

IDENTIFICACIÓN AUTOMÁTICA DE LAS NOTICIAS MÁS RELEVANTES EN REDES SOCIALES EN COSTA RICA

AUTOMATIC IDENTIFICATION OF THE MOST SIGNIFICANT NEWS IN SOCIAL MEDIA IN COSTA RICA

*Jairo Andrés Navarro Rodríguez**

*Co-autor: Édgar Casasola Murillo***

RESUMEN

Este artículo muestra el resultado de la implementación de un monitor que busca las noticias más relevantes en tres perfiles de Facebook de medios de comunicación costarricenses. Se crea un modelo de datos capaz de almacenar la información consultada en cada uno de los perfiles, como publicaciones, comentarios, y «me gusta». Con la información recolectada se define una estrategia para la clasificación de las noticias, estableciendo para cada una de ellas un valor de relevancia. Los resultados son mostrados en tiempo real mediante una página web, clasificando las noticias de mayor a menor relevancia y que, a su vez, pueda llevarlo hasta la noticia seleccionada. Se realiza una comparación manual de los resultados presentados por el sitio web y los datos presentados por cada uno de los perfiles, con el fin de evaluar que tan buena o mala fue la estrategia de clasificación. Finalmente, basándose en los resultados se plantea una discusión de las mejoras que se podrían implementar a futuro en la estrategia de clasificación.

Palabras Clave: monitor, relevancia, estrategia, perfiles, noticias.

ABSTRACT

This article shows the result of the implementation of a monitor that looks for the most relevant news in three Facebook profiles from media in Costa Rica. A data model capable of storing information consulted on each of the profiles, such as posts, comments, and “likes” is created. Therefore, a strategy will be establishing a relevant value for each classification with all the collected information. The results are displayed in real time through a web page, *ranking* from highest to lowest relevance of each selected news. A manual comparison of the results will be presented by the website and the data presented by each of the profiles, in order to evaluate how good or bad was the strategy of classification. Finally, based on the results, it is proposed a discussion of improvements that could be implemented in the future in the strategy of classification.

Key Words: monitor, relevance, strategy, profiles, news.

* Universidad de Costa Rica. Correo electrónico: jandresnr@gmail.com

** Universidad de Costa Rica. Correo electrónico: edgar.casasola@ucr.ac.cr

Recepción: 15/1/2016. Aceptación: 16/3/2016

1. Introducción

En la actualidad se cuenta con gran cantidad de páginas de noticias en línea, las cuales han ido sustituyendo poco a poco los medios impresos. Este tipo de páginas tienen la ventaja de publicar noticias constantemente y en cualquier momento del día. Este tipo de publicaciones elimina la necesidad de tener que esperar hasta el día siguiente para divulgar un acontecimiento de última hora, tal y como sucede con los medios impresos, o para comunicar una noticia mediante la radio o la televisión.

Una noticia puede ser relevante para una persona y no para otra. En este caso, debemos analizar diferentes variables que se deben tomar en cuenta como: el tema de la noticia, el tipo de noticia, la época en la que se genera la noticia (si hay algún evento en particular a nivel nacional o mundial que se dé en el momento de la publicación) y los gustos de la persona en algún tema en particular. Todas estas variables son importantes considerarlas al momento de decidir qué tan relevante puede ser o no una noticia publicada.

Actualmente no se cuenta con una herramienta que permita identificar cuáles son las noticias publicadas en línea que están de moda, considerando variables como las mencionadas anteriormente, y que tomen en cuenta otros aspectos como la cantidad de comentarios en un lapso de tiempo determinado o la cantidad de "me gusta" con los que cuenta la publicación.

Para crear una herramienta con la capacidad de detectar las noticias más relevantes y que están generando mayor interés en las personas, resulta necesario establecer una estrategia que permita identificar cuál noticia es más importante que otra, para la creación de esta estrategia nos basamos en el modelo PageRank para determinar la relevancia de un documento (Baeza y Ribeiro, 2008).

Con base en esto, nace la necesidad de implementar un monitor automático en tiempo real de las noticias en Costa Rica, que contemple una estrategia de clasificación de noticias que vaya de la más relevante a la menos relevante.

En las secciones siguientes, presentamos el análisis realizado sobre tres perfiles de noticias

en línea de la red social Facebook, para extraer elementos relacionados sobre estas. Se muestra un modelo de datos en donde se almacena la información obtenida, la cual sirve de insumo para generar una estrategia de clasificación. Además, se explica la estrategia de clasificación definida, la cual solo toma en cuenta elementos o variables cuantitativas. Asimismo, se exponen los resultados obtenidos después de la implementación del monitor y, por último, se concluye sobre el trabajo realizado.

2. Metodología

Para llevar a cabo la investigación, se utilizaron como base tres de los perfiles más populares de la red social Facebook en Costa Rica (Arce, 2012). De cada uno de los perfiles se extrajeron los comentarios, «me gusta» y atributos para cada una de las noticias publicadas.

Para la extracción de estos datos se utilizó una librería gratuita llamada RestFB (Allen, 2010-2016). Esta librería se integra con el lenguaje de programación Java, la cual se utilizó para el desarrollo del monitor. Además, se empleó el API (Application Programming Interface) Graph API de Facebook. Este API permite la ejecución de consultas (Jesse y Tarjan, 2015), las cuales son enviadas desde RestFB para la extracción de la información de los perfiles (publicaciones, comentarios, autor comentario, «me gusta», autor, fecha).

Después de obtener los elementos de cada una de las publicaciones de los perfiles, se creó el modelo de datos. Este modelo de datos se usó para definir la forma en la cual se almacenó la información. Posteriormente, con la información almacenada en un repositorio de datos, se analizó y definió una estrategia para clasificar las noticias, en función de la más relevante a la menos importante. Esta clasificación se llevó a cabo utilizando únicamente datos cuantitativos.

Una vez definida la estrategia de clasificación, se implementó un monitor automático en tiempo real que empleó esta estrategia y que sirvió como insumo para mostrar los resultados a través de una página web con las noticias clasificadas según el nivel de importancia. Concluido

el desarrollo del monitor, se realizó una prueba en la cual se compararon manualmente los resultados obtenidos por el monitor con lo publicado en cada uno de los perfiles.

3. Implementación del Monitor

La implementación del monitor para la identificación de las noticias más relevantes se divide en cinco fases. La primera fase de análisis consiste en la extracción de noticias y sus elementos. La segunda fase, la generación del modelo de datos para almacenar la información. La tercera fase, la creación de la estrategia de clasificación. La cuarta fase, la creación del monitor; y la última fase, la creación de un sitio web para mostrar los resultados.

3.1. Análisis y extracción de noticias y sus elementos

Para la extracción de las noticias publicadas y sus elementos (cantidad de comentarios, cantidad de «me gusta», comentarios, autor de comentario) se seleccionaron tres perfiles de

noticias visitados normalmente por los costarricenses. Los tres perfiles consultados son: crhoy.com noticias (35 publicaciones diarias promedio), Telenoticias 7 (70 publicaciones diarias promedio) y nacion.com (60 publicaciones diarias promedio) (Arce, 2012).

Una vez identificados los tres perfiles, se crearon las diferentes consultas a utilizar para extraer la información requerida. Por medio de Graph API, se valida cada una las consultas que serán ejecutadas cada media hora por el monitor.

Para recolectar la información necesaria de cada perfil se necesitaron tres consultas diferentes. La primera, para extraer las noticias nuevas publicadas; la segunda, para obtener y contabilizar la cantidad de comentarios; y la tercera, para contabilizar la cantidad de «me gusta».

3.1.1. Consulta para extraer las noticias

Para la extracción de las nuevas noticias, se utilizó la consulta **265769886798719/posts**. La cifra numérica representa la identificación del perfil seguido del comando «posts», el cual obtiene el ID de las noticias publicadas.

```

{"data": [
  {
    "actions": [
      {
        "name": "Comment",
        "link":
"https://www.facebook.com/265769886798719/posts/1096608880381478"
      },
    ]
    "actions": [
      {
        "name": "Comment",
        "link":
"https://www.facebook.com/265769886798719/posts/1096606900381676"
      },
    ]
  },
]
    
```

FIGURA 1

Resultado de la consulta de las publicaciones de un perfil.

Cada una de las URL en azul representa una noticia, la cual se identifica por la última serie

de números del enlace. La primera ruta representa la noticia que se muestra a continuación:



FIGURA 2

Ejemplo de publicación devuelta con la consulta de publicaciones de un perfil.

3.1.2. Consulta para extraer los comentarios

Una vez obtenidas las noticias de los perfiles, se realiza la consulta para extraer cada uno de los comentarios y el total de estos. Se utilizó la consulta **1096608880381478/comments?summary=true**.

La cifra numérica representa la identificación de la noticia, seguido del comando «comments» para obtener los comentarios y, por último; se le indica que muestre un resumen. La siguiente figura muestra un resumen que contiene la cantidad total de comentarios.

```
{
  "data": [
    {
      "created_time": "2016-01-28T06:01:14+0000",
      "from": {
        "name": "Sha Gomez Brenes",
        "id": "10153393986175101"
      },
      "message": "Meli Flores desde cuando cambiaron los gansos por un perrito?"
    },
    {
      "created_time": "2016-01-28T06:27:58+0000",
      "from": {
        "name": "Espinoza Tutti",
        "id": "888710391155057"
      },
      "message": "Me alegro muchísimo tanto por ellas, como por el perrito. Ojalá todos tuvieran un hogar."
    },
    "summary": {
      "order": "ranked",
      "total_count": 6
    }
  ]
}
```

FIGURA 3

Resultado de la consulta de los comentarios de una publicación.

La consulta devuelve la fecha, la hora de creación, el mensaje y el autor. Al final de los mensajes devueltos por la consulta, retorna la cantidad total de mensajes que tiene la noticia. Cada uno de los mensajes, así como la cantidad total, son guardados para el análisis posterior de la estrategia de clasificación.

3.1.3. Consulta para extraer los “me gusta”

De la misma forma como se obtuvieron los comentarios de cada noticia, así se consigue los «me gusta» de las estas. Para obtener la cantidad total de «me gusta», se utilizó la siguiente consulta `1096608880381478/likes?summary=true`. La cifra numérica representa la identificación de la noticia, seguido del comando «likes» para obtener los usuarios que han dado «me gusta» a la noticia, y por último; se le indica que muestre un resumen. La siguiente figura muestra un resumen que contiene la cantidad total de «me gusta».

```
{
  "data": [
    {
      "id": "707699109271152",
      "name": "Alonzo Monge Quesada"
    },
    {
      "id": "10202652614255791",
      "name": "Ricardo Barquero"
    },
    {
      "id": "10152100243907469",
      "name": "Geo B. Torres"
    }
  ],
  "summary": {
    "total count": 64
  }
}
```

FIGURA 4

Resultado de la consulta de los “me gusta” de una publicación.

La consulta devuelve un identificador y el nombre del usuario que dio «me gusta» a la noticia. De la información, devuelta sólo es almacenada la cantidad total de «me gusta».

3.2. Modelo de Datos

Establecida la forma en que se extraerán los datos, se procede con la creación del modelo de datos para almacenar la información. De manera que se establece un modelo de tres tablas. En la primera tabla, se almacena la identificación de la noticia, el nombre de la noticia, la fecha de consulta, la cantidad de «me gusta», la cantidad de comentarios y la cantidad de veces que se ha consultado la información de la noticia. Esta información es almacenada periódicamente cada treinta minutos. Ni la cantidad de comentarios ni la de «me gusta» son almacenados de forma acumulativa.

Cuando se ingresa una noticia por primera vez, se almacenara la cantidad de comentarios y «me gusta» que tenga hasta el momento. La siguiente ocasión en que se registre la misma noticia solo se almacenará la diferencia de comentarios y «me gusta» conforme al registro anterior.

La segunda tabla almacena los comentarios de cada noticia. Al igual que en la primera tabla, cada vez que se consulta por la misma noticia solo se agregan los comentarios que han ingresado después de la última consulta realizada.

La tercera tabla se usa para guardar el *ranking* de las noticias, de la más relevante a la menos importante, según el criterio de clasificación. Se almacena el ID de la publicación, el peso de comentarios, el peso de «me gusta», el peso de tiempo, el peso de variables, la relevancia principal, la relevancia secundaria y la relevancia absoluta.

3.3. Creación de la estrategia de clasificación

Para establecer la estrategia de clasificación, se definieron tres variables por analizar:

- Cantidad de comentarios
- Cantidad de «me gusta»
- Tiempo de haber sido publicada la noticia

Estas variables fueron analizadas cada una por separado, con el fin de establecer para cada una de ellas un peso, según la cantidad de datos obtenidos por corrida del monitor.

Además, de los pesos de cada variable, se establecieron dos reglas para determinar las noticias de cada uno de los tres perfiles, las cuales van a ser tomadas en cuenta para el *ranking* de noticias relevantes. Estas reglas van a delimitar la cantidad de noticias evaluadas con cada corrida del monitor. Las reglas que se definieron son las siguientes:

- Se evaluarán solo las últimas 100 noticias de cada perfil. Es decir, como máximo se van analizar 300 noticias por corrida del monitor.
- De las 300 noticias por analizar, solo se van a tomar en cuenta aquellas con comentarios o un «me gusta» en las últimas 48 horas.

Una vez definidos los parámetros para evaluar las noticias, se establece la fórmula para clasificarlas por relevancia. La fórmula utilizada para generar este *ranking* se define de la siguiente forma:

$$R1 = WC + WL + WT$$

En donde:

- R1:** Relevancia principal.
- WC:** Peso de comentarios.
- WL:** Peso de *like*.
- WT:** Peso de tiempo.

Se establece como relevancia de una noticia la suma del peso de los comentarios, más el peso de «me gusta», más el peso de tiempo. El resultado final puede ser igual entre noticias, ya sea del mismo perfil o no. Para tomar la decisión de cuál noticia es más importante respecto a otra donde su relevancia dada por la fórmula anterior es igual, se utiliza una segunda fórmula de desempate.

Para desempatar noticias con el mismo valor de relevancia, se toma en cuenta un nuevo peso dado sobre las tres variables de análisis

(comentarios, «me gusta», tiempo). La fórmula utilizada para desempatar la relevancia es definida de la siguiente forma:

$$R2 = (R1, PL + PC + PT)$$

En donde:**R2:** Relevancia secundaria

- R1:** Relevancia principal
- PL:** Posición *like* (la posición va de 1 a «n», en donde «n» representa el número total de R1 que están desempatando)
- PC:** Posición comentarios (la posición va de 1 a «n», en donde «n» representa el número total de R1 que están desempatando)
- PT:** Posición tiempo (la posición va de 1 a «n», en donde «n» representa el número total de R1 que están desempatando)

Se establece como relevancia secundaria el par (relevancia principal, posición *like* + posición comentarios + posición tiempo), y en donde la relevancia principal está dada por la fórmula R1.

Una vez calculada la relevancia de todas las noticias analizadas, se indexan los resultados en la base de datos según su orden de relevancia de mayor a menor.

3.3.1. *Peso de clasificación para los comentarios*

Para definir el peso que se debe asignar a los comentarios de una publicación, se analizaron individualmente las noticias publicadas de cada perfil, durante un período de cinco días. Este análisis arrojó como resultado que las noticias con más comentarios tenían entre 150 y 200 comentarios.

Con respecto a los resultados, se crea una escala de pesos de 0 a 10, en donde 0 representa una noticia sin comentario y 10 representa una noticia con 199 o más comentarios. Con base en esto, se establece la siguiente tabla de pesos en función a la cantidad de comentarios:

TABLA 1

Pesos de los comentarios para una noticia según la cantidad

Cantidad de comentarios	Peso
0 comentarios	0
1 a 22 comentarios	1
23 a 44 comentarios	2
45 a 66 comentarios	3
67 a 88 comentarios	4
89 a 110 comentarios	5
111 a 132 comentarios	6
133 a 154 comentarios	7
155 a 176 comentarios	8
177 a 198 comentarios	9
199 o más comentarios	10

El peso asignado a la noticia es el valor **WC** de la formular **R1** (relevancia principal).

3.3.2. *Peso de clasificación para los «me gusta»*

Para definir el peso que se debe asignar a los «me gusta» de una publicación, se analizaron individualmente las noticias publicadas de cada perfil, durante un periodo de cinco días. Este análisis reveló que las noticias con más comentarios tenían entre 350 y 400 «me gusta».

De acuerdo con los resultados, se dispone una escala de pesos de 0 a 10, en donde 0 representa una noticia sin «me gusta» y 10 representa una noticia con 398 o más «me gusta». De esta manera, se observa la cantidad de «me gusta» en la siguiente tabla de pesos:

TABLA 2

Pesos de los “me gusta” para una noticia según la cantidad

Cantidad de “me gusta”	Peso
0 «me gusta»	0
1 a 44 «me gusta»	1
45 a 88 «me gusta»	2
89 a 132 «me gusta»	3
133 a 176 «me gusta»	4
177 a 220 «me gusta»	5
221 a 264 «me gusta»	6
265 a 308 «me gusta»	7
309 a 352 «me gusta»	8
353 a 396 «me gusta»	9
398 o más «me gusta»	10

El peso asignado a la noticia es el valor **WL** de la formular **R1** (relevancia principal).

3.3.3. *Peso de clasificación por tiempo*

Para definir el peso de tiempo para una noticia, se debe registrar el número de consultas que se han hecho sobre esta, cada vez que se ejecuta el monitor. El tiempo definido para la ejecución del monitor es de cada treinta minutos. La noticia tendrá registrada en un día, después de su ingreso, un total de 48 consultas; es decir, la primera vez que se ingresa una noticia tendrá un 1 en el número de consultas, la segunda vez un 2 y así sucesivamente, cada vez que se ejecute el monitor.

Para la generación del *ranking* se toman en cuenta solo las 100 noticias más recientes del perfil, esto implica que una noticia puede ser evaluada dentro de las primeras 48 horas de

ingreso (96 entradas de ejecución). Para generar la escala, se toma como valor máximo un 96 (cantidad total de ejecuciones en los primeros dos días).

A diferencia de las escalas de peso para comentarios y «me gusta», se asigna un mayor peso si el tiempo de ingreso de la noticia es menor. La escala de peso definida para el tiempo es la siguiente:

TABLA 3

Pesos de las consultas realizadas de una noticia según la cantidad

Cantidad de Ejecuciones	Peso
1 a 12 ejecuciones	9
13 a 24 ejecuciones	8
25 a 36 ejecuciones	7
37 a 48 ejecuciones	6
49 a 60 ejecuciones	5
61 a 72 ejecuciones	4
73 a 84 ejecuciones	3
85 a 96 ejecuciones	2
96 o más ejecuciones	1

El peso asignado a la noticia es el valor **WT** de la formular **R1** (relevancia principal). Este peso disminuye en 1 cada 6 horas (2 ejecuciones = 1 hora).

3.3.4. *Peso de clasificación por variables*

Esta clasificación de pesos se utiliza cuando más de una noticia tiene la misma relevancia principal. Para asignar un segundo valor de importancia entre las noticias con igual relevancia principal, se genera un consecutivo de 1 a «n», en donde «n» es la cantidad de noticias a desigualar. Para las variables de comentarios y «me gusta», se pone la posición según el valor de la variable, estableciendo la posición «n» a la noticia con mayor valor, la posición n-1 a la noticia con el segundo valor más alto, así sucesivamente hasta llegar a 1.

Para la variable de tiempo también se asigna la posición según el valor, pero en este caso

se asigna la posición «n» a la noticia con menor valor, la posición n-1 a la noticia del segundo menor valor, así hasta llegar a 1.

En caso de que después de calcular la relevancia secundaria, más de dos noticias tengan la misma relevancia secundaria se asignará un peso a la variable de la siguiente forma:

TABLA 4

Pesos de las variables de clasificación

Variable	Peso
«me gusta»	3
Comentario	2
Tiempo	1

3.4. **Creación de monitor y página de resultados**

Para la creación del monitor, se utiliza el lenguaje de programación de Java, el cual se comunica con Graph API mediante la librería RestFB. A través de esta comunicación, se consulta la información de los tres perfiles para obtener las publicaciones y los elementos relacionados a estos.

Toda la información consultada es almacenada en el repositorio de datos. Esta información se consulta cada treinta minutos. Una vez almacenada toda la información, se realiza el proceso para determinar la relevancia. En este punto se aplica la estrategia de clasificación y sus resultados son almacenados de nuevo en el repositorio de datos.

Para mostrar los resultados de clasificación de relevancia, se creó un sitio web donde se muestran las noticias clasificadas de mayor a menor relevancia. Debido a que cada ejecución del monitor registra un *ranking* nuevo con la clasificación de las noticias, se pueden hacer consultas de las últimas ejecuciones del monitor. Esto permite mostrar al usuario las noticias actuales más relevantes y las noticias que fueron más relevantes en ciclos anteriores de ejecución.

3.5. Evaluación de resultados

Para la evaluación de resultados se tomó la información generada por el monitor en dos días diferentes. Se hizo una selección al azar de tres horarios (2:00 a.m., 10:00 a.m., 5:00 p.m.) y se evaluaron los primeros diez resultados generados por el monitor. En total se hicieron seis evaluaciones.

3.5.1. Resultados Horario 2:00 a.m.

Al realizar la evaluación manual, en los resultados, para ambos días en este horario, se detectó que el monitor puso noticias que tenían entre doce y dieciocho horas de haber sido publicadas. Este comportamiento se debe a que son noticias que tenían una importante cantidad de «me gusta» y comentarios, y de ahí que el peso de la variable de tiempo no fue tan influyente.

También, se observó que los comentarios y la cantidad de «me gusta», al hacer la consulta de las publicaciones en los perfiles, eran los mismos al momento de aplicar la estrategia de clasificación.

3.5.2. Resultados Horario 10:00 a.m.

Los resultados para los dos días analizados en este horario incluyó noticias que tenían de dos a treinta y seis horas de haber sido publicadas. Esta amplitud de horario respecto al horario de las 2:00 a.m. se debe a que las publicaciones de noticias se dan en menor cantidad durante la noche. Además, se logró detectar que de los tres perfiles, el de Telenoticias 7 es el que mantiene una publicación más constante de noticias y, en el caso de los otros dos perfiles, las publicaciones de noticias durante la noche disminuyen considerablemente respecto a las publicaciones durante el día.

Otro punto interesante es que de las noticias registradas como las diez más relevantes, y que tenían apenas dos horas de publicación, ingresaron debido a que hay un incremento en las noticias publicadas a partir de las 6:00 a.m.. Además del incremento en publicaciones de noticias nuevas, también hay un incremento de las personas quienes comentan y dan «me gusta» a las noticias a partir de las 8:00 a.m.

Respecto al análisis individual de cada noticia, se observó que la cantidad de «me gusta» y de los comentarios cambia en las noticias más recientes, en relación con los utilizados durante el cálculo de relevancia. Esto puede incidir en la cantidad de noticias que se muestran como las diez más relevantes y que fueron publicadas hace más de veinticuatro horas.

3.5.3. Resultados Horario 5:00 a.m.

Los resultados en este horario nos permitieron tener como noticias más relevantes las diez que tenían de dos a doce horas de haber sido publicadas. En comparación con los otros dos horarios, se observa que se tienen noticias más recientes. Este hecho se debe a que la cantidad de noticias publicadas entre las 6:00 a.m. y las 6:00 p.m. es mayor y, además, es cuando hay más personas comentando y dando «me gusta» a las publicaciones.

En este horario, al igual que el de las 10:00 a.m., se observó el mismo comportamiento en el cual, tanto la cantidad de «me gusta» como la de comentarios de las diez noticias más relevantes, se incrementaban rápidamente. Debido a este comportamiento y el correr el monitor cada media hora, influye bastante en los resultados obtenidos, pues una noticia que está recibiendo mucha actividad por parte de los lectores, es mejor clasificada hasta media hora después.

Otro punto importante de observar es que en las diez noticias más relevantes, principalmente para las ejecuciones analizadas en el día, no se observan noticias que tengan menos de dos horas de publicadas y que puedan ser relevantes.

4. Conclusiones

Se logró recopilar toda la información necesaria para la definición de una estrategia de clasificación de relevancia. A pesar de que no se utilizó toda la información obtenida de los perfiles, al estar almacenada, nos permite en un futuro tomar en cuenta estos datos para crear otro método que mejore los resultados obtenidos por la estrategia de clasificación actual.

Además, se identificó una estrategia de clasificación de relevancia utilizando únicamente variables cuantitativas. Aunque los resultados obtenidos por el monitor son aceptables con base al método seleccionado, esto se puede mejorar; con el fin de que se tomen en cuenta noticias que hayan sido publicadas más recientemente.

También, los resultados evidenciaron que las noticias con menos de dos horas de publicación no son consideradas como relevantes, debido a que el peso dado a la variable de tiempo no es tan influyente con respecto al peso dado a las variables de cantidad de comentario y cantidad de «me gusta». Esto plantea la necesidad de dar un mayor peso al tiempo, con el fin de, que se detecte con más facilidad si una noticia puede llegar a ser potencialmente de mucha relevancia; aún cuando la cantidad de comentarios y «me gusta» no sea bastante buena.

En cuanto a la estrategia, esta se puede mejorar para obtener las noticias más relevantes tomando en cuenta aspectos como: el tipo de noticia, los gustos de las personas, los temas y el análisis de comentarios, los cuales no fueron considerados. Sin embargo, el monitor implementado abre espacio para recopilar información muy valiosa, la cual está segregada en varios perfiles de la red social y que, por separado, es muy difícil de analizar.

La implementación del monitor permitió la recopilación de información muy valiosa que está segregada en varios perfiles de la red social Facebook y que por separado es muy difícil de analizar. Al tener información de estos perfiles de los medios de comunicación más populares de Costa Rica recopilados en forma conjunta, da

la posibilidad de hacer estudios muy interesantes como análisis de mercado de productos específicos, gustos y preferencias de las personas, análisis de índole social, políticos y culturales, todo esto mediante el análisis de la opinión pública de las personas.

5. Bibliografía

Allen, M. y N. Bartels. 2010-2016. RestFB, RestFacebook. Recuperado de: <http://restfb.com/>. Consulta: 29 de diciembre 2015.

Arce, José Luis. 2012 “Medios de comunicación de masas en Costa Rica: Entre la digitalización, la convergencia y el auge de los “new media””. En: *Hacia la Sociedad de la Información y el Conocimiento, capítulo Medios de Comunicación de Masas en Costa Rica*: 283-308.

Baeza-Yates R y B. Ribeiro-Neto. 2008 *Recuperación de la Información Moderna: Conceptos y Tecnología detrás de la búsqueda*. Ciudad de México, Editorial Addison Wesley.

Weaver, Jesse y Paul Tarjan. 2015. Facebook datos ligados a través de la API de Graficos. Web Semántica. Recuperado de: <https://developers.facebook.com/docs/graph-api>. Consulta: 29 de diciembre 2015.

