detección de entidades basado en Wikipedia

El algoritmo del sistema es el siguiente:

1. Se descargan todas las entradas de la Wikipedia y sus relaciones
2. Con la información de las relaciones entre entradas, se deriva un número: el número de *'inlinks'* o de enlaces entrantes a cada artículo. Éste se usa para definir un número mínimo de '*inlinks'* necesario para identificar una entrada en un texto. De esta manera, eliminamos en gran medida el 'ruido' de entradas poco importantes en el grafo de la Wikipedia
3. Se persiste esta información (lista de artículos junto con su número de *inlinks* en una base de datos)
4. Esta información se complementa en la base de datos con la información de la DBPedia[13] relativa a cada artículo: de esta manera distinguimos entidades (p.ej: 'Mariano Rajoy') de conceptos (p.ej: 'Crisis')
5. Una librería carga esta base de datos en memoria, y detecta en el texto de entrada todos los artículos de la Wikipedia presentes. En el caso de que haya solapamiento (p.ej: 'Puerta del Sol' y 'Sol'), se elige 'el mejor' según una métrica que combina el número de enlaces entrantes y la longitud del literal que ha hecho *match.*
6. Finalmente, la librería devuelve los conceptos/entidades detectados, 'aumentados' con la información de la DBPedia (clasificación de la entidad, categoría, etc.) y el número de enlaces entrantes

En la figura anterior, a modo de ejemplo, se observa como entrada el texto 'acampada en sol', y la extracción de los conceptos 'acampada' y 'sol'

A partir de este sistema, conseguimos detectar las entidades y los conceptos,

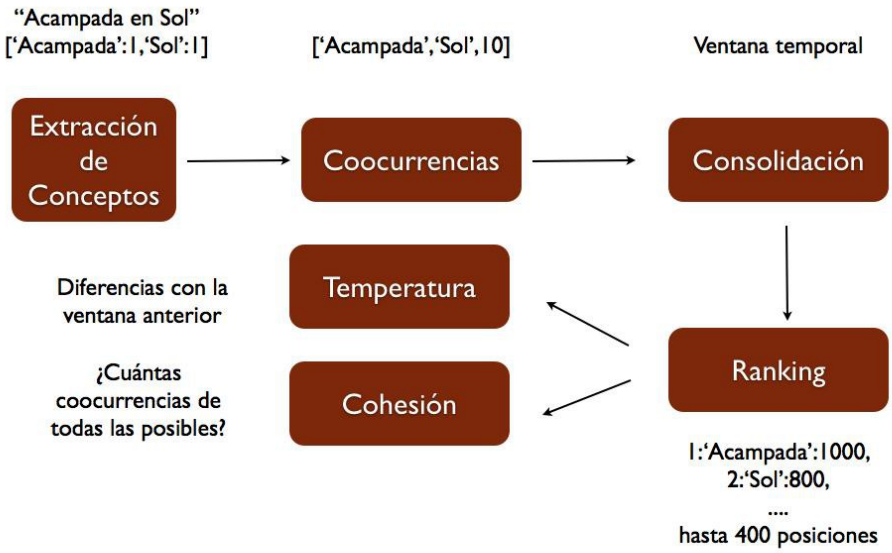
[13] http://es.dbpedia.org/ (22/09/2013)

descansando en un 'instrumento' como la Wikipedia (con sus ventajas: actualización, frescura y cobertura). Pero, ¿cómo detectamos cuando un concepto o entidad se está 'poniendo de moda'? ¿Cómo detectamos, también, la diversidad/unicidad del vocabulario usado en las diversas fases del 15M en las redes?

Para responder a estas preguntas, se hacen dos tipos de análisis adicionales, toda vez que contamos con la lista de conceptos/entidades presentes en un tweet:

1. **Uso de ventanas temporales**: se definen ventanas de tiempo (en el estudio de un día de duración) y, para cada ventana (en este caso, para cada día), identificados sus tweets, se extraen todos sus conceptos y entidades, se consolidan, y se hace un listado con los 400 más frecuentes
2. **Análisis de movimiento entre ventanas temporales:** puesto que para cada ventana contamos con un listado de los conceptos más frecuentes, podemos calcular, para cada concepto/entidad, su diferencia en posición con la ventana anterior. Esta métrica la hemos venido a llamar 'velocidad', e intuitivamente, nos ofrece una medida de la 'temperatura' de un concepto': Si ha pasado a ser más frecuente (velocidad positiva) o menos (velocidad negativa). . De la misma manera, también podemos hablar de la 'aceleración' de un concepto/entidad: si un concepto/entidad cada vez lleva más velocidad
3. **Análisis de coocurrencias**: si cada vez que un concepto aparece junto a otro en un mismo tuit, anotamos un incremento de una unidad en el número de coocurrencias del par de conceptos, en realidad la estructura resultante es una red de conceptos, con los concepto más frecuentemente asociados, más próximos en la red. Por otro lado, podemos analizar la dispersión léxica (si se usan conceptos muy relacionados entre sí, o no) mediante la métrica de red denominada 'densidad de red'[14]

El proceso completo de análisis cuantitativo, incluyendo la extracción de conceptos, el enventanado, el análisis de coocurrencias y cohesión, queda reflejado en la siguiente figura (Figura 3)



[14] http://en.wikipedia.org/wiki/Social_network (22/09/2013)

Figura 3: Esquema del sistema completo de extracción de conceptos y análisis de temperatura y cohesión del vocabulario

Resultados

- **Resultados técnicos**

El análisis emocional en el corpus de test arroja una métrica F1 de 82.4%. Esta métrica es un estándar que combina la precisión y la cobertura. A primera vista, es un resultado bastante alto, pero hay que recordar que estas mediciones dependen del contexto y es muy probable que los diccionarios estén sobreadaptados al corpus utilizado (mensajes relacionados con el 15M).

Se hace necesario un trabajo de campo sobre el conjunto general de Twitter en castellano, y como se hablará en el epígrafe “Trabajos Futuros”, la vía más razonable es liberar el software, los diccionarios y el corpus para mejorar y generalizar la medición via *crowdsourcing.*

- **El 15M en Twitter. Análisis de la carga emocional del lenguaje**

(El resto de este epígrafe está basado en la contribución del autor a [1])

Las principales emociones detectadas en el corpus de tweets son, en orden decreciente: Empoderamiento, Indignación, Miedo y Felicidad.

Se ha introducido la figura *“Carga Emocional”* en el análisis, que hace referencia a la proporción de mensajes originales con componente emocional detectada algorítmicamente respecto al total de mensajes. En la siguiente figura (Figura 4) se puede observar la evolución de esta figura a lo largo del período Abril-Julio de 2011.

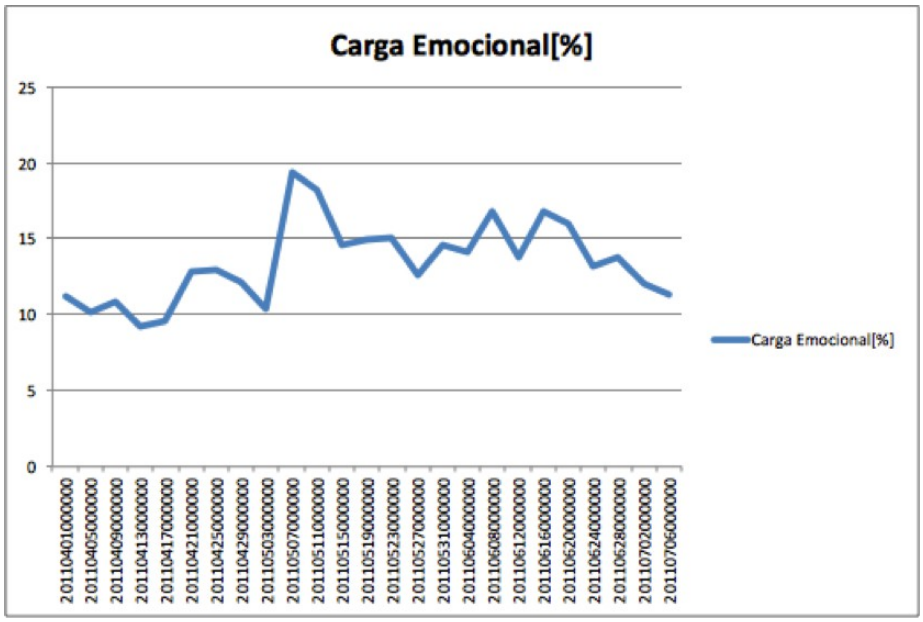


Figura 4: Porcentaje de carga emocional por fechas

A destacar:

1. En media, alrededor del 15% de los mensajes tienen carga emocional. Como referencia a efectos comparativos, un muestreo al 10% de los mensajes capturados durante todo el año 2012 con origen geográfico dentro del territorio Español arroja una carga emocional constante en torno al 7%; (la mitad que en nuestra muestra de estudio), con lo cual **se puede afirmar que el "mensaje 15M" en Twitter tiene una acusada componente emocional.**
2. **La carga emocional se dispara en las primeras semanas de mayo de 2011, alcanzando un pico del 19%**, para después mantenerse en torno 15% hasta finales de junio de 2011

- **"Indignación" y "Empoderamiento/Esperanza" como polos emocionales del 15M**

En la siguiente figura (Figura 5), se observa la evolución a lo largo del período elegido de las emociones predominantes. El eje vertical muestra el volumen asociado a cada emoción (número de tweets originales). Claramente se observa un fuerte pico alrededor de la segunda semana de mayo de 2011 en torno al "empoderamiento", seguido de una meseta que durará hasta finales de mayo.

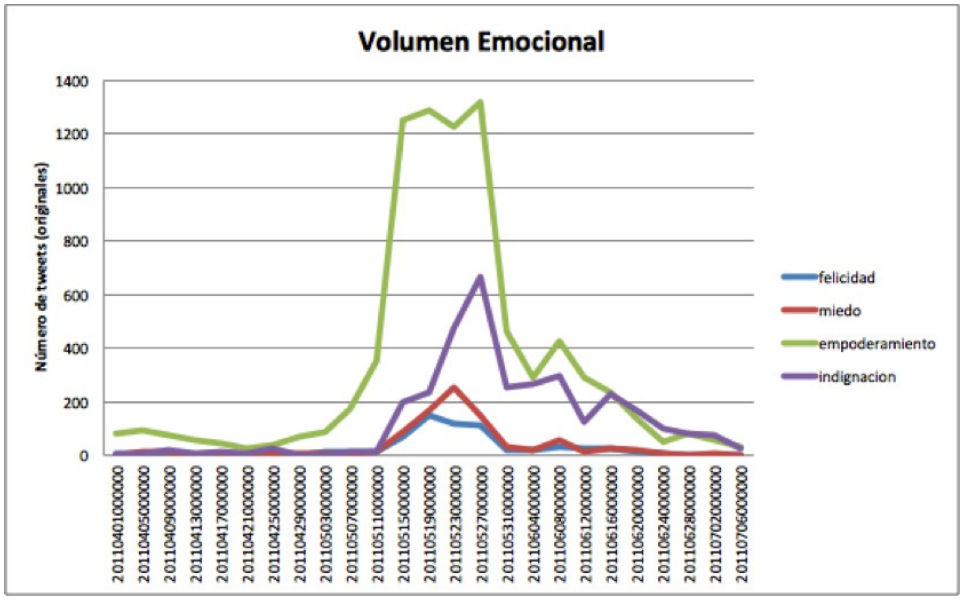


Figura 5: Volumen de tweets asociados a cada emoción por fechas

Dado que el volumen mostrado en la anterior gráfica es absoluto, y no relativo al volumen total de *tweets* originales, **cabe pensar que estas curvas en realidad nos muestran simplemente el incremento de actividad en Twitter en los períodos señalados.**

A continuación se reproduce la misma información (Figura 6), pero con el **eje vertical normalizado al total de *tweets*** en cada punto del tiempo:

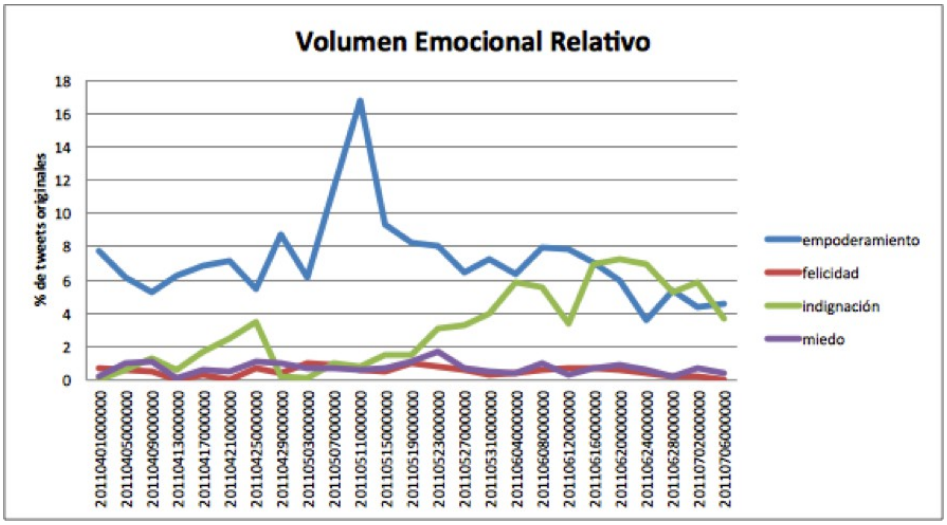


Figura 6: Volumen de tweets asociados a cada emoción por fechas (normalizado al total)

En esta figura se hace todavía más evidente el hecho de que en realidad, **existen dos polos emocionales muy fuertes en el "lenguaje 15m": el "Empoderamiento" y la "Indignación".** Es reseñable, por un lado, que el punto más fuerte de empoderamiento aparece **el día 11 de mayo de 2011: Un 17% de los *tweets* hablaban de empoderamiento (uno de cada 6 mensajes, aproximadamente)** y por otro, que **la indignación entra tímidamente en abril para llegar a ser la emoción predominante en el último mes.**

Respecto a la viralización, los resultados no son concluyentes (no hay indicios estadísticamente significativos de que el contenido emocional aumente la viralización, como se puede observar en el interactivo que se creó como consecuencia de esta investigación[15].

- **La cohesión del lenguaje como síntoma de la sincronización de mensaje**

La métrica de cohesión pretende medir la 'unicidad' o cohesión del mensaje: Mensajes semánticamente muy diferentes (los conceptos expresados en los mensajes son muy dispares) tienden a arrojar magnitudes bajas (en torno a cero), y mensajes muy parecidos (se habla de conceptos muy similares en todos) ofrecen una magnitud cercana a uno.

En el fondo, se trata de ver los conceptos del mensaje (básicamente nombres propios y sustantivos) como una red cuyos enlaces son proporcionales al número de veces que coocurren dichos conceptos en un mismo mensaje. Un conjunto de mensajes cuyo contenido sea exactamente el mismo, daría como resultado una cohesión equivalente a la unidad.

En la siguiente figura (Figura 7) se reproduce la evolución de esta métrica a lo largo del período de estudio en Twitter; y como se observa, **la cohesión es muy alta durante el mes de mayo de 2011, indicando claramente la cohesión del mensaje y la sincronización de mensaje durante este período**.

[15] http://assets.outliers.es/15memociones/ (22/09/2013)

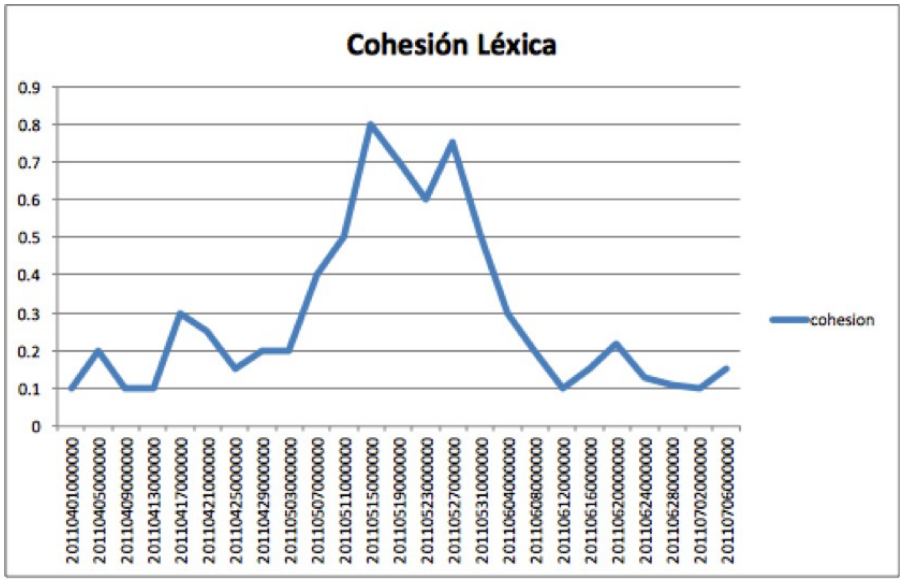


Figura 7: Cohesión léxica por fechas

A continuación se reproduce gráficamente (Figura 8) la 'red' de conceptos usados alrededor del 15 de mayo de 2011, donde se observa directamente la gran cantidad de relaciones en la red (básicamente todos los conceptos están relacionados) que da lugar a una cohesión léxica tan alta (alrededor de 0.7)

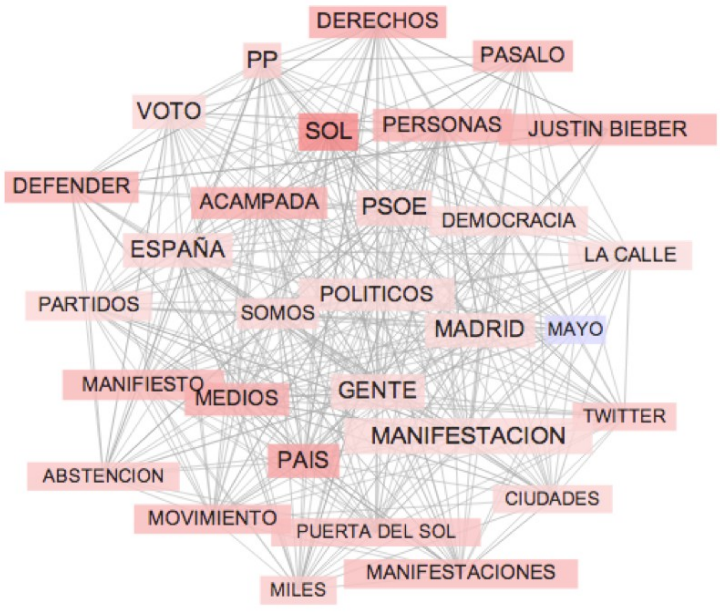


Figura 8: Red de conceptos el 15 de mayo de 2011 en Twitter[16]. Cohesión = 0.7

A efectos de comparación se reproduce (Figura 9) la red léxica a principios de abril de 2011 (Cohesión = 0.1)

[16] http://assets.outliers.es/15mvocabulario/ (22/09/2013)

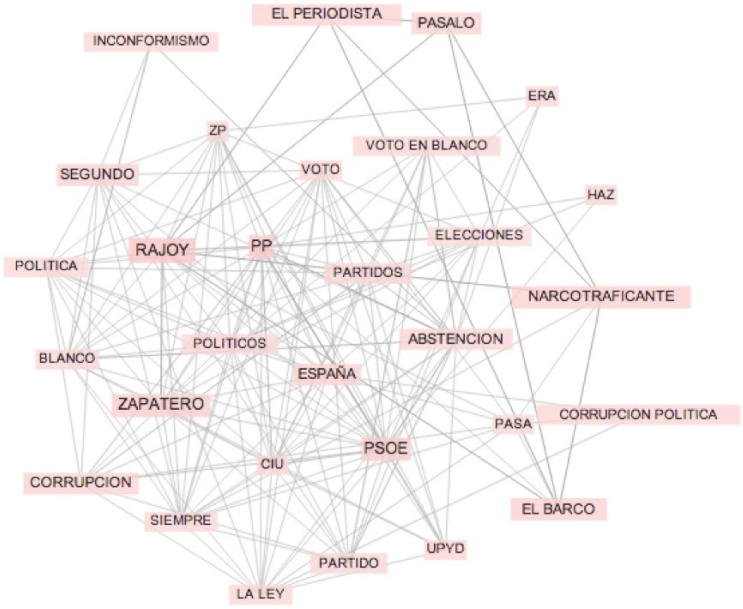


Figura 9: Red de conceptos en la primera semana de abril de 2011 en Twitter. Cohesión = 0.1

- **La temperatura del vocabulario en los días centrales**

La temperatura del vocabulario hace referencia al 'ritmo' con el que se crea un lenguaje. Así, en un período donde circulan memes y mensajes con conceptos muy novedosos respecto al período anterior se obtienen una temperaturas léxicas muy altas. Por otro lado, un período donde los mensajes son exactamente iguales a los aparecidos en el anterior, tendría una temperatura léxica exactamente igual a cero.

En los primeros epígrafes se explica en profundidad cómo se llega algorítmicamente a esta magnitud (que se normaliza a un mínimo de cero y a un máximo de 1.000). Como se observa, **en torno a la segunda semana de mayo de 2011 se produce una verdadera revolución en el "mensaje 15m", generándose lenguaje a mucha velocidad (temperatura cercana a 800) y produciéndose una aceleración característica de los puntos de inflexión en los sistemas realimentados.** En concreto, se puede observar el lenguaje en la red del apartado anterior, que gira en torno a las manifestaciones, acampadas, a la gente ("somos", "personas") y verbos asociados a la acción ("defender", "pásalo"), fenómeno muy acoplado con la aparición del "Empoderamiento" en la medición emocional alrededor del mismo período. Se puede deducir que **la mayor creatividad léxica, por tanto, se concentra en el mes de mayo de 2011**, con una vuelta a un discurso más estático pasado este mes.

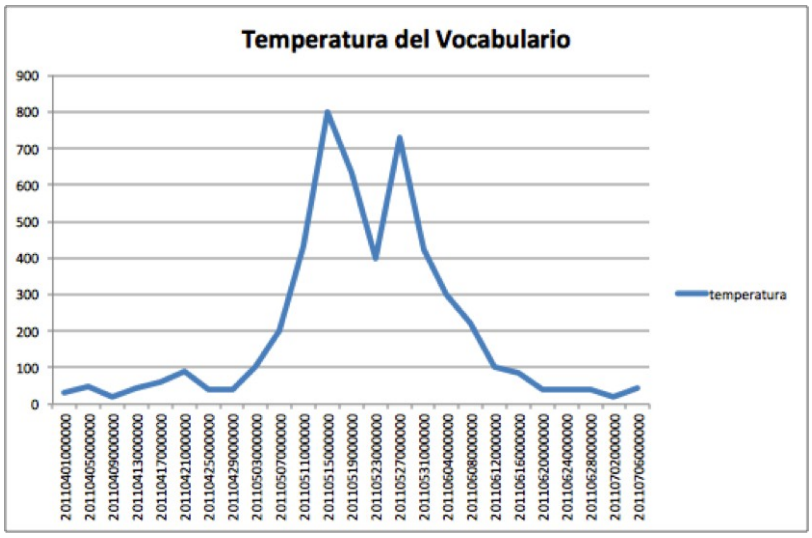


Figura 10: Temperatura del Vocabulario por fechas

Trabajos futuros

Las principales vías de mejora y progreso en este trabajo son las siguientes:

1. Relacionar los conceptos y las entidades con las emociones, para detectar patrones y relaciones entre ambas y responder a la pregunta: "¿Cuáles son los sujetos/objetos de estas emociones?"
2. Visualizar la propagación de las emociones en una red, con el objeto de obtener *insights* de partida para avanzar más en el frente de la relación viralización-contenido emocional.
3. Liberación completa de corpus, software de medición y diccionarios

Referencias

[1] TORET, J et al. (2013). *Tecnopolítica: la potencia de las multitudes conectadas. El sistema red 15M, un nuevo paradigma de la política distribuida (Informe de investigación)*. Barcelona: IN3 Working Paper Series. Universitat Oberta de Catalunya