



TESIS DE GRADO
EN INGENIERÍA INDUSTRIAL

MINERÍA DE OPINIÓN
PARA LA VOZ DEL CONSUMIDOR

Autor: Juan Pablo Stocca
Legajo: 45219

Directora de Tesis: Prof. María del Carmen Galindez

2010

A mis abuelos, quienes me dieron la posibilidad de estudiar en el ITBA.

Quería agradecer especialmente a todas las personas que colaboraron conmigo para hacer este trabajo:

A mi tutora, la profesora María del Carmen Galindez, por su buena voluntad y predisposición para ayudar.

A mis compañeros, Guillermo Stockdale y Martín Ribotta, por su acompañamiento y apoyo.

A mis colegas, Gonzalo Suárez, Davi Pioli y Jenny Suvada, por el apoyo y el aporte de sus conocimientos técnicos.

A mis hermanas y padre, Mariana Stocca, Carolina Stocca y Davor Stocca, por su colaboración para mejorar la calidad del trabajo.

RESUMEN

Este trabajo pretende capturar la voz del consumidor dentro del mercado de celulares “smartphone” de Estados Unidos. Su objeto de estudio es el Droid Incredible, un producto de HTC. Por otro lado se comparará el rendimiento de este producto con algunos de sus principales competidores: iPhone 4 de Apple, Droid X de Motorola y Bold 9700 de BlackBerry.

La innovación que presenta esta tesis es el uso de la Minería de Opinión para lograr los resultados. La Minería de Opinión es una aplicación específica de la Minería de Texto, un proceso que permite obtener información de alto valor a partir de texto desestructurado basado en el lenguaje natural. La información de texto que se usará como entrada al análisis será información pública de la Web, extraída a través de la técnica de Web Scraping.

Con el uso de estas herramientas se encontrarán los principales problemas que el consumidor enfrenta cuando usa el producto, siendo los más importantes la poca vida útil de la batería y una sistema de comunicación errático. Por otro lado se hará un análisis especial sobre la velocidad de respuesta del teléfono, un atributo valorado en la competencia y con no tan buenos resultados en el Droid Incredible.

En el trabajo se explicará como el estudio de la voz del consumidor es una etapa clave en el mejoramiento de la calidad de los productos. La información proporcionada permitirá a la organización tomar acciones para corregir sus problemas, pues el método planteado corresponde a un proceso real de mejoramiento de la calidad.

Palabras clave: *Voz del consumidor, Calidad, Minería de Opinión, Minería de Texto, Web Scraping, Smartphone.*

ABSTRACT

This work intends to capture the voice of the consumer within the US smartphone market. The object of this study is the Droid Incredible, a product manufactured by HTC. On the other hand, the analysis will compare the performance of this product with some of its main competitors: iPhone 4 by Apple, Droid X by Motorola and Bold 9700 by BlackBerry.

The innovation that this thesis presents is the use of sentiment analysis to achieve the results. Sentiment analysis is a specific application of Text Mining, a process that allows obtaining high quality information from unstructured text data. The text information that will be used as an input to this analysis will be extracted from the public domain of the Web, through a Web Scraping technique.

With the use of these tools the main issues that the consumer faces will be found, from which the most important are the short battery life and an erratic communication system. On the other hand there will be a special analysis on phone responsiveness, an attribute valued in the competition but with poorer results for the Droid Incredible.

This work will explain how the voice of the consumer is a key phase of product quality enhancement. Since the proposed method is part of a real process, the information provided will allow the organization to take actions to correct their issues.

Key Words: *Voice of consumer, Quality, Sentiment analysis, Text Mining, Web Scraping, Smartphone*

ÍNDICE

1. INTRODUCCIÓN	1
1.1. LA VOZ DEL CONSUMIDOR	1
1.2. WEB SCRAPING	4
1.3. MINERÍA DE OPINIÓN	6
1.4. VISIÓN GENERAL DEL PROYECTO.....	10
1.5. ALCANCE DEL PROYECTO.....	11
2. RECOLECCIÓN DE LA INFORMACIÓN	15
2.1. OBSERVACIÓN DE LA WEB	15
2.2. EXTRACCIÓN DE LA INFORMACIÓN.....	19
2.3. CONSIDERACIONES SOBRE LA INFORMACIÓN EXTRAIDA.....	24
3. PROCESO DE LA MINERÍA DE TEXTO	27
3.1. VALORACIÓN DEL PRODUCTO	27
3.2. DESARROLLO DE LA MINERÍA DE TEXTO	31
3.3. PRODUCTIVIDAD Y MEJORA CONTINUA	38
4. ANÁLISIS DE LA INFORMACIÓN	41
4.1. VOZ DEL CONSUMIDOR SOBRE EL PRODUCTO	41
4.2. PROFUNDIZACIÓN DE UNA PROBLEMÁTICA	52
5. COMENTARIOS FINALES.....	59
5.1. UTILIDAD DE LA INFORMACIÓN PARA LA ORGANIZACIÓN	59
5.2. FACTIBILIDAD DE LA MINERÍA DE TEXTO	60
5.3. CONCLUSIONES	63
6. ANEXOS.....	65

7. BIBLIOGRAFÍA 67

1. INTRODUCCIÓN

El presente capítulo introducirá los temas a tratar durante el proyecto. Se definirá la minería de texto (a partir de ahora MT) y las distintas aplicaciones que ésta tiene, pero principalmente su relación con la minería de opinión (ahora MO) y el estudio de la voz del consumidor. También se definirá la técnica de Web Scraping, que será utilizada en el proyecto como input al proceso de MT. Al final del capítulo se especificarán las limitaciones y el alcance del proyecto. El objetivo final será obtener la voz del consumidor dentro del mercado de dispositivos móviles. Se comenzará presentando este tema en la siguiente sección.

1.1. LA VOZ DEL CONSUMIDOR

1.1.1. Definición

La voz del consumidor representa las expectativas, preferencias y aversiones del mismo. Dentro del contexto de la organización, la voz del consumidor es una técnica de investigación de mercado, a través de ella se definen los deseos y necesidades del cliente, los cuales son organizados en una estructura jerárquica. Ésta técnica involucra tanto estudios cualitativos como cuantitativos. Generalmente se estudia la voz del consumidor durante el lanzamiento de nuevos productos para la mejor comprensión del impacto sobre los clientes. Esta información luego es utilizada para la definición de las actualizaciones del producto o, a su vez, en la determinación de las especificaciones.

Hay muchas formas de capturar la voz del consumidor, por ejemplo a través de encuestas, focus group, servicio al cliente, información de la garantía, call center, reparación. Es muy importante la participación del área de diseño y desarrollo de producto en este proceso pues es la que utilizará la información posteriormente para definir las especificaciones del producto.

Resultados encontrados en la voz del consumidor:

- Un entendimiento detallado de los deseos y necesidades del consumidor.
- Un lenguaje común que los usuarios emplean.
- Un input crítico para determinar las actualizaciones y/o las especificaciones de nuevos productos.
- Una palanca para detonar la innovación.

1.1.2. Su relación con la voz del proceso

La voz del proceso es una representación de la variabilidad, estabilidad y límites del mismo. Esto es resultado del control estadístico de procesos (CEP). El objetivo del CEP es controlar y mejorar, sobre todo mediante la técnica del gráfico de control. Escapa al alcance del proyecto profundizar sobre el CEP, simplemente se pretende mostrar la relación entre la voz del consumidor y la voz del proceso

La importancia del estudio de la variabilidad está en la percepción del cliente. Mientras históricamente las organizaciones se han concentrado en controlar la media del proceso, el foco de los consumidores está en la variabilidad que encuentran en los productos que adquieren. Por otro lado, la estabilidad dará al proceso predictibilidad y consistencia. Estos atributos son de interés para los consumidores que quieren saber qué esperar de una organización.

Los límites de control del proceso establecidos por el CEP deben caer dentro de los límites de especificación del proceso, y es ahí donde entra en juego la voz del consumidor. Esta relación da lugar a la participación del área de diseño y desarrollo de procesos en la interpretación de la voz del consumidor. Cuando se entiendan los deseos y necesidades del cliente se podrán establecer los límites de especificación del proceso. En la figura 1.1 se pueden visualizar los conceptos desarrollados en este párrafo:



Figura 1.1. La voz del proceso

1.1.3. Fuentes de voz del consumidor

En esta sección se listarán diferentes alternativas para capturar la voz del consumidor, luego en la sección 1.2.1 se explicará en mayor detalle la fuente que se usará a lo largo del proyecto.

Encuestas

Se ha considerado a la encuesta como la base de cualquier investigación de mercado. Se utiliza la encuesta para recolectar información acerca de los hábitos de compra de los distintos grupos demográficos. En muchos casos estas encuestas son incentivadas, lo cual resulta en una muestra más amplia y representativa de todos los grupos demográficos. Con el surgimiento de Internet, el método de la encuesta se ha agilizado y se pueden optimizar los resultados.

Focus group

El focus group es una técnica de investigación cualitativa, en la cual un grupo de personas responde acerca de sus actitudes y opiniones sobre un producto determinado. El grupo conformado suele ser muy interactivo e invita a los participantes a expresarse libremente. Con esta técnica, a través de pruebas y/o discusión libre, se puede evaluar el potencial del producto antes de su lanzamiento.

Servicio de atención al cliente

Este servicio es el que se le da a un consumidor antes, durante y después de una compra, para obtener su satisfacción. El objetivo es lograr que el consumidor adquiera un producto operacional en tiempo y forma. Durante la prestación de este servicio se consigue información sobre las dificultades y problemas que el consumidor ha encontrado en el producto y/o en su uso.

Garantía

La garantía asegura el nivel de calidad y rendimiento que un producto debería tener. Si en la práctica esto no resultara como indica la garantía el cliente tiene la posibilidad de devolver el producto para que sea reparado o reemplazado por uno nuevo si fuera necesario. Cuando esto ocurre, se identifica la causa de devolución del producto, sea por una falla o por no cumplir con las expectativas del cliente.

Call center

El call center es un centro de administración de llamados telefónicos, capaz de procesar una amplia cantidad de conexiones. A través de él una empresa puede manejar un flujo importante de información sobre soporte técnico y consultas. La mayoría de las grandes organizaciones utilizan el call center como una forma de interacción con los consumidores.

1.2. WEB SCRAPING

La técnica de la MT será definida en la sección 2.3, pero antes es importante entender que la entrada a este proceso es información basada en el lenguaje natural, información de texto. Para el análisis que se realizará en el presente trabajo, esa fuente de texto será obtenida de la web.

1.2.1. La web como fuente

Con el paso del tiempo, la cantidad de información que existe en la web crece exponencialmente. Esto ha cambiado la forma en que la gente accede a la información. Hoy en día si un individuo está interesado en averiguar dónde ir de vacaciones, lo hace a través de Internet; si desea investigar un tema, lo hace a través de internet; si quiere comprar un producto también lo puede hacer en la web. La interacción del ser humano con Internet es cada vez mayor y está en todo los niveles, incluso entre los consumidores. El diagrama de la figura 1.2 muestra el proceso típico que un consumidor lleva a cabo a la hora de comprar un producto:



Figura 1.2. Proceso de compra del consumidor

Las 5 etapas de este proceso el consumidor las puede efectuar en Internet. Tal vez el primer paso, el surgimiento de la necesidad de adquirir un producto que satisfaga una necesidad específica, sea el más ajeno a Internet, aunque no del todo. Contrariamente, la segunda instancia es la de mayor relevancia en la web. Es muy común hoy en día que el consumidor acceda a investigar en foros de discusión o en páginas de comercio electrónico, donde usuarios que ya han comprado el producto evalúan el producto, expresan su grado de satisfacción y opinan sobre diferentes cuestiones. Hay una infinidad de sitios con información del consumidor. A las 2 ya mencionadas le podemos agregar las redes sociales, blogs e incluso youtube. Más adelante en este trabajo se analizará el

rendimiento del “Droid Incredible”, un teléfono celular de HTC. Basta con escribir el nombre de este producto en el buscador de Facebook y se encontrará una enorme cantidad de información del consumidor.

En el tercer paso el usuario decide qué producto va a comprar, teniendo en cuenta diferentes variables como precio, marca y calidad. Esta información se ha recogido en el paso previo, por lo tanto se observa que la decisión puede ser tomada a partir de una investigación en la web. Asimismo, la compra efectiva del producto puede ser realizada a través de Internet con una tarjeta de crédito. Por último el consumidor expresa sus sentimientos después de la compra, los cuales implican un grado de satisfacción y también de insatisfacción. En este caso el usuario también puede hacer su evaluación en la web. Por ejemplo, un individuo compra una cámara digital en Amazon (un sitio muy conocido de comercio electrónico), y después puede emitir una evaluación en ese mismo sitio. Luego esta información sirve para el siguiente consumidor que se encuentre en la segunda fase del proceso de compra. Aunque es pertinente hacer una observación: la cantidad de gente que utiliza la web para la segunda instancia es mayor a la cantidad de gente que usa la web para la quinta instancia. Esto quiere decir que puede suceder que un consumidor haga su investigación a través de Internet y luego la evaluación a través de otro medio.

En los párrafos anteriores se ha explicado como un individuo puede llevar a cabo todo el proceso de compra a través de Internet. Según la oficina de censos de Estados Unidos el comercio electrónico en ese país ha crecido un 419 % en los últimos 10 años. Sin dudas, la web pública constituye una fuente inconmensurable de información sobre el consumidor. En particular sobre la voz del consumidor, que queda manifestada en la quinta etapa del proceso de compra. Uno de los objetivos de este trabajo es demostrar que la web pública contiene información muy valiosa para la organización. El problema radica en que esta información contiene un alto grado de desestructuración y su análisis se torna dificultoso. La MT, precisamente, logra estructurar la información de manera que tenga sentido para el análisis.

1.2.2. Definición de Web Scraping

En la sección anterior se manifestó la dificultad que presenta la desestructuración de la información en la web. Este no es el único problema, también sucede que esa información no está al alcance de la mano ni lista para ser usada, sino que está repartida a través de la web con diferentes formatos.

Web Scraping es una técnica de extracción del contenido de un sitio de la web, con el objetivo de cambiar su formato para que éste pueda ser utilizado en otro

contexto. En este proyecto ese contexto es el de la MT. En el proyecto se utilizará Web Scraping como medio para obtener la información que será analizada a través de la MT.

Hay diferentes métodos para recolectar información de la web. El más conocido por todos es el famoso “copiar y pegar”. “Copiar y pegar” es la forma más simple de extraer contenido de la web, y al mismo tiempo es una forma intuitiva de explicar la técnica de Web Scraping. Precisamente Web Scraping consiste en automatizar estas acciones y hacer en pocos segundos lo que a una persona le tomaría un día. Existen otros métodos como la programación HTTP y el uso de expresiones regulares. Las expresiones regulares son patrones que describen un conjunto de cadenas de texto, son muy usadas en Web Scraping.

En el presente trabajo se usará Kapow Web Data Server, una herramienta informática especialmente diseñada para Web Scraping. Esta herramienta requiere poco conocimiento de programación pues cuenta con una interface donde el usuario puede diseñar un programa llamado robot. Éste se compone de una serie de pasos, y cada paso hace una acción específica, por ejemplo “cargar página” o “extraer objeto”. El robot emula el comportamiento humano en la web y luego extrae el contenido deseado del código HTML del sitio.

1.3. MINERÍA DE OPINIÓN

1.3.1. Definición

La MO es en realidad una aplicación específica de la Minería de texto (MT). Antes de definir la MT sería propicio entender el dominio en la que ésta opera. La figura 1.3 da una referencia de este dominio:

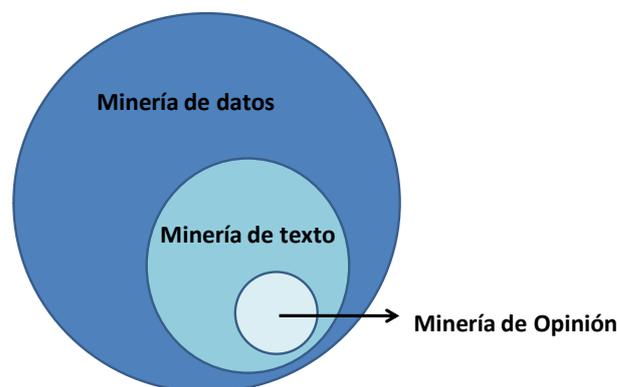


Figura 1.3. Dominio de la minería de texto

La minería de datos es el proceso de extraer patrones de un extenso conjunto de datos. Es una herramienta cada vez más influyente en el negocio moderno gracias a su capacidad de transformar datos en conocimiento valioso para

mejorar la toma de decisiones. El ámbito al cual pertenece es difícil de determinar, pues tiene relación con distintas materias como estadística, ciencias de la computación, inteligencia artificial, gestión de bases de datos y otras. Su metodología es inversa a aquella que se utiliza en estadística, donde se parte de una hipótesis o modelo y se debe verificar si los datos se ajustan al modelo. En minería de datos, se obtiene un modelo o teoría a partir de los datos. Típicamente, la minería de datos implica 4 tipos de tareas:

- Agrupar: encontrar grupos de datos semejantes
- Clasificar: aplicar estructuras conocidas sobre los datos
- Regresión: buscar un modelo que represente los datos
- Asociación: buscar relaciones entre las variables.

La MT es un proceso dentro de la minería de datos. La diferencia es que la primera trata con el lenguaje natural. Consiste en derivar conocimiento de alta calidad, a priori ignorado, a partir de un texto. Alta calidad se refiere a relevancia, novedad y atractivo. Las tareas típicas son categorización del texto, agrupación del texto y extracción de conceptos. A través del proceso de la MT se transforma el texto en datos sobre los cuales se puede aplicar minería de datos. Como la mayor parte de la información (casi el 80%) está almacenada en texto, la MT tiene un gran valor comercial¹ y un potencial ilimitado.

El tipo de conocimiento que se derive de la MT puede variar según la aplicación que se le quiera dar. En este caso particular, se pretende determinar la postura y sentimientos del escritor con respecto a un tema específico. A esto se le llama MO, y en este trabajo esa opinión representa la voz del consumidor. Sin embargo, de aquí en adelante se referenciará este tema como MT, pues ya se ha determinado la aplicación que ésta tendrá en el proyecto. Existen otras aplicaciones para la MT, las cuales serán tratadas en la siguiente sección.

1.3.2. Aplicaciones de minería de texto

Detección automática de Fraude

Analistas pueden observar un extenso conjunto de números de tarjeta de crédito para encontrar desvíos del gasto normal. Un ejemplo clásico es la compra de una pequeña cantidad de combustible seguida de un pasaje de avión, lo que exhibe una actitud sospechosa. La primera operación garantiza que la tarjeta está activa.

¹ Ah-Hwee Tan. 1999. Text Mining: Promises and Challenges.

Procesamiento automático emails/ mensajes

La ventaja en este caso sería la posibilidad de clasificar correo indeseado (junk email) a través de la identificación de palabras y frases comunes entre estos.

Biomedicina

Se han identificado enfermedades extrañas relacionando indirectamente texto común en diferentes conjuntos de literatura biomédica. Otro ejemplo es la investigación del genoma. Se han encontrado términos que coexisten en distintos artículos sobre proteínas permitiendo predecir interacciones entre ellas.

Minería de opinión (MO)

Fue tratado en la sección 1.3.1

Aplicaciones académicas

Es de gran utilidad para editoriales que cuentan con una extensa base de datos de información que necesita ser indexada para su posterior recuperación.

Seguridad, Desarrollo de SW, Marketing

Son otras de las áreas donde puede aplicarse la MT.

1.3.3. Un ejemplo didáctico

A continuación se tiene una reseña del HTC Droid Incredible, copiada del sitio de Verizon (Prestador del servicio de telefonía celular de Estados Unidos):

“This has to be the best cell phone I've ever owned. this is my first android smartphone and I'm in love with it. It dose everything I need to keep my life going with a phone music player navigator and portable internet device all in one I so glad I upgraded to an android smartphone its the best yet so far and upgrading from the EnV Touch of course it beats it hands down

- *Pros:*
larger touch screen camera featuers
- *Cons: battery life”*

Esta reseña puede ser sintetizada con la tabla 1.1:

Comentarios positivos	Comentarios negativos
+El mejor teléfono que he tenido	-Contras: Vida útil de la batería
+Me alegro de haber cambiado a un Smartphone	
+Supera al Env. Touch.	
+Pros: touchscreen más grande, atributos de la cámara	

Tabla 1.1. Los sentimientos del consumidor

En esta tabla hay tanto opiniones positivas como negativas. Si bien lo positivo es importante para la organización, la voz del consumidor se interesa particularmente en lo negativo, pues es lo que servirá como input a la hora de mejorar la calidad de un proceso. Por lo tanto, a lo largo de este proyecto se hará hincapié en los aspectos negativos de los comentarios.

A través de la MT se pueden desarrollar patrones que permitirán capturar texto y agruparlo en una categoría específica. El objetivo es identificar la frecuencia con la que un consumidor enfrenta un problema particular o se siente insatisfecho. En este caso, se podría tener una categoría llamada “Vida útil de la batería”, y capturar comentarios que indiquen la insatisfacción del cliente sobre este tema. Los patrones de la MT podrían ser:

- [contras+vida útil de la batería]
- corta vida util
- poca vida util
- vida útil más larga
- [batería+no alcanza]
- [batería+pocas horas]

Cada vez que se identifique alguno de estos patrones, la MT lo contabilizará dentro de la categoría antes mencionada. Por lo tanto este tipo de técnica requiere de un alto entendimiento del lenguaje utilizado por los consumidores, y por otro lado un amplio conocimiento de las características del producto. Ambas nociones aportan conceptos que la MT debe reconocer como tales antes de poder procesar la información de manera efectiva.

Si una misma reseña contuviera dos (2) o más comentarios negativos acerca de un mismo tema, esto no implicaría que el conteo para esa tema vaya a ser el mismo. Todo depende de la finalidad del análisis. Si se quiere saber de qué trata la reseña, se puede ajustar la MT para tal propósito y en ese caso el conteo será uno (1). Si en cambio se quiere saber la frecuencia del concepto dentro de la reseña, entonces el conteo será dos (2) o más.

1.4. VISIÓN GENERAL DEL PROYECTO

Definidos los temas que se tratarán a lo largo del presente trabajo, en esta sección se pretende dar una definición al proceso que será ejecutado para capturar la voz del consumidor y para conseguir los objetivos que serán planteados en la sección 1.5.

1.4.1. Diagrama de proceso

En la figura 1.4 se presentan las operaciones y en última instancia el proceso general que tendrá lugar en el proyecto:

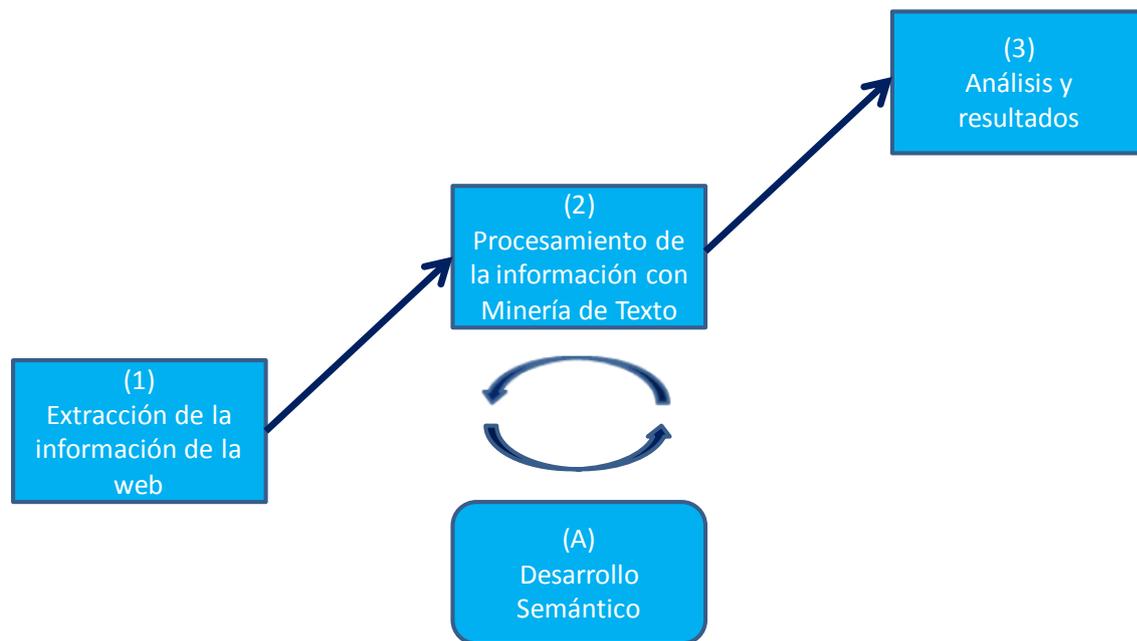


Figura 1.4. Diagrama de proceso

Este diagrama es el que sigue una organización en condiciones de producción. Se observa que el proceso cuenta de 3 pasos y también de un procedimiento alternativo. El desarrollo semántico es la construcción del motor que dará funcionamiento a la MT. En producción esto no se necesita hacer en forma reiterativa pues dejarían de existir las ventajas de la MT, principalmente la de evitar tener que leer una gran cantidad de información. Siguiendo la analogía del motor, una vez que éste está en funcionamiento, se necesitarán mantenimientos esporádicos para asegurar un adecuado rendimiento. En este caso ese mantenimiento sería la incorporación de nuevos conceptos y la mejora de la robustez y la calidad de la MT, tema que será tratado en mayor detalle en el capítulo 3.

En este trabajo, sin embargo, habrá un capítulo dedicado a este proceso alternativo, pues el enfoque de este proyecto no es el de la producción, sino el de la anterior etapa de desarrollo: se construirá el motor y luego se lo pondrá en marcha.

Para un mejor entendimiento de cómo fluye la información en el proceso descrito al inicio de esta sección, es oportuno atender a la figura 1.5

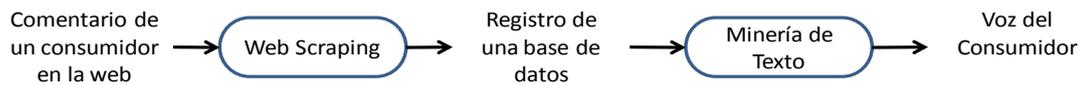


Figura 1.5. Diagrama de flujo de información

1.4.2. SIPOC

El SIPOC que se presenta en la tabla 1.2 (suppliers, input, process, output, clients) permitirá entender la aplicabilidad del proyecto desde el punto de vista de los proveedores y de los clientes del proceso.

Proveedores	Entradas	Proceso	Salidas	Clientes
Consumidores comentando en distintos sitios de la web: <ul style="list-style-type: none"> • Foros • Comercio electrónico • Redes Sociales • Otros 	<ul style="list-style-type: none"> • Discusiones • Reseñas • Opiniones 	El explicado en la sección 1.4.1, en particular MT.	Información de alto valor expresando la voz del consumidor	<ul style="list-style-type: none"> • Diseño de producto • Ingeniería de proceso • Marketing

Tabla 1.2. SIPOC

1.5. ALCANCE DEL PROYECTO

En este segmento se definirán las fronteras de la tesis.

1.5.1. Objetivos

Esta tesis tiene 2 objetivos principales:

1. A través del uso de la MT, identificar los sentimientos, actitudes, opiniones y posturas de un grupo de individuos hacia un determinado objeto de estudio.

2. Demostrar que el dominio público de la web constituye una valiosa fuente de información sobre los consumidores para la organización.

1.5.2. Objeto de estudio

Se estudiará principalmente el **Droid Incredible**, un teléfono celular de la compañía HTC. Es un dispositivo móvil de alta gama que vende el carrier Verizon en Estados Unidos.

Por otro lado se tomarán otros dispositivos móviles de la competencia para realizar una comparación en el capítulo 4. Por necesidad deben pertenecer a la misma categoría. Estos serán el iPhone4, el Droid X de Motorola, y el BlackBerry Bold 9700.

La elección de este producto proviene principalmente de dos razones. En primer lugar, la cantidad de discusión que se genera en Internet sobre este tipo de producto es mayor que la de cualquier otro producto o servicio. En segundo lugar, la experiencia del autor en el tema permitirá una mejor calidad de trabajo.

De todas formas es importante aclarar que este proceso puede ser aplicado a cualquier objeto que genere discusión en la web. Por ejemplo GPS, cámaras digitales, restaurantes, hoteles, automóviles.

1.5.3. Fuentes del análisis

Por la naturaleza del proyecto las fuentes serán íntegramente obtenidas de la web pública. Además estas fuentes estarán en idioma inglés, puesto que el producto se vende en Estados Unidos, sus respectivos consumidores hablan este idioma. Por supuesto, el análisis que se realice será en español.

Otras fuentes, por ejemplo las mencionadas en la sección 1.1.3 igualmente podrían ser sometidas al mismo procedimiento que se plantea en este trabajo y en simultáneo. Lo que sucede es que usualmente esa información es totalmente confidencial para una organización, y eso implica una complicación para el autor. Pero si se deseara, cualquier fuente en base al lenguaje natural podría ser considerada.

1.5.4. Ideas presentadas en el trabajo

La mayoría de las afirmaciones e ideas presentadas en este trabajo son producto de la experiencia del autor en el tema, en particular las que se enunciarán los capítulos 3 y 4. El proceso empleado es una adaptación de un proceso real del cual el autor tuvo una participación en su elaboración y

aplicación. Por otro lado, las ideas que corresponden a terceros están debidamente referenciadas en el texto.

1.5.5. Herramientas del estudio

- IBM SPSS MODELER. Es un SW de minería de datos y MT que permite modelar la información en forma intuitiva para el usuario, sin tener que hacer uso de la programación. Es la herramienta principal del trabajo, con la que se realizará el análisis y con la que se buscará conseguir los objetivos planteados.
- KAPOW WEB DATA SERVER. Como se mencionó en la sección 1.2.2 se utilizará para extraer información de la web.
- ORACLE DATABASE. Donde se almacenará la información.
- MICROSOFT OFFICE. Se utilizará para presentar la documentación y los resultados.

1.5.6. Limitaciones

- Las fuentes a utilizar no son representativas del total de los consumidores. Esto se debe a la simple razón de que no todos los consumidores expresan sus inquietudes en la web. Otros individuos tienen otro tipo de comportamiento y resuelven sus inconvenientes con otra alternativa. De todas maneras, la web efectivamente representa una importante cantidad de usuarios.
- La MT no es una ciencia exacta. Por un lado, las formas en que una persona puede expresar su opinión son infinitas, siempre existirán sentimientos que escapan a los patrones semánticos desarrollados en la MT. Por otro lado, y por la misma razón, la precisión jamás será del 100%. Algunos patrones pueden fallar en su propósito. Por ejemplo, si alguien dice “¿Me gusta este producto? En realidad no” o “Me gusta el producto. Mentira, estoy siendo irónico”. En estos casos, si existiera un patrón [gustar + producto] procesaría la información en forma equivocada.

2. RECOLECCIÓN DE LA INFORMACIÓN

En este capítulo se hará una exploración de la Web en la búsqueda de sitios de interés que contengan información del Droid Incredible y de sus competidores. Se hará una selección de sitios y se definirán los objetos relevantes a extraer, como usuario, fechas, opinión que tendrán que ser objetos comunes a todos los sitios. La información será almacenada en una base datos para su posterior procesamiento y análisis. Luego, será necesario aplicar ciertas manipulaciones sobre la información para favorecer la aplicabilidad y representatividad del estudio. Para entender en qué punto del proceso se sitúa el presente capítulo, la figura 2.1 hace referencia al diagrama de proceso presentado en la sección 1.4.1:

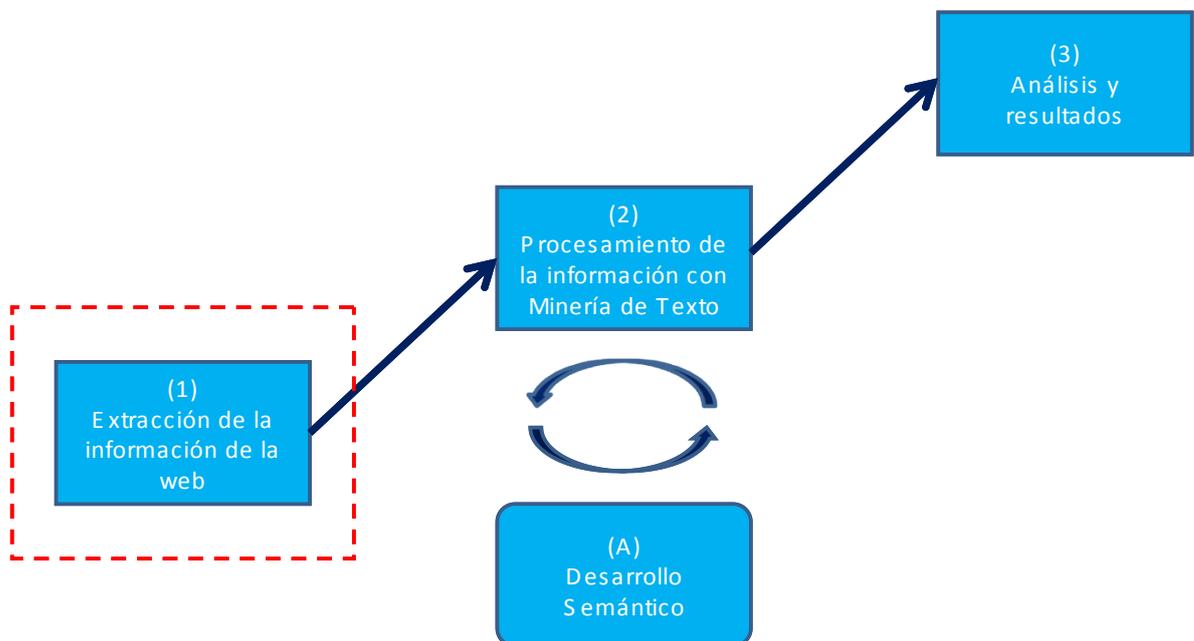


Figura 2.1. Localización del capítulo 2 en el proceso

2.1. OBSERVACIÓN DE LA WEB

2.1.1. Sitios con CGU de interés

CGU es un acrónimo de Contenido Generado por el Usuario. Existen diferentes tipos de sitios donde los usuarios pueden generar contenido expresando sus opiniones e inquietudes. Puede ocurrir que un mismo usuario acceda a varios al mismo tiempo, pero en general cada sitio atrae un perfil específico de consumidores. Una observación profunda de la Web resultó en el hallazgo de los siguientes tipos de sitios:

Foros

Probablemente el tipo de sitio con mayor proporción de CGU, prácticamente la totalidad del sitio es generado por lo que sus usuarios escriben. En los foros, se plantean temas para discutir donde los usuarios pueden obtener respuestas a sus preguntas, plantear inquietudes, ayudar a otros o simplemente informarse sobre cierto asunto.

Todos los foros que existen en Internet siguen un mismo formato:

- Una página principal clasificando el contenido del foro. Por ejemplo, por marcas de teléfonos celulares: HTC, Apple, Nokia, Motorola, BlackBerry, etc.
- Accediendo a una de las categorías enunciadas en el punto anterior, se visualiza un listado de hilos de discusión (thread).
- El listado de hilos está ordenado cronológicamente de más reciente a más antiguo, y cada hilo tiene su correspondiente título para que los usuarios sepan de qué trata.
- Dentro de cada hilo hay una compilación de mensajes emitidos por los usuarios acerca del tópico planteado por el título del hilo.
- Estos mensajes se llaman “posts” y también están ordenados cronológicamente, pero de más antiguo a más reciente.

Los foros poseen mucha información del consumidor, pero lo que sucede es que existe un exceso: muchas veces las discusiones se desvirtúan y se transforman en una especie de chat sobre las nimiedades de la vida cotidiana. Esto último genera ruido para la MT.

Sitios de comercio electrónico

También conocidos como e.commerce. Son sitios donde se puede efectuar la compra y venta de productos o servicios. Luego de efectuar la compra el usuario tiene la posibilidad de hacer una reseña sobre el producto que ha adquirido, tal como la que se usó como ejemplo en la sección 1.3.3

Usualmente las reseñas cuentan con las siguientes particularidades:

- Un puntaje del producto (del 1 al 5 o del 1 al 10)
- Una breve evaluación
- Pros y contras

Sitios sobre telefonía celular

Contienen información sobre dispositivos móviles, como noticias, lanzamientos, especificaciones. Además cuentan con la alternativa de hacer una reseña como las encontradas en comercio electrónico, aunque en menor cantidad.

Facebook

Al igual que los foros, prácticamente todo el contenido es generado por los usuarios. Es la red social de mayor envergadura y uso en la actualidad². Principalmente se usa para relacionar gente, pero también los usuarios se pueden conectar con organizaciones, universidades y agrupaciones. Por ejemplo HTC tiene una página dentro de Facebook y mucha gente se relaciona con la empresa a través de mensajes en el muro. El muro es un espacio en el perfil de cada usuario que permite a sus amigos dejar mensajes.

Sobre el muro, muchos consumidores plantean preocupaciones o alzan preguntas que bien pueden ser resueltas por la empresa o por otros usuarios. También existen dentro de Facebook páginas para los productos, como el Droid Incredible, con exactamente la misma funcionalidad. Esto se traduce en información valiosa sobre los consumidores.

2.1.2. Selección de los sitios

Se hizo una selección extensa de 24 sitios, listados en la tabla 2.1. Es importante aclarar que el autor reconoce la existencia de otras páginas con información sobre teléfonos celulares, pero la adición de las mismas no agregaría mucho valor al análisis. Por otro lado, hay que recordar que para cada uno de estos sitios se debe desarrollar un robot (sección 1.2.2) que extraerá información de ellos. Estos robots toman horas de trabajo, y la adición de otros sitios representaría mucho trabajo con poca utilidad. En la tabla 2.1 se muestran los sitios seleccionados para el análisis.

Se percibe en la selección de sitios que algunos pertenecen al fabricante (HTC, Motorola o Apple), otros pertenecen a los más importantes carriers de Estados Unidos (Verizon, AT&T y T.Mobile) y también hay muchos sobre Android. En el capítulo 3 se describirán las características del Droid Incredible, pero en este punto es conveniente aclarar que Android es el sistema operativo de este celular. Así pues, se advierte que la selección ha tenido en cuenta los factores más relevantes y entonces, dentro del dominio de la web, conforma una

² Morejon, R. 2010. Top 10 Social-Networking websites and forums.

muestra representativa de información sobre el Droid Incredible y sus competidores. A continuación se puede apreciar la tabla 2.1

Sitio	Tipo	Cantidad de información	Relación Calidad/Cantidad
Amazon	e.commerce	Baja	Alta
Android Community	Foro	Baja	Mediana
Android Forums	Foro	Alta	Mediana
Apple support	Foro	Alta	Mediana
AT&T F	Foro	Mediana	Mediana
AT&T R	Celulares	Baja	Alta
Best Buy	e.commerce	Baja	Alta
BlackBerry Support	Foro	Alta	Mediana
Cellphone Forums	Foro	Baja	Mediana
Cnet	e.commerce	Mediana	Alta
Facebook	Red Social	Mediana	Baja
Gsmarena	Celulares	Mediana	Baja
Howard Forums	Foro	Alta	Mediana
HTC Support	Foro	Alta	Mediana
Motorola Support	Foro	Alta	Mediana
Letstalk	e.commerce	Baja	Alta
Phonearena	Celulares	Baja	Alta
Phonescoop	Celulares	Baja	Alta
Shopping	e.commerce	Baja	Alta
Talk Android	Foro	Baja	Mediana
T.Mobile Forum	Foro	Alta	Mediana
Verizon F.	Foro	Baja	Mediana
Verizon R.	Celulares	Alta	Alta
Viewpoints	e.commerce	Baja	Alta

Tabla 2.1. Sitios seleccionados

En la tabla se advierten dos campos que fueron introducidos a modo ilustrativo y son puramente cualitativos. Cantidad de información se refiere a la cantidad de comentarios por teléfono generados por el usuario: baja para el orden de las decenas, mediana para el orden de las centenas y alta para el orden de los miles. Relación calidad/cantidad se refiere a la frecuencia con que estos comentarios expresan la voz del consumidor.

2.2. EXTRACCIÓN DE LA INFORMACIÓN

En esta sección se explicará qué es lo que se extraerá de la web. Sin embargo no se profundizará sobre el proceso de Web Scraping ni el desarrollo de robots. Aunque cabe aclarar que ese trabajo fue realizado igualmente para poder llevar a cabo el trabajo.

2.2.1. Identificación de los atributos relevantes para el análisis

De la selección de sitios en la tabla 2.1, se pueden encontrar dos formatos en común: el foro y la reseña. Un caso aparte es el de Facebook, que se asemeja a la discusión en el foro pero tiene su propio formato. Entonces en definitiva se tienen 3 conjuntos de datos, de los cuales se extraerá una serie de atributos. Se comenzará analizando la reseña, la figura 2.2 es una reseña sacada de Cnet. En ella se marcan los elementos relevantes con recuadros rojos.

(1) ★★★★★

(2) "Great smartphone"

(3) by BenMNSTATE

(4) on April 29, 2010

(5) Pros: FAST processor, lots of memory, nice big screen, android 2.1.

(6) Cons: Battery life is good for a smartphone, but isn't going to last multiple days like a simple flip phone.

(7) Summary: I got the Droid Incredible after having a Blackberry tour and Motorola Droid. This is the best by far. If only for the processing speed alone, it's worth the money. No lag or slowdowns, even with multiple apps running. The camera is at least as good as the Motorola droid. The screen is awesome. Responsive and bright. I was at first very worried about typing without a keyboard (especially after coming from a blackberry), but this screen is so big that I'm not finding it to be a problem. Android is a GREAT OS. I loved it on the Motorola Droid and I love it on the Droid Incredible. Once you get used to it you really see how outdated the blackberry OS is.

Battery life is pretty good. In my test day I was streaming music from pandora, browsing the internet, and texting. I got a full day of use and that will probably increase once the battery cycles a few times. Yes, I would love a smartphone that could go five days without a charge, but really the Droid Incredible's battery life is on par with what your going to find in smartphone these days. Battery tech is still catching up to smartphone energy demands. That being said, is it really that hard to plug in a phone every night?

Overall, GREAT smartphone. If you want a touch screen only smartphone with Android, there aren't any better (for now).

Reply to this review | Read reply (1) | Was this review helpful? (23) 👍 (3) 👎

Figura 2.2. Reseña del Droid Incredible sacada de Cnet

Se observa que la reseña cuenta con una calificación (1), un título (2), un usuario (3), una fecha de evaluación (4), Pros (5), Contras (6) y finalmente la reseña propiamente dicha (7). Estos son los elementos a extraer, siendo el último el de mayor interés.

MO para la voz del consumidor

En los foros se tienen dos instancias, primero el listado de hilos, luego los posts dentro de un hilo. La figura 2.3 muestra como se visualizan los hilos en el foro de Android:

	Thread / Thread Starter	Rating	Last Post	Replies	Views
	Sticky: HTC Incredible Guide: Updated 10:30am EST, Aug 31, 2010 lekky		08-30-2010 07:41 PM by lekky	0	3,761
	FYI - Swype Beta Updated for FROYO (1 2) TitanAlum		Today 09:23 PM by livspop	63	2,558
	If you want an extended battery...just threaten! :)(1) truestory.d:(2)		Today 08:57 PM by elborak	25:(5)	979:(6)
	Moved: Handcent Themes Piercing Heavens		-	-	-
	Let get serious about Incredible camera picture quality (1 2 3) mrspeedmaster	☆☆	Today 07:40 PM by Muerte_X	118	6,506

Figura 2.3. Hilos en el foro de Android

Y en la figura 2.4 se muestra como se visualizan los posts dentro del hilo:

Figura 2.4. Un post del foro Android

Los recuadros rojos marcan los elementos a extraer: título del hilo (1), usuario generador del hilo (2), fecha del último post (3), último usuario en escribir un post dentro del hilo (4), cantidad de respuestas al hilo (5), cantidad de vistas al hilo (6), fecha del post (7), nombre de usuario (8), cantidad de posts del usuario (9) y finalmente el post propiamente dicho (10). El último elemento es el de mayor interés para el proyecto.

Finalmente se tiene el muro de los productos dentro de Facebook. La figura 2.5 enseña los comentarios en el muro del Droid Incredible.

HTC Droid Incredible Is anyone here using the Innocase Active or the Steinheil Ultra Crystal Clear? If so what are your thoughts?
El 11 de septiembre a las 14:01 · Denunciar



Figura 2.5. Comentarios en el muro

De aquí los elementos a extraer serán simplemente el usuario (1), el comentario (2) y la fecha del comentario (3).

2.2.2. Almacenamiento de la información

En la sección 2.2.1 se presentaron los sitios que se utilizarán en el proyecto y los elementos que serán extraídos de cada uno de ellos. Lo más relevante es el comentario que los usuarios hacen, pues la MT se aplicará sobre éstos. Para cada uno de estos sitios se desarrolló un robot con Kapow Web Data Server. El robot almacena la información en una base de datos, y luego con SPSS Modeler se accede a la base de datos y se utiliza la información su análisis.

Para almacenar la información se generaron tres tablas en la base de datos. Una tabla para los foros, una tabla para las reseñas y una tabla para Facebook. Más adelante, cuando se aplique la MT, esta información será unificada y analizada simultáneamente. El propósito de dividirla en tres es tener una base de datos ordenada, donde la información de un mismo tipo se encuentra separada de los otros tipos. La tabla 2.2 muestra los campos que cada una de estas tablas tiene.

Foros	Reseñas	Facebook
Usuario	Usuario	Usuario
Comentario	Comentario	Comentario
Fecha del comentario	Fecha del comentario	Fecha del comentario
Nombre del teléfono	Nombre del teléfono	Nombre del teléfono
Sitio	Sitio	Sitio
Título	Título	Tipo
Usuario Generador del hilo	Pros	
Ultimo usuario que posteo	Contras	
Fecha del último post	Calificación	
Cantidad de respuestas al hilo		
Cantidad de vistas al hilo		
Cantidad de posts del usuario		

Tabla 2.2. Campos de las tablas

2.2.3. Resultados de la extracción

	Droid X	Droid Incredible	iPhone 4	B 9700	Total
Amazon	44	115	0	17	176
Android Community	54	70	0	0	124
Android Forums	24943	48109	0	0	73052
Apple support	0	0	37075	0	37705
AT&T F	0	0	11811	117	11928
AT&T R	0	0	0	29	29
Best Buy	38	22	95	11	166
BlackBerry Support	0	0	0	8717	8717
Cellphone Forums	15	96	190	19	320
Cnet	108	265	286	29	688
Facebook	320	1233	70	47	1670
Gsmarena	93	35	2588	706	3422
Howard Forums	2528	2228	9134	1	13891
HTC Support	0	6566	0	0	6566
Motorola Support	13910	0	0	0	13910
Letstalk	18	15	0	49	82
Phonearena	12	27	15	4	58
Phoncoop	31	59	34	5	129
Shopping	0	0	15	5	20
Talk Android	49	2	0	0	51
T.Mobile Forum	0	0	0	1888	1888
Verizon F.	0	3788	0	0	3788
Verizon R.	2269	2800	0	0	5069
Viewpoints	14	11	299	18	342
Total	44446	65441	61612	11662	183161

Tabla 2.3. Cantidad de registros obtenidos

La tabla 2.3 resume la cantidad de registros almacenados en la base de datos, por sitio y por producto. Cada registro contiene un comentario de un usuario.

Los resultados que se presentan en esta sección pueden ser engañosos, se podría caer en el error de pensar que algunos sitios aportan información poco significativa con respecto a otros. La tendencia, en general, es que los foros presentan mucha mayor información que las reseñas. Lo que sucede es que por cada reseña hay un usuario, en cambio en los foros un usuario puede realizar varios comentarios. En los foros existe el principio de la desigualdad de la información, un problema que será tratado en la sección 2.3.1. Lo que importa en este punto es entender que si bien la diferencia entre los comentarios puede ser muy amplia, la cantidad de usuarios comentando es del mismo orden.

En las figura 2.6 se presentan los porcentajes de participación de cada de sitio, separando por un lado sitios con desigualdad (foros y Facebook) y por otro, las reseñas.

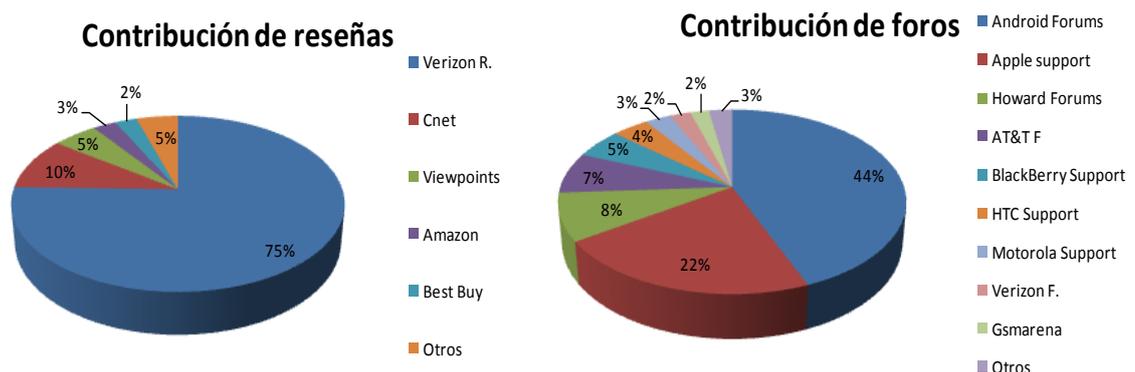


Figura 2.6. Contribución de los sitios.

Entre los foros se destaca la contribución Apple Support y Android con el 66% de participación. Entre las reseñas de destaca Verizon, con 5069 registros, tomando un 75% del contenido.

Si se observa nuevamente la tabla 2.3 se nota que el Droid Incredible es el que más información aporta entre los productos, mientras que Bold 9700 tiene un menor aporte. Esto indica en cierta forma la popularidad de los productos. De todas formas hay que tener en cuenta que los productos fueron lanzados en distintas fechas. La extracción de la información se realizó desde el 21 de abril de 2010, fecha de lanzamiento del Droid Incredible, hasta el 20 de Septiembre de 2010. Es decir, los primeros 5 meses del producto en el mercado. La tabla 2.4 sintetiza los períodos considerados.

Producto	Lanzamiento	Comienzo de extracción	Fin de extracción
Droid Incredible	21/04/2010 (en Verizon)	21/04/2010	20/09/2010
Droid X	15/07/2010 (en Verizon)	15/07/2010	20/09/2010
iPhone 4	24/06/2010 (en AT&T)	24/06/2010	20/09/2010
Bold 9700	16/11/2009 (en T.Mobile)	21/04/2010	20/09/2010

Tabla 2.4. Delimitación temporal de la información extraída

Todos los comentarios extraídos en este capítulo se encuentran en el Anexo 1.

2.3. CONSIDERACIONES SOBRE LA INFORMACIÓN EXTRAIDA

2.3.1. Desigualdad en el contenido generado por los usuarios (CGU)

El CGU se está convirtiendo en la más popular y valiosa información disponible en la Web. Poco se ha estudiado la forma en que esta información se produce. Xavier Ochoa, en su trabajo “Quantitative analysis of user generated content on the Web”, demuestra estadísticamente la aplicabilidad del principio de la desigualdad sobre el CGU. Este principio es comparable al de Pareto, quien diría “El 80% del contenido es atribuible a un 20% de los usuarios”.

En su trabajo, Xavier Ochoa ajusta la producción del CGU a una distribución de Lotka o de Weibull, asemejándose a la generación de artículos científicos o académicos. Es decir que la producción no tiene comportamiento gaussiano, el concepto de usuario promedio no se refleja en la Web. Mientras la mayoría de los usuarios generan una sola publicación en un sitio, unos pocos son responsables de la mayoría de las publicaciones de ese mismo sitio.

Debido a estas implicaciones, es conveniente estar atentos y tener cuidado a la hora de utilizar la información que se ha extraído de la Web. En la sección 2.3.2 se explicará como eludir estos inconvenientes que podrían llevar a conclusiones equivocadas.

2.3.2. Conciliación de la información

Dado el problema de la desigualdad, es necesario hacer algunas manipulaciones sobre la información extraída. Usando la herramienta de MT, se tomaron los registros de la base datos y se aplicó sobre ellos una serie de operaciones lógicas con el propósito de unificar la información por usuario. Es decir, todos los registros pertenecientes a un usuario particular, se unieron para

que cada usuario aporte exactamente una opinión. Esto implica que algunos tendrán opiniones más largas que otros pero de esta manera se evita sesgar los resultados hacia ciertos usuarios. Por ejemplo, si un consumidor se queja 90 veces sobre la velocidad de internet en el teléfono, la herramienta de MT entendería que existen 90 documentos que hablan sobre ese problema. En cambio, si se agrupa la información por usuario, existirá un documento con 90 quejas, pero la MT analiza ese documento únicamente y dirá que se tiene solamente una queja sobre la velocidad de internet. Hacer la MT sobre los usuarios en vez de los comentarios da mayor precisión al análisis y evita caer en la desigualdad.

Si bien esta limpieza dará mayor robustez al análisis final al mismo tiempo enfrenta una limitación: un mismo individuo puede contar con dos (o más) usuarios para igual cantidad de sitios. Por ejemplo, el consumidor puede realizar una reseña del iPhone4 en Best Buy con un nombre de usuario, y luego discutir en el foro de Apple bajo otro nombre de usuario. Esta limitación representara ruido para el análisis pero no hay forma de evitarla. Se podría evitar tomando únicamente un sitio pero a costa de sacrificar una muestra mayor y más diversa de información.

3. PROCESO DE LA MINERÍA DE TEXTO

En el presente capítulo se hará inicialmente una descripción detallada del producto. Una vez registrados los atributos se procederá al desarrollo semántico que funcionará como motor de la MT, esto implica elaboración de librerías de conceptos y reglas TLA (patrones semánticos). Toda esta semántica dará lugar a la descripción precisa de los problemas que enfrenta el consumidor. Cada problema será subordinado a una categoría objetivo, dentro de la cual la MT clasificará las opiniones recolectadas en el capítulo 2. La figura 3.1 permite visualizar en qué punto del trabajo se sitúa este capítulo.

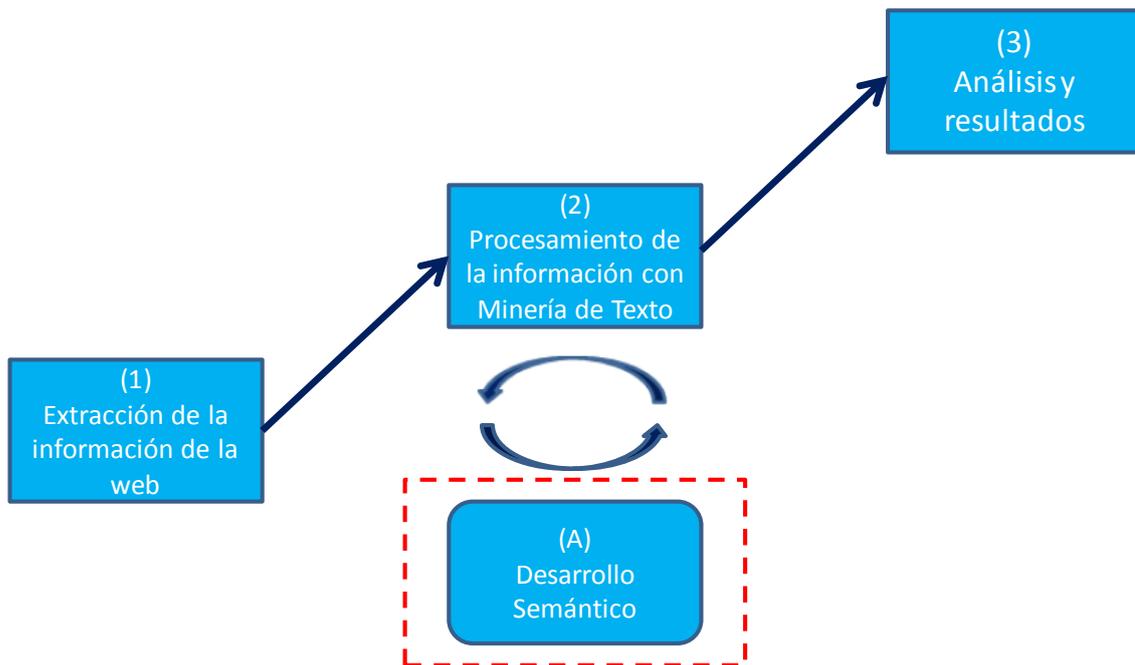


Figura 3.1. Localización del capítulo 3 en el proceso

3.1. VALORACIÓN DEL PRODUCTO

La importancia de registrar las características del producto radica en que la MT debe reconocer cuáles son los conceptos que debe capturar cuando procese la información. Por ejemplo, la MT debe reconocer el concepto “GPS” antes de poder clasificar este término en una de las categorías objetivo. Un adecuado entendimiento del producto dará lugar a una MT de alta calidad.

3.1.1. Descripción del producto

El Droid Incredible es un celular de alta gama de la empresa HTC. HTC es una empresa de origen taiwanés que se dedica a la producción de smartphones, teléfonos celulares que ofrecen mayores habilidades computacionales que los celulares ordinarios. En Estados Unidos este producto es vendido por Verizon,

el proveedor del servicio de red (o carrier), mientras que en Argentina no está disponible principalmente porque la empresa HTC no tiene operaciones en el país (además porque la tecnología CDMA del producto lo haría inutilizable aquí).

El Incredible es un teléfono con pantalla táctil, es decir mediante el tacto de su superficie permite la entrada de datos y órdenes al dispositivo. El sistema operativo del teléfono es Android, el sistema desarrollado por Google en 2007. Desde su lanzamiento, en 2009, ha crecido considerablemente su popularidad al punto tal de ser uno de los principales competidores del mercado. En la figura 3.2 se observa la participación en el mercado de Estados Unidos en julio de 2010.

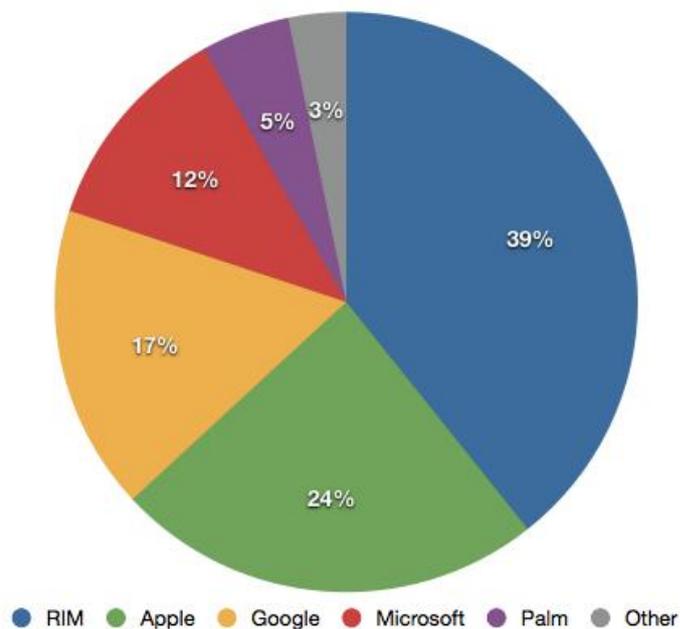


Figura 3.2. Participación de OS en el mercado hacia julio 2010 (Fuente: comScore)

Se percibe que Android es el tercer competidor en el mercado y las tendencias indican que seguirá creciendo. De los productos competidores elegidos a modo comparativo, el Droid X también es un teléfono Android, el iPhone4 es el último dispositivo de Apple y BlackBerry (RIM) Bold 9700 pertenece al principal segmento. Cabe aclarar que estos productos son bastante similares entre sí y son los de más alta gama al momento de comenzar el trabajo, de aquí la elección de los mismos. La principal diferencia la hace el Bold 9700, que no cuenta con pantalla táctil, sino con un teclado físico.

3.1.2. Identificación de las características

En la tabla 3.1 se pueden visualizar todas las especificaciones del producto. Estos son los conceptos que la MT deberá capturar.

Característica principal	Característica secundaria	Detalles
Tecnología de Red	CDMA	800/1900
Datos	Datos CDMA	1xEV.DO rev.A
Diseño	Forma	Barra de chocolate
	Dimensiones	117.5 x 58.5 x 11.9 mm
	Peso	130g
	Antena	Interna
	Teclas laterales	Control de volumen
	Navegación	Modo táctil
Pantalla	Tipo	Color
	Tecnología	AMOLED
	Resolución	480X800 pixeles
	Tamaño	9,4 cm
	Pantalla Táctil	Capacitiva
	Táctil múltiple	Sí
	Sensor de proximidad	Sí
	Sensor de luz	Sí
Batería	Tipo	Li.Ion
	Capacidad	1300 mAh
	Duración en llamada	5,21 horas
	Duración en suspensión	146 horas
Software	Android OS	2.1 (actualizable)
Hardware	Procesador	Snapdragon 1000 MHz
	Memoria	512 RAM / 512 ROM
Cámara	Tipo	Inmóvil
	Resolución	8 mega pixeles
	Flash	Dual LED
	Atributos	Auto foco, Geo tagging
	Video grabación	Sí
Multimedia	Reproductor de música	MP3, AAC, WMA, WAV, M4A, AMR, OGG, MIDI
	Reproductor de video	MPEG4, WMV, 3GP, 3G2
	Radio	FM
	Streaming	Sí
Navegación en Internet	Sí	HTML, flash Lite
Servicio de localización	A.GPS	Con navegación
Agenda de contactos	Capacidad	Depende de la memoria
	Características	Ring ID, Imagen ID, Búsqueda por apellido y nombre, múltiples números por usuario, agrupación de llamados.

Característica principal	Característica secundaria	Detalles
Organizador	Calendario	Sí
	Alarma	Sí
	Tareas	Sí
	Vista de documentos	Sí
	Otros	Calculadora, reloj, notas.
Mensajería	SMS	Sí
	MMS	Sí
	Plantillas	Sí
	Email	POP3/IMAP/SMTP
	Mensajería Instantánea	Google Talk
Memoria	Expansión	microSD/microSDHC 32GB
	Incorporada	8192 MB
Conectividad	Bluetooth versión 2.1	Headset (HSP)/Handsfree (HFP)/Transferencia de archivos (FTP)/Object Push (OPP)/Distribución avanzada de audio (A2DP)/Acceso a la agenda (PBAP)
	Wi.Fi	802.11b/802.11g/802.11n
	USB 2.0	MicroUSB
	Auriculares	3.5mm
	Conector del cargador	microUSB
	Salida de TV	Sí
	Sincronización	Con computadora y OTA.
Notificaciones	Ringtone polifónicos	Sí
	Ringtone de música	MP3
	Vibrador	Sí
	Luces de servicio	Sí
	Modo avión	Sí
	Modo silencio	Sí
	Altavoz	Sí
Otras características	Marcado de voz	Sí
	Comandos de voz	Sí
	Grabación de voz	Sí
	Sensores	Acelerómetro, brújula
	TTY/TDD	Sí

Tabla 3.1. Especificaciones del Droid Incredible (Fuente: Phonearena.com)

Sobre cada uno de los conceptos listados en la tabla 3.1 el consumidor puede emitir una opinión. Algunas características, como por ejemplo la cámara, están dentro de la percepción de la mayoría de los usuarios. En cambio otras, como la velocidad del procesador, solamente son consideradas por algunos usuarios con perfil más técnico.

3.2. DESARROLLO DE LA MINERÍA DE TEXTO

El trabajo realizado en la presente sección será determinante para la concreción de los objetivos y resultados. Todo se logra partir del desarrollo de una sólida MT que permita procesar la información en forma adecuada. Es importante destacar que los procedimientos que se lleven a cabo en este punto corresponden al entorno del IBM SPSS Modeler, sin embargo otros entornos presentan funcionalidades similares.

3.2.1. Desarrollo de las librerías de conceptos

Antes que nada, es oportuno recordar lo mencionado en la sección 1.5.3: el lenguaje hablado por el segmento de mercado considerado es el inglés. Por lo tanto todos los términos a los que se hace referencia están en ese idioma. A los efectos prácticos no hay diferencia con hacer este mismo procedimiento para el español, pero es necesario hacer esta aclaración.

Uno de los primeros pasos a la hora de realizar MT es la extracción de conceptos. Es un proceso que indica cuáles son las palabras de mayor recurrencia dentro de un conjunto de archivos. Entonces, en el caso de este trabajo, se tomó una muestra aleatoria de 1000 comentarios de la información extraída en el capítulo anterior sobre el Droid Increíble. De esta manera se corroboró que las características listadas en la tabla 3.1 de la sección 3.1.2, resultaron ser términos relevantes. También, para cada característica, se encontraron sus conceptos asociados. Por ejemplo, la característica Email tiene asociados varios términos: bandeja de entrada, correo, mensaje, contactos, etc. Para que la herramienta reconozca a qué género pertenecen cada uno de los conceptos, éstos deben ser guardados en un módulo llamado librería.

El segundo paso es justamente la introducción de los conceptos en una librería. Al realizar esta operación hay que tener en cuenta términos equivalentes. Por ejemplo, en la sección “Mensajería” de la tabla 3.1 se tiene el concepto SMS, del cual es sinónimo “Mensaje de texto”. El mal uso de lenguaje también debe ser considerado, pues una individuo podría decir “mensaje de texto” o bien podría decir SSM por un error de tipeo. (Haciendo referencia al inicio de esta sección, lo que verdaderamente se está introduciendo en la herramienta es “Text message”).

El IBM SPSS Modeler trae consigo unas librerías genéricas con gran variedad de términos, por lo que algunos de las características de la tabla 3.1 y sus términos asociados ya se encuentran allí, aunque la gran mayoría no. Algo que también incluye, resultando de gran utilidad, son términos relacionados al

sentimiento. Por ejemplo “no me gusta”, “me encanta”, “odio”, “me molesta”, etc. Estos son necesarios para capturar la voz del consumidor. Por otro lado queda excluido el lenguaje coloquial, palabras que los usuarios emplean en un contexto informal y que deben ser incorporadas también.

Finalmente se desarrollaron dos librerías principales: una de investigación y otra de características. Cada una de estas librerías cuenta con diferentes géneros, y cada género refiere a una clase específica de conceptos. En la tabla 3.2 se listan todos los géneros y algunos ejemplos de sus conceptos.

Librería: Características		Librería: Investigación	
Género	Conceptos	Género	Conceptos
Batería	Carga, vida, power	Ocurrencia	Cuándo, desde, frecuente
Aplicaciones	Juegos, office, programas	Consultas	
Accesorios	Auriculares, cable USB, cargador para auto	Positivo	Amor, felicidad, rápido
Bluetooth	Señal, apareamiento	Negativo	odio, triste, lento, roto
Llamada	Altavoz, micrófono, marcado de voz	Acciones	tomar, enviar, recibir
Memoria	Interna, opcional, 8GB		
E.mail	Bandeja de entrada, correo, contactos		
GPS	Brújula, mapas, dirección		
Teclado	Teclas, querty, botón		
Contactos	Agenda, números de teléfono, contactos		
Software	Android, 2.1, 2.2		
Internet	Ancho de banda, navegador, HTML		
Teléfono	Unidad, dispositivo, celular, Droid Incredible		
Carrier	Verizon, T.Mobile, AT&T		
Mensajería	SMS, chat		
Cuentas	Contraseña, código, licencia, autenticación		
Redes Sociales	Facebook, myspace, Twitter, YouTube		
Red de comunicación	3G, Wi.Fi, señal		
Audio	Sonido, volumen, voz, silencio, interferencia,		
Cámara	Auto foco, fotos, flash		
Video	TV, reproductor, grabador		
Música	Reproductor, canción,		
Apariencia Física	Tamaño, peso, color		
Conectividad	Sincronizar, conectar, enlazar		
Sensores	Acelerómetros, sensor de proximidad		

Tabla 3.2. Librerías desarrolladas

Cabe aclarar que el contenido total de las librerías es muy extenso para ser exhibido en el trabajo, por eso simplemente se muestran algunos conceptos que cada uno de los géneros contiene para que se entienda cómo éstos están compuestos.

3.2.2. Desarrollo de las reglas TLA

TLA es un acrónimo de Text Link Analysis (análisis de relación de texto). Estas reglas han sido descritas por el autor como patrones semánticos. El patrón buscará hacer juego entre dos o más conceptos diferentes, dentro de una cadena de texto.

Las reglas TLA relacionan conceptos de un género con conceptos de otro género diferente. El IBM SPSS Modeler permite relacionar hasta 6 géneros, aunque lo más usual es relacionar entre 2 y 3 géneros. La relación está basada en la proximidad, lo que implica que si los conceptos son próximos entre sí, entonces existe relación entre ellos. En este contexto el concepto de proximidad refiere a la cantidad de términos que separa a un concepto de otro. Quien diseña las reglas TLA es quien define la proximidad, basándose en una cantidad máxima de términos aceptables. Por ejemplo, se podría buscar una relación entre el género 'Música' de la librería de Características listada en la sección anterior con el género 'Positivo' de la librería de Investigación, estableciendo una proximidad de 3 términos. Esto implica que, si en un texto la separación entre cualquier concepto del género 'Música' y cualquier concepto del género 'Positivo' es menor o igual a 3 términos, entonces la regla lo identificará. Una ilustración tal vez permita entender mejor el ejemplo. Si un usuario dijera:

“El reproductor de música suena muy bien”

La regla antes mencionada funcionaría y detectaría este comentario por una simple razón: El concepto 'reproductor de música' pertenece al género 'Música'; el concepto 'Muy bien' pertenece a 'Positivo'; y la diferencia entre estos conceptos es solamente 1 término: 'suena'. En cambio, si el consumidor dijera:

“No entendía cómo funcionaba el reproductor de música, pero luego entendí y me pareció excelente”

En este caso la regla mencionada no funcionaría, y no detectaría este comentario positivo, puesto que la separación entre los conceptos es mayor a 3 términos: 'pero luego entendí y me pareció' suma 6 términos. Por lo tanto la regla considera que en este caso no existe relación entre los conceptos.

La calidad de las reglas TLA depende principalmente de dos factores.

- Primero, a mayor proximidad entre conceptos mayor precisión.
- Segundo, a mayor cantidad de géneros relacionados mayor precisión

La elección de los parámetros a usar depende del tipo de trabajo y la precisión que requiera. En este trabajo se relacionarán entre 2 y 3 géneros, y se aceptará una proximidad de 3 términos. Cabe destacar, que cuanto mayor sea la precisión exigida en la regla, mayor será la pérdida de objetivos. Es decir, serán más los resultados que no sean detectados, debido a la mayor precisión establecida. En el ejemplo se vio como el espaciado de 3 términos hizo que la regla fallara en capturar una frase objetivo.

El desarrollo de las reglas queda contenido dentro de la interfaz del IBM SPSS Modeler. Lo que interesa en este punto es entender su funcionamiento pues a partir de ellas se va a capturar el sentimiento y los problemas que el consumidor enfrenta. Básicamente se relacionan grupos de palabras. En el ejemplo dado, se relaciona el género 'Música' con el género 'Positivo', pero como ya se explicó en la sección 1.3.3 en esta tesis sólo interesan los rasgos negativos. Por lo tanto a través de las reglas TLA se buscarán relaciones entre los géneros de librería de Características y el género 'Negativo' de la librería de Investigación. A veces para establecer la relación se necesita la participación de un tercer género, que podría ser 'Ocurrencia' o 'Acción'.

3.2.3. Desarrollo de categorías

Habiendo desarrollado las librerías y las reglas, se procede al desarrollo de las categorías. Las categorías indican de qué trata el texto, se podría decir que son temas de interés. A priori se desconoce de qué trata el texto, pero al correr los la herramienta de MT, ésta clasificará cada comentario dentro de la categoría que corresponda. Lógicamente, también habrá comentarios sin categorizar, pues muchos de ellos tratan de asuntos que quedan fuera del alcance del objetivo a cumplir: capturar la voz del consumidor. Se recuerda que la mayor parte de la información proviene de foros, y es frecuente que en estos sitios las discusiones se desvirtúen de la temática inicial.

Siendo la voz del consumidor el tema de interés en este trabajo, cada categoría representará una dificultad que el consumidor enfrenta. La MT procesa la información indicando todos los problemas relevantes. Sin embargo es pertinente distinguir los diferentes tipos de problemas. A priori la información a analizar presenta las siguientes problemáticas:

1. Fallas en la operación del dispositivo.
2. Insatisfacción sobre alguna característica del teléfono.
3. Consultas sobre la operación del celular.
4. Consultas sobre el grado de satisfacción de otros usuarios.
5. Consultas sobre el precio y las condiciones de contrato.

Para el análisis se tomarán en cuenta únicamente las primeras dos problemáticas, puesto que el presente trabajo está orientado hacia la mejora de la calidad y el diseño del producto. Como se mencionó en la sección 1.1.2, cuando se entiende la voz del consumidor se define la voz del proceso. Las consultas no aportan valor para este propósito, aunque sí aportan valor dentro de otros sectores de la organización.

Teniendo en cuenta todo lo considerado en este capítulo hasta este punto, se construyeron las categorías que luego servirán para el análisis, listadas en la tabla 3.3.

Fallas		Insatisfacciones	
Pantalla	Sensor de proximidad	Sistema Operativo	Tamaño de la fuente
No enciende	Cámara	Vida útil de batería	Cámara
Teclado/botones	Tarjeta SIM	Calendario	Tamaño
Batería/carga	Música	Velocidad de respuesta	Pantalla
Audio	Conexión a PC	Contactos	Tarjeta de memoria
Señal	Linterna	SMS	Cargador de auto
No responde	Auto orientación	Teléfono muy complejo	Calidad de fotos y video
Contactos/agenda	Aplicaciones	Teléfono de la competencia preferido	Bluetooth
Se calienta	Vibrador	Red Social	Wi.Fi
Conexión a Internet	Arranque	Correo de voz	Carrier
Conexión a Wi.Fi	Widgets	Diseño	Precisión del GPS
Se apaga	Correo de voz	Email	Aplicaciones
SMS	Tapa de batería	Tapa de batería	Llamadas
Memoria	Headset	Teclado	Interfaz
Micrófono	Descarga	Wallpaper	Velocidad de Internet
Email	Red Social	Alarma	Accesorios
Llamadas	Calendario	Cobertura de red	
Bluetooth	Ringtones		
Se Reinicia solo	Cierre a fuerza		

Tabla 3.3. Categorías Desarrollados

La construcción de las categorías no es proceso lineal, sino más bien iterativo. Se tomó la misma muestra de 1000 comentarios que en la sección 3.2.1 y se trabajó sobre ellos. Al principio se parte de categorías más generales y luego se van partiendo esas principales en categorías más específicas. Se observa en la tabla que muchas categorías están repetidas, pero la concepción no es la misma para las dos. Una falla ocurre cuando el teléfono no funciona de acuerdo a especificaciones, en cambio una insatisfacción ocurre cuando el teléfono funciona de acuerdo a especificaciones pero por debajo de la expectativa del consumidor. Una cosa es “no funciona el teléfono” (falla) y otra cosa es “no me gusta el teléfono” (insatisfacción). La repetición de los nombres de las categorías alude a estos dos puntos de vista. Por ejemplo, la imposibilidad de enviar un email es una falla, en cambio la incomodidad del usuario al usar la aplicación que envía emails es una insatisfacción. Por eso la categoría Email está repetida.

Cada categoría está compuesta por una serie de descriptores, que en su mayoría son reglas TLA. Entonces, recapitulando el trabajo de la sección 3.2, se partió de palabras simples (conceptos), las cuales fueron clasificadas por género dentro de una librería, éstos fueron relacionados con reglas TLA, que a su vez forman parte de una categoría: el problema que se desea identificar. Para entender mejor la relación, la figura 3.3 es una impresión de pantalla sacada del IBM SPSS Modeler. La impresión muestra la categoría “Vida útil de batería” junto a sus descriptores.

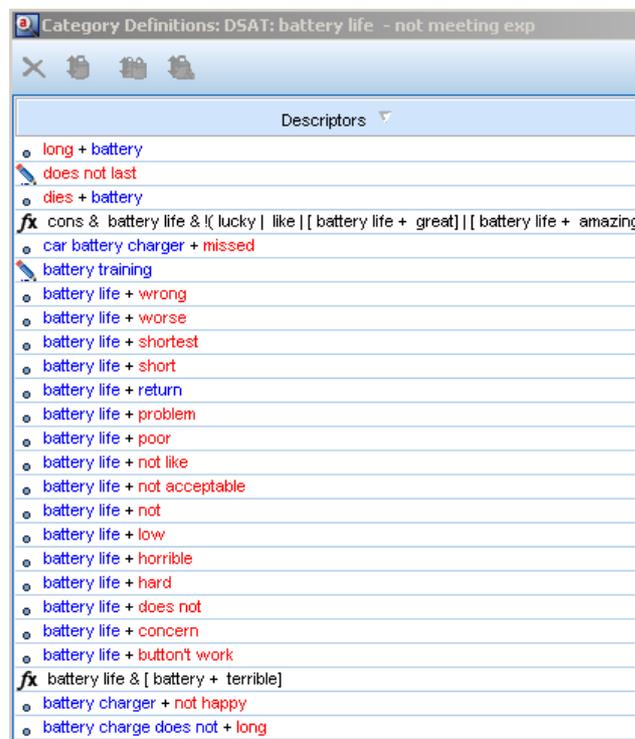


Figura 3.3. Captura de la categoría “Vida útil de batería”

Se observan 3 tipos de descriptores dentro de la categoría:

- Precedidos de un punto: Son reglas TLA. En ellas se puede ver los conceptos que la regla ha relacionado. Todas éstas surgen de relacionar el género “Batería” con el género “Negativo”.
- Precedidos de un lápiz: Son simplemente conceptos.
- Precedidos de Fx: Es una regla lógica en la que se pueden combinar conceptos y reglas TLA. El símbolo “&” representa la conjunción, el símbolo “|” representa la disyunción y el símbolo “!” representa la negación. En la figura 3.3 se puede ver una regla lógica que captura comentarios con el concepto “contra” y el concepto “Vida útil de batería” exigiendo que en ese comentario no exista, por ejemplo, [Vida útil de batería + genial] (Una regla TLA). La regla lógica no exige relaciones entre palabras, como en las TLA. Solo exige que un comentario contenga (o no) ciertos conceptos y/o reglas TLA.

Las diferencias entre categoría, regla, género y concepto son difíciles de percibir. Lo importante es que la categoría representa un problema, y todo lo demás es un conjunto semántico que dará significado a esa categoría. Habiendo entendido esta distinción ahora es oportuno explicar cómo funcionan las categorías. Para ese propósito se hace referencia al diagrama de flujo de la figura 3.4

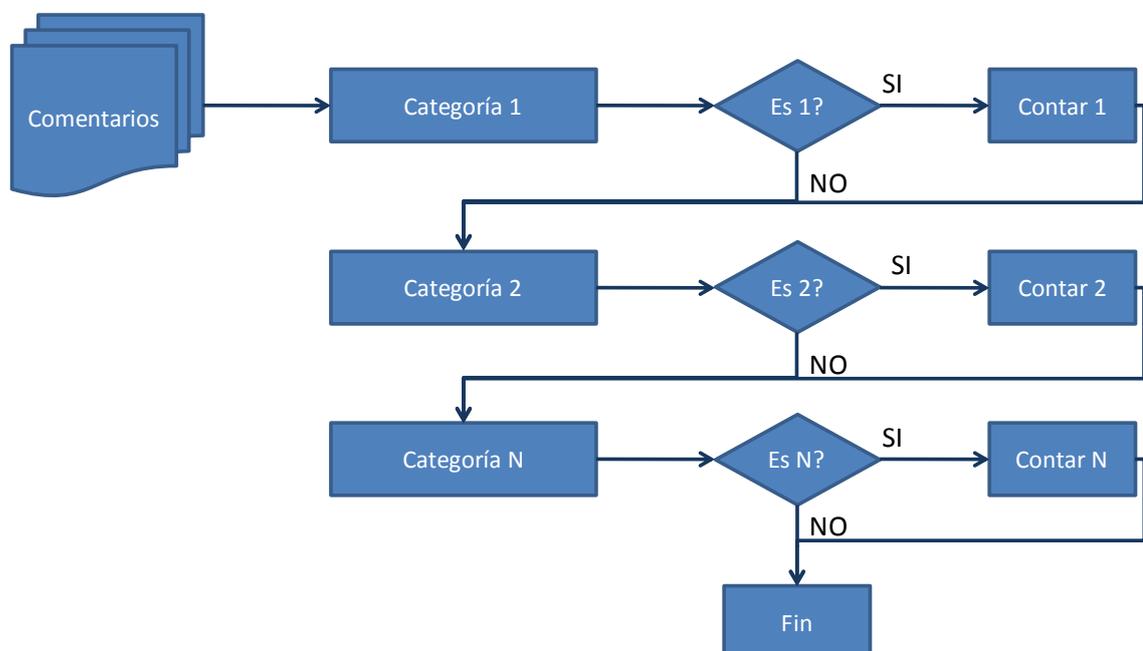


Figura 3.4. Flujo de la información a través de las categorías

Cada comentario pasará a través de los descriptores de todas las categorías. Si el comentario hace juego con alguno de los descriptores entonces pertenece

a esa categoría y se contabilizará dentro de la misma. Como se ve en el diagrama de la figura 3.4 un mismo comentario puede pertenecer a más de una categoría. Al mismo tiempo puede suceder que un comentario no pertenezca a ninguna categoría (Si siempre responde “No” en el diagrama) y quedará sin categorizar. El resultado final es un conteo por categoría y un conteo por información sin categorizar. Para este trabajo eso significa un conteo por problema y de esta manera se podrá saber cuáles son los principales inconvenientes que el consumidor padece. Se recuerda que los comentarios fueron agrupados por usuario como se explicó en la sección 2.3.2.

3.3. PRODUCTIVIDAD Y MEJORA CONTINUA

3.3.1. Eficacia, eficiencia y productividad

La gran ventaja de la MT radica en la eficiencia. La MT es capaz de procesar y clasificar las opiniones de 1000 consumidores en 1 minuto. Web Scraping extrae esa información en 29 minutos. Esto significa que la acción “procesar y clasificar las opiniones de 1000 consumidores” requiere de 1 individuo, una computadora, y 30 minutos. Lo que da una velocidad de procesamiento de 1,8 segundos por consumidor. Si esto se compara con otras soluciones se observa a simple vista que la diferencia es abismal. Por ejemplo, según Benchmark portal, la velocidad de procesamiento de llamados en un call center es de 6 minutos por cliente. Al mismo tiempo, si se desea realizar un focus group o una encuesta, esto requerirá de mucho mayor tiempo, siempre del orden de los minutos para cada consumidor. Con la MT la eficiencia y la productividad aumentan considerablemente, más aún si el trabajo implica tener que leer la información. La MT es una solución para una organización que cuenta con información de texto pero no cuenta con el tiempo para leerla. Por supuesto, estas comparaciones se refieren a las operaciones en producción, no en desarrollo. Así como en este capítulo se explicó el desarrollo semántico necesario para la MT, un focus group o un call center también requiere de una etapa de construcción antes de estar en funcionamiento.

Toda esta velocidad de procesamiento de la información tiene un costo: la eficacia. Como se dijo en la sección 1.5.5, la MT no es una ciencia exacta. La precisión de los resultados no es del 100%. Se puede trabajar para mejorar la precisión pero nunca será perfecta. Ésta es una desventaja ante los otros métodos antes mencionados. Todo se resume a qué nivel de precisión la organización está dispuesta a aceptar. A priori, la semántica desarrollada tiene una precisión del 70%, esto significa que clasificará 7 comentarios bien y 3 mal. La siguiente sección explica cómo trabajar para mejorar esa precisión y tener una MT más robusta.

3.3.2. Calidad de la MT y mejora continua

Cuando se habla de calidad en MT, se hace referencia principalmente a dos parámetros. En primer lugar, la precisión. En la sección 3.3.1 se indicó que a priori se tiene un 70% de precisión. En este proyecto se hizo un trabajo adicional para que la precisión sea aproximadamente del 80%. Esto es, 8 comentarios bien clasificados, 2 comentarios mal clasificados.

En segundo lugar, la potencia de la MT. Eso se mide en función de la información sin categorizar. Como se explicó en la sección 3.2.3, siempre habrá información que quedará sin categorizar (temas fuera del interés). Generalmente esto es algo esperable en un proceso de MT. El problema es que puede suceder que un comentario objetivo no sea capturado por ninguna de las categorías. Es decir que no existe un descriptor que sea capaz de hacer juego en ese texto. En este trabajo la potencia es del 95%, es decir de 100 comentarios sin categorizar, 5 deberían haber sido categorizados.

Estos parámetros, sin embargo, pueden mejorar en el tiempo. Esto siempre y cuando se use la MT en producción: procesamiento continuo de la información. Este puede ser el caso de una organización donde se requiera semanalmente un análisis del rendimiento del producto en el mercado. Cada cierto período de tiempo (por ejemplo, mensualmente o semanalmente) se puede hacer un mantenimiento, en el cual se mejoran los descriptores que capturan malos comentarios y se agregan nuevos descriptores para capturar comentarios objetivos que no fueron categorizados. Esto hace que la calidad y robustez del proceso mejore en el tiempo. El diagrama de la Figura muestra este proceso de mejora continua.

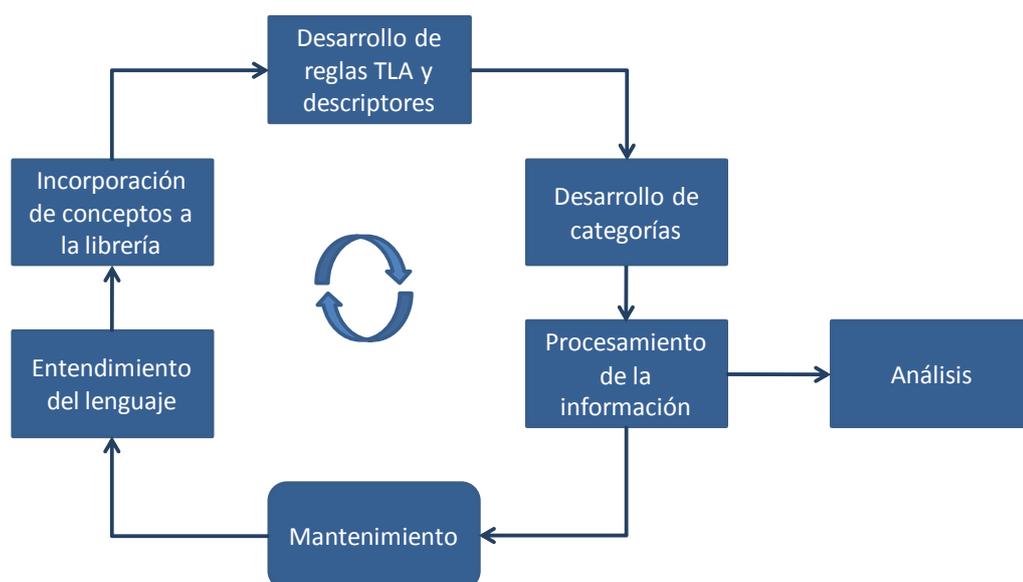


Figura 3.5. Proceso de mejora continua en MT

4. ANÁLISIS DE LA INFORMACIÓN

En este capítulo se presentarán los resultados de la aplicación de MT sobre la información extraída de la Web. Como se planteó en los objetivos, se pretende capturar la voz del consumidor, por lo tanto se hará clara identificación de los principales problemas que el consumidor enfrenta cuando usa su teléfono, el HTC Droid Incredible. Por un lado se tienen las fallas, por otro lado la insatisfacción. En la segunda parte de este capítulo se profundizará sobre una de las problemáticas expuestas en primera instancia, para entender en detalle los síntomas que describen el problema. La Figura 4.1 indica en qué punto del trabajo se sitúa este capítulo.

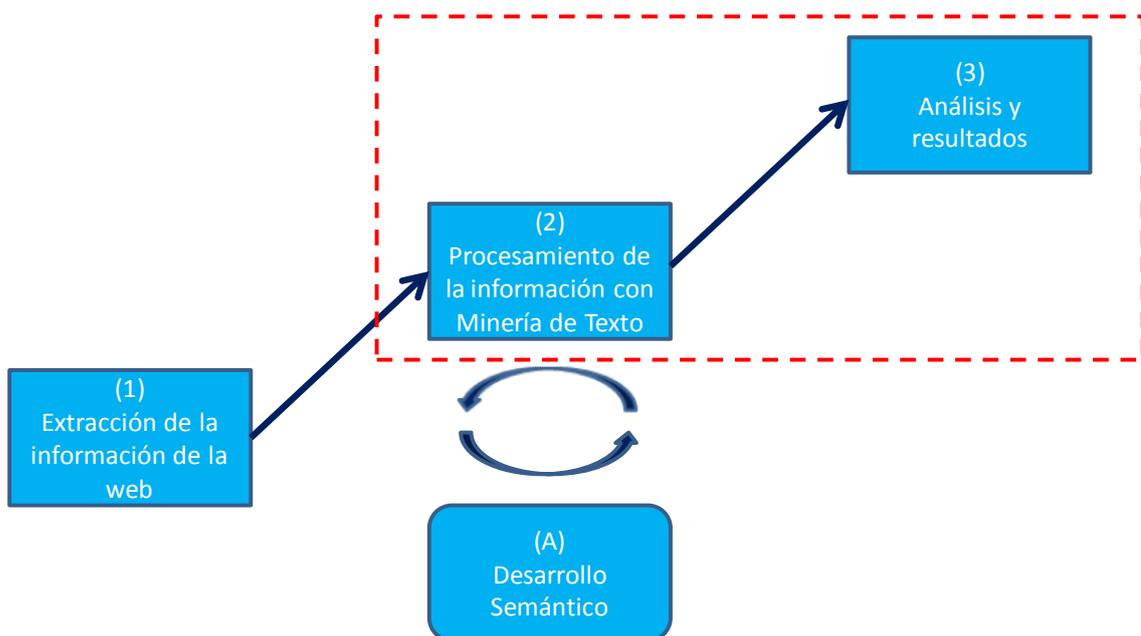


Figura 4.1. Localización del capítulo 4 en el proceso

4.1. VOZ DEL CONSUMIDOR SOBRE EL PRODUCTO

Usando la MT desarrollada en el capítulo anterior se procesó la información con el IBM SPSS Modeler sobre un total de 11022 usuarios que comentaron en 24 sitios de la web sobre el Droid Incredible, entre el 21 de abril de 2010 y el 21 de Septiembre de 2010. De esos 11022, 1986 encontraron una falla en su teléfono y 3264 expresaron su insatisfacción, representando el 18% y el 29,6% del total. Puede ocurrir que alguno de estos usuarios haya indicado tanto una falla como una insatisfacción. El diagrama de Venn de la figura 4.2 puede clarificar esta particularidad.

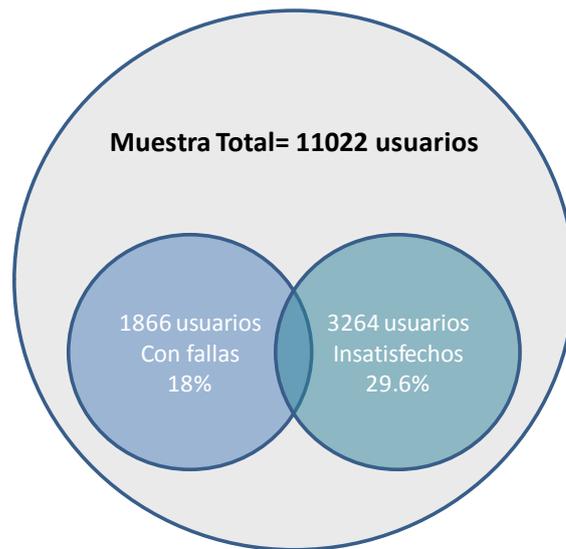


Figura 4.2. Diagrama de Venn sobre la muestra total

La conjunción del conjunto de usuarios con fallas y el conjunto de usuarios insatisfechos reúne 718 individuos, lo que implica que la cantidad total de consumidores experimentando algún tipo de problema (falla, insatisfacción o ambas) es de 4412, el 40% de la muestra total. A priori esta suma puede parecer un demasiado alta, sin embargo es muy coherente con los resultados esperados. En un estudio basado en la voz del consumidor, se parte de la hipótesis que 67% de los comentarios serán positivos, y el 33% de ellos serán negativos. Esta afirmación proviene de la experiencia del autor en tema bajo estudio. De acuerdo a esta postura, el resultado se encuentra cercano al valor esperado pero tal vez sea un tanto elevado.

4.1.1. Fallas

Ante todo, se recuerda lo dicho en la sección 3.2.3, una falla ocurre cuando el producto no funciona de acuerdo a especificaciones. Siguiendo esta concepción en la figura 4.3 se presenta un diagrama de Pareto con las principales fallas del Droid Increíble según la muestra considerada. Los descriptores de estas 10 categorías se encuentran en el Anexo II del presente trabajo.

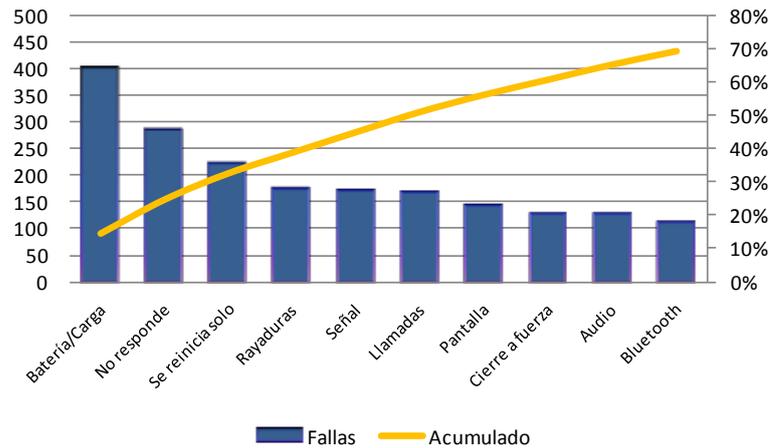


Figura 4.3. Principales fallas encontradas

La observación de este diagrama de Pareto, permite distinguir dónde se localizan el 70% de las fallas, pero no queda en claro cuál es la falla que se produce. A continuación se dan los detalles para cada una de ellas.

1. **Batería/Carga.** El cargador no carga adecuadamente su batería, o la batería no logra retener la carga.
2. **No responde.** El teléfono no responde a las órdenes del usuario, está congelado, tildado o bloqueado.
3. **Se reinicia solo.** El celular se reinicia sin haber recibido esta orden del usuario.
4. **Rajaduras.** Se han generado rajaduras o grietas en el producto, sobre todo en la pantalla.
5. **Señal.** Ausencia de cobertura de red que imposibilita la comunicación.
6. **Llamadas.** El consumidor no puede realizar llamadas, no puede recibirlas o se corta la comunicación en medio de una conversación.
7. **Pantalla.** Principalmente se acumula polvo debajo de la pantalla. Por otro lado, ocasionalmente la pantalla se torna blanca.
8. **Cierra a fuerza.** Al querer iniciar una aplicación el usuario recibe un mensaje de error diciendo que la aplicación no se puede iniciar y deberá ser cerrada.
9. **Audio.** Durante una llamada, el usuario no puede escuchar al receptor o viceversa. También se perciben sonidos extraños como silbidos o estática.
10. **Bluetooth.** No funcionan los comandos de voz al usar Bluetooth.

Examinando este desglose, se distinguen distintos tipos de fallas. En primer lugar, software y hardware, las más comunes por supuesto. A priori es difícil determinar si una falla es de software o de hardware. Esto requiere trabajar sobre el producto, lo cual escapa a los conocimientos técnicos y a las posibilidades del autor. La principal falla de este tipo es la carga, encontrada en

398 usuarios. Por otro lado se identificó una falla física: la poca resistencia del producto hace que se generen rajaduras. Finalmente se advierte una falla de diseño en la pantalla, no debería acumularse polvo debajo de ésta.

Un caso que llama la atención es el del sistema de comunicación. Se observan fallas en la señal, en los llamados, y en el audio durante los llamados. Al parecer la comunicación telefónica tradicional es uno de los puntos débiles de este producto. Una ironía, ya que supuestamente ésta es su principal función. Esto alude a una vieja ley sobre construcción de máquinas: A menor cantidad de funciones, mejor funcionamiento de cada una de ellas. Es posible que todos estos problemas estén causados por la falta de señal, aunque es imposible de determinar con este estudio.

La identificación de las causas raíces que generan estas fallas queda fuera del alcance de este trabajo. Precisamente, el objetivo de este trabajo es encontrar el resultado que funcionará como entrada al proceso de detección de causas raíces. Con este resumen trabajaría el equipo de ingeniería de procesos para aplicar las modificaciones necesarias que corregirían estos problemas.

4.1.2. Insatisfacción

A diferencia de la falla, la insatisfacción ocurre cuando el teléfono funciona de acuerdo a especificación pero no está a la altura de las expectativas del consumidor. En la figura 4.4 se encuentran las principales causas de insatisfacción.

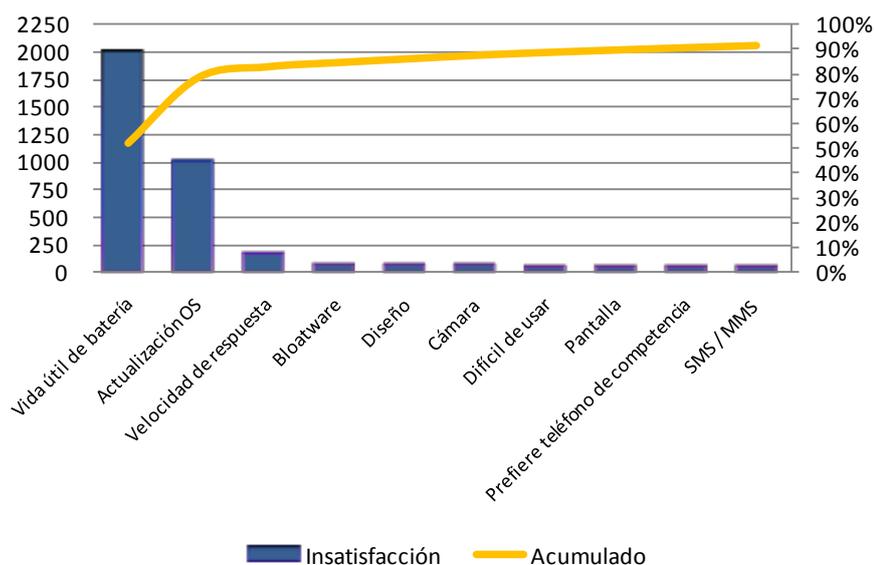


Figura 4.4. Principales insatisfacciones encontradas

Los descriptores de estas 10 categorías se encuentran en el Anexo II de este documento. A continuación se describe en detalle qué significa cada uno de estos problemas.

1. **Vida útil de batería.** La batería no dura suficiente.
2. **Actualización OS.** El consumidor se frustra a la espera de una actualización que solucione su problema, o simplemente pretende tener la versión más reciente de Android. Algunos experimentan nuevos problemas luego de una actualización.
3. **Velocidad de respuesta.** La Interfaz es lenta desde el punto de vista del usuario.
4. **Bloatware.** Una característica que irrita a los consumidores porque no permite personalizar el aparato a gusto (por ejemplo, no permite borrar aplicaciones preinstaladas).
5. **Diseño.** Al usuario no le gustan algunos aspectos de diseño, como la forma y la disposición física de los botones.
6. **Cámara.** Algunas fotos salen borroncadas y el celular no tiene botón para sacar fotos (sólo se puede dar esta orden desde la pantalla táctil)
7. **Difícil de usar.** El teléfono es demasiado complejo y no es amigable para su uso.
8. **Pantalla.** La pantalla es un poco chica para el consumidor y por otro lado es difícil mirarla bajo el sol.
9. **Prefiere teléfono de la competencia.** El usuario considera que el teléfono no está a la altura de teléfonos de la competencia.
10. **SMS/MMS.** Escribir los mensajes se torna dificultoso.

Queda a las claras que la principal insatisfacción es con la vida útil de la batería. 1979 usuarios dijeron estar descontentos con esta característica, representando un 18% de la muestra total, y el 52% de la insatisfacción. No debe confundirse con la principal falla, relacionada a la carga de la batería. Una cosa es una carga incorrecta, otra cosa es una duración por debajo de las expectativas. Sin embargo los principales problemas están relacionados a la batería, lo cual debe ser una señal de alarma para HTC si pretende mejorar la calidad de sus productos.

La poca vida útil de batería es un problema común para todo teléfono con pantalla táctil, aunque nunca en una forma tan desproporcionada como ésta. Los usuarios que no lo toleran siempre optarán por BlackBerry, ya que la mayoría de sus teléfonos no cuenta con pantalla táctil, sino con un teclado físico, lo cual permite una mayor vida útil de la batería. Justamente, la mayoría de los usuarios de la categoría “Prefiere teléfono de la competencia”, prefirieron a BlackBerry. En segundo lugar se encuentra iPhone y luego otras marcas.

4.1.3. Actualización del sistema operativo

Un párrafo aparte merece la actualización del sistema operativo, pues es un punto muy importante. Con la sofisticación de estos aparatos, una actualización permitirá resolver muchos de los problemas que el consumidor enfrenta. Luego del lanzamiento de los productos, la organización recopila información del mercado para saber cuáles son los problemas más relevantes. Modificando el sistema operativo, muchos de estos problemas pueden ser mejorados e incluso corregidos. La actualización la hace el carrier a través del medio OTA (“Over The Air”), es decir por el aire. Haciendo uso de la red 3G del carrier se distribuye la última versión del sistema operativo a los consumidores. Así logran mejorar la calidad del producto. Lógicamente, la nueva versión puede originar nuevos problemas, los cuales pueden ser corregidos con posteriores actualizaciones. Por otro lado, el nuevo sistema operativo trae consigo nuevas funcionalidades que atraen al usuario. El consumidor conoce este sistema de mejora, por eso espera y reclama por una actualización para resolver su problema o simplemente para mejorar el rendimiento de su teléfono. La figura 4.5 hace referencia a este sistema de mejoramiento de la calidad.

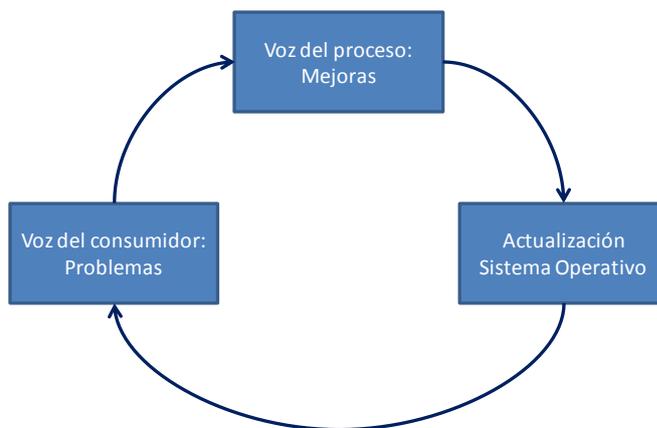


Figura 4.5. Mejoramiento de la calidad del producto

El presente trabajo se enmarca dentro de este proceso que se muestra en la figura 4.5 a través de la investigación de la voz del consumidor. Este sistema actualmente funciona en las empresas que fabrican celulares, como HTC, Apple, Motorola, BlackBerry, y otras; y además es de conocimiento público. Este proyecto plantea el uso de la MT como un método eficiente para analizar información desestructurada y contribuir a la organización en la mejora de los productos.

La actualización puede ser de dos tipos: puede ser un mantenimiento o puede ser un salto de versión en el sistema operativo. El mantenimiento tiene el objetivo de eliminar errores, en cambio el salto de versión trae consigo nuevas funcionalidades y potenciamiento del rendimiento general. El 27 de agosto HTC hizo una actualización del sistema operativo para el Droid Incredible, donde los usuarios subieron de Android 2.1 a Android 2.2. La versión 2.2 traía consigo las siguientes mejoras y funcionalidades:

1. Optimización de la velocidad del sistema operativo, la memoria y el rendimiento.
2. Mejoramiento de la velocidad de las aplicaciones gracias a implementación de JIT.
3. Integración del motor javascript V8 de Chrome en el Navegador.
4. Aumento del soporte con Microsoft Exchange.
5. Mejora del lanzador de aplicaciones con accesos directos a las aplicaciones del teléfono y navegador.
6. Anclaje a red con USB y funcionalidad hotspot con Wi-Fi.
7. Opción para deshabilitar el acceso de datos.
8. Versión actualizada de la aplicación Market.
9. Rápido cambio entre múltiples lenguajes de teclado.
10. Comandos de voz y transferencia de contactos con Bluetooth.
11. Soporte para contraseñas numéricas y alfanuméricas.
12. Soporte para subir archivos con el Navegador.
13. Soporte para mostrar GIFs animados en el Navegador.
14. Soporte para instalar aplicaciones a la memoria expandible.
15. Soporte para Adobe Flash 10.1

Llama la atención el punto 10 de esta lista, pues uno de los problemas encontrados indicaba la imposibilidad de emplear comandos de voz con Bluetooth. ¿Se habrá resuelto realmente el problema? La figura 4.6 muestra la evolución en el tiempo de la discusión sobre este problema.

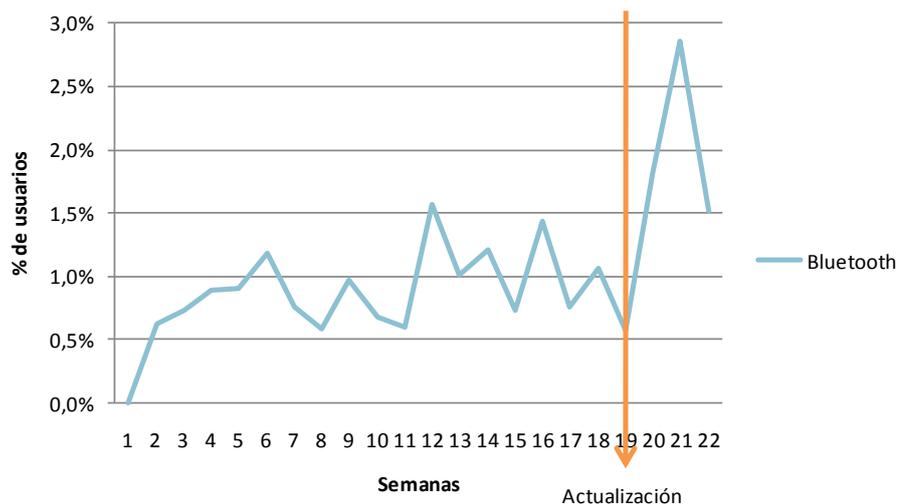


Figura 4.6. Evolución temporal de discusión sobre bluetooth

Un buen indicador es la cantidad de gente comentando sobre este tema antes de la actualización y después de la actualización. En el gráfico el eje vertical representa la cantidad de usuarios experimentando el problema sobre la cantidad total de usuarios comentando sobre el producto.

En la semana 19 luego del lanzamiento del producto, se produjo la actualización y curiosamente en ese punto se eleva bruscamente la discusión sobre el bluetooth. Esto puede llevar a la interpretación de que el problema no se ha solucionado. La realidad es que falta información para determinar si el problema ha sido resuelto o no. De todas formas, hay que tener cuidado al interpretar la figura 4.6. De hecho, este comportamiento es el esperado: no es casualidad que justo luego de la actualización aumente la discusión así.

Inicialmente la gente toma conciencia sobre la existencia del problema, por lo que la discusión aumenta hasta la semana 12. Luego la gente toma conciencia sobre la existencia de la solución: una actualización de sistema operativo que permitirá utilizar los comandos de voz con Bluetooth. Así, la discusión disminuye hasta la semana 19, en la cual se produce la actualización. Lo que ocurre es que el proceso de actualización no se hace en un solo día. En este caso comienza en la semana 19 (27 de Agosto), y se envía la nueva versión a los usuarios por tandas, por ejemplo a 10000/día. Lo normal es que este proceso dure unas 2 semanas. Durante este tiempo la discusión aumenta pues luego del 27 de agosto muchos usuarios observan que su problema persiste. Al cabo de dos semanas la mayoría ya tiene la actualización y ve que el problema está resuelto. Justamente luego de la semana 21 la discusión baja repentinamente. Como se dijo en el párrafo anterior, se necesitan más datos para sacar una conclusión definitiva, pero la tendencia final parece indicar que el problema estaría resuelto.

4.1.4. Comparación con la competencia

En este punto se comparará al Droid Incredible con los competidores que se han elegido. El propósito es saber si los problemas son sistemáticos a través de los productos, o si pertenecen exclusivamente al Droid Incredible. El primero de los casos pone a HTC en condiciones de competencia dentro del mercado. El segundo caso representa una debilidad y se deberá prestar especial atención a aquellos problemas.

Para que la comparación sea justa, se debe utilizar la misma MT, a excepción de dos categorías. En primer lugar los otros dispositivos funcionan con otro sistema operativo. Habrá que agregar los descriptores correspondientes en esta categoría. En segundo lugar “Prefiere teléfono de la competencia” no tendrá lugar en la comparación, pues la concepción de esta categoría no aplica sobre productos de la competencia.

Antes de ver la comparación es preciso hacer una aclaración. En estos gráficos de Pareto se exhiben los problemas encontrados únicamente en el Incredible. Si la competencia no los tiene o los experimenta en menor medida, no significa que la competencia tiene un mejor rendimiento general, sino un mejor rendimiento sobre esos problemas específicos. La competencia podría tener otros inconvenientes que no serán analizados en este trabajo. En primer término, la figura 4.7 muestra la comparación entre las fallas.

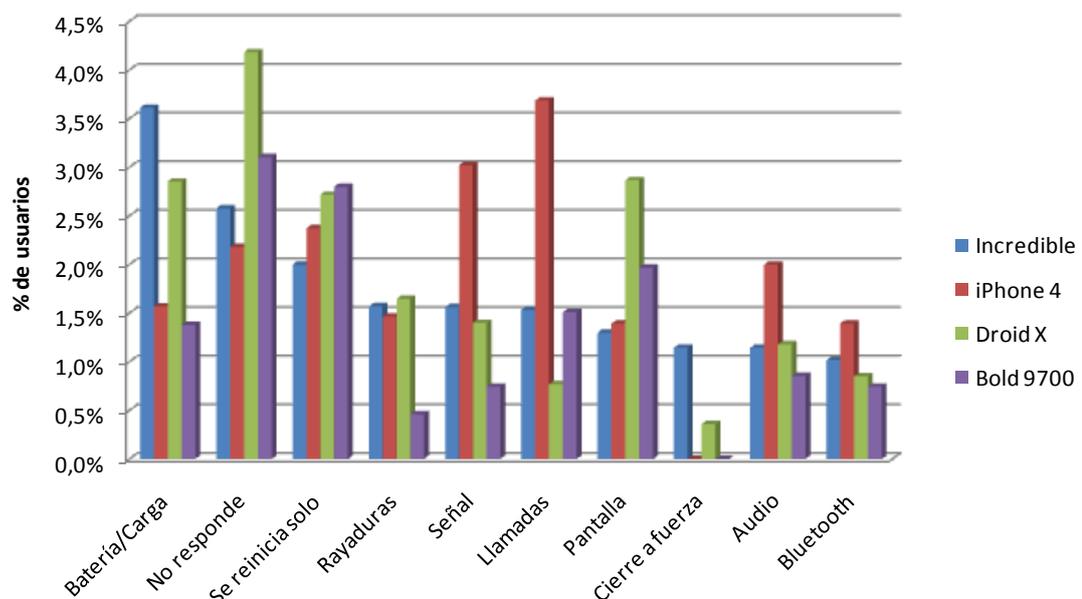


Figura 4.7. Comparación de fallas

En este diagrama, el eje vertical representa el porcentaje de usuarios que experimentó la falla sobre el total de usuarios comentando. Como ya se mencionó, para el Droid Incredible se tienen 11022 consumidores opinando dentro del período considerado. Para ese mismo período, se tienen 16522 usuarios de iPhone 4, 7294 de Droid X y 4576 de Bold 9700.

Se advierte en la figura 4.7 que la mayoría de las fallas de Droid Incredible son recurrentes en la competencia. De hecho, la competencia tiene incluso peores resultados en 8 categorías. Está claro que el Droid Incredible tiene serios problemas con su batería, principal falla y con mayor recurrencia que la competencia. Por otro lado es curioso el caso del “Cierre a Fuerza”, la octava categoría en el Pareto. Ni iPhone ni BlackBerry exhiben este problema, pero sí el Droid X. De ahí se desprende la idea de que es un problema del sistema operativo Android, el cuál es común a ambos teléfonos.

Cuando se analizaron las fallas en la sección 4.1.1 se comentó sobre el problema en el sistema de comunicación del Droid Incredible. Sin embargo, según el gráfico iPhone tiene un rendimiento más pobre en este aspecto, saliendo último en “Señal”, “Llamados” y “Audio”. Esta conclusión coincide con la historia real del aparato. Cuando salió el iPhone 4 a fines de junio de 2010 rápidamente sus usuarios experimentaron un grave problema: si el teléfono era sostenido en cierta forma la antena quedaba bloqueada, lo cual provocaba pérdida de señal. El CEO de Apple, Steve Jobs prometió a sus clientes regalar un estuche para evitar este problema. Fue un caso muy conocido que dio mala prensa a la empresa. La escandalosa historia, si bien es preocupante para Apple, da coherencia al análisis y robustez a la MT desarrollada.

Por último el Droid X perdió en “No responde” y en “Pantalla” y el Bold 9700 en la categoría “Se reinicia solo”. En cuanto a la categoría “Rayaduras”, se observa un aparente empate entre los 3 teléfonos con pantalla táctil y ventaja del BlackBerry. Esto tiene sentido, pues la pantalla táctil debe enfrentarse constantemente a la presión de los dedos, los cuales pueden estar sucios o la pantalla misma con polvo, generando rajaduras al deslizarse los dedos. El Bold 9700 cuenta con un teclado Qwerty, y la operación del teléfono es mediante la presión de sus teclas.

Si bien las fallas del Incredible aparentan ser comunes a este tipo de teléfonos, HTC deberá trabajar si pretende diferenciarse de la competencia. El problema del “Cierra a fuerza” recae mayormente sobre Google, quien diseña el sistema operativo. Se recuerda que Android es un sistema relativamente nuevo, y todavía necesita ser pulido, mientras que BlackBerry e iPhone operan hace un tiempo mayor. Como se verá a continuación las debilidades de HTC se

concentran en torno a la insatisfacción. En la figura 4.8 se tiene la comparación desde esta perspectiva.

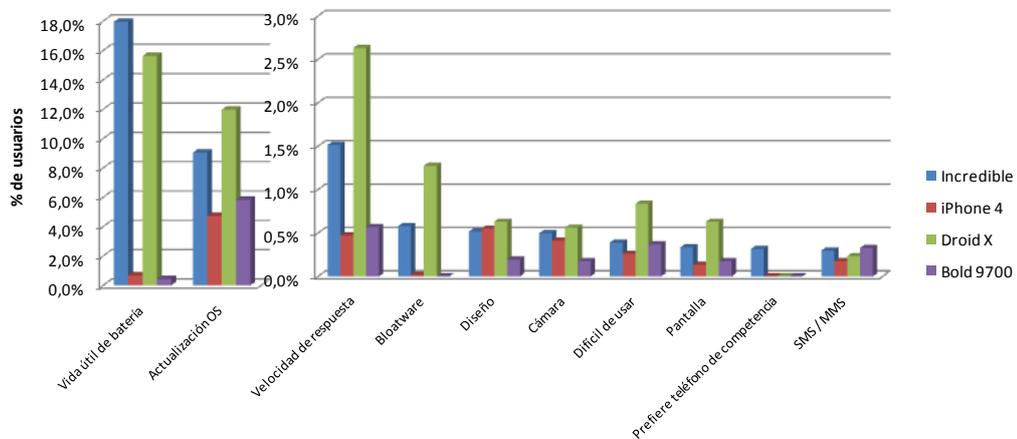


Figura 4.8. Comparación de insatisfacción

En general se observa que el Droid X tiene un rendimiento peor en casi todos los aspectos. Sin embargo, el Droid X es como un primo del Incredible. Si bien es de Motorola, poco a poco la comparación de los usuarios pasa por el sistema operativo. Por lo tanto hay que tener cuidado interpretando este gráfico. El Droid X puede servir como una referencia, siendo un producto de similar funcionalidad, pero HTC debería prestar mayor atención a lo que sucede con iPhone y BlackBerry. Con esta postura, se observa una gran falencia en la Vida útil de la batería, y debilidades en la velocidad de respuesta y la existencia del Bloatware. En los demás aspectos el rendimiento es similar y la incidencia es menor al 0,5%, por lo que su importancia relativa los minimiza. El caso de la Actualización del sistema operativo es anecdótico, pues consiste básicamente de usuarios reclamando por la última versión disponible. La diferencia de porcentaje habla del grado de tolerancia que tienen los consumidores hacia la marca.

El gran desafío para HTC gira en torno a la batería. Fallas en la carga y una vida útil que genera una considerable disconformidad en los consumidores, donde el 18% de ellos no está contento con esta característica. La vida útil de la batería es algo difícil de medir, depende del uso que cada usuario le dé al celular. Una medida muy común es el tiempo en llamada, es decir cuánto dura la batería usando el teléfono únicamente para hablar. En el gráfico de barras de la figura 4.9 se compara este indicador bajo dos fuentes distintas. Por un lado, la especificación y por otro lado un estudio realizado por Cnet.

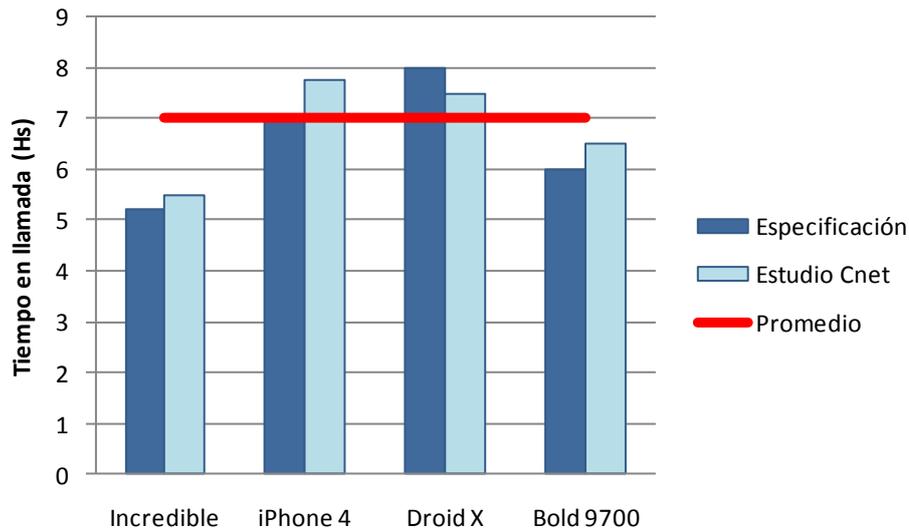


Figura 4.9. Vida útil de las baterías

El estudio de Cnet consiste simplemente en probar el teléfono y medir el tiempo en llamada. El promedio que se indica no corresponde a los cuatro productos, sino que es un promedio global entre celulares. El Incredible tiene el peor rendimiento, encontrándose un 21,4% debajo del promedio. Esto explica el descontento de los consumidores, pero al mismo tiempo alza una pregunta: Si HTC especifica la reducida duración de su batería, ¿por qué sus clientes luego se quejan de esta característica? Es como comprarse una camisa roja para luego decir “A mí no me gusta el color rojo”. Si un atributo no satisface al individuo lo lógico sería no comprar un producto con ese atributo. De cualquier manera, el trabajo de una organización no pasa por demostrar tener razón o que los clientes actúan irracionalmente, sino que el trabajo debe apuntar a la satisfacción de los consumidores. Si HTC pretende mejorar el grado de satisfacción deberá fabricar productos con mayor vida útil de batería.

En la sección 4.2 se profundizará sobre el tema de la velocidad de respuesta, la tercera categoría en el Pareto.

4.2. PROFUNDIZACIÓN DE UNA PROBLEMÁTICA

En esta sección se profundizará sobre la velocidad de respuesta del Droid Incredible. El objetivo es la identificación de los síntomas para que los ingenieros de proceso puedan interpretarlos. Así como un médico puede determinar la enfermedad a partir de los síntomas de un paciente, el ingeniero de procesos puede identificar la causa raíz del problema a partir de los síntomas que exhibe el producto.

Para lograr este fin se deberá repetir el mismo proceso de MT desarrollado en el capítulo 3, pero a un nivel micro. Los comentarios son menos, los conceptos son menos, las categorías serán menos, etc. El procedimiento es el mismo pero la escala y duración del trabajo es muchísimo menor. Mientras el proceso del capítulo 3 implica meses de trabajo, el de esta sección implica días.

4.2.1. Contexto

El crecimiento de Android desde el lanzamiento de su primer sistema operativo a comienzos de 2009 ha sido abismal, a tal punto que se ha convertido en uno de los principales competidores del mercado. La figura 3.2 dejó en evidencia esta afirmación, y muestra a BlackBerry como el principal competidor del mercado. Sin embargo la tendencia indica que BlackBerry está siendo desplazado, dejando a iPhone y Android como ganadores de la pelea (aunque sea por el momento). La figura 4.10 muestra la lealtad de los consumidores hacia estos productos

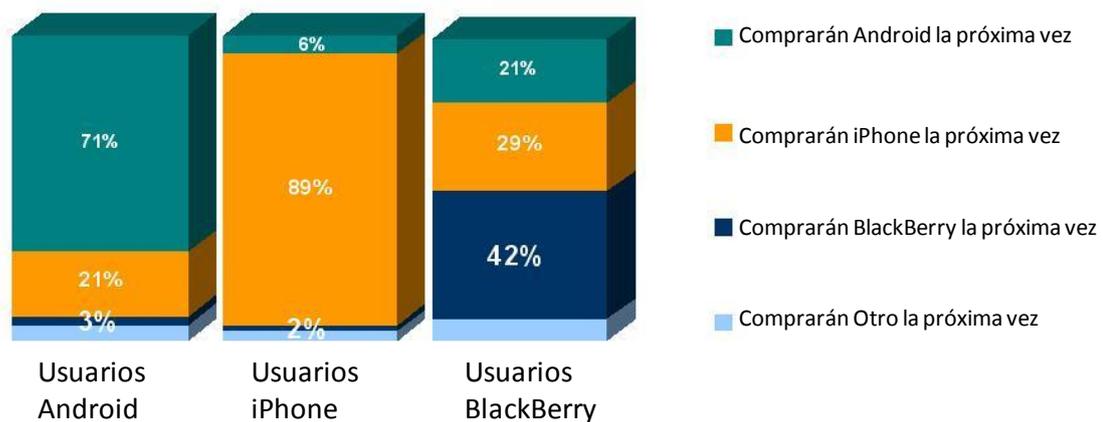


Figura 4.10. Lealtad de los consumidores (Fuente: Nielsen)

Se ve que en el futuro muchos de los usuarios de BlackBerry estarán migrando a iPhone o Android. Por otro lado, algunos de los consumidores de Android estarán cambiando a iPhone, el cual muestra la mayor lealtad. Una lealtad basada en la gran calidad (demostrada) de estos productos, pero también en un pre-concepto de que iPhone es lo único y lo mejor. Sea cierto o no este pre-concepto, el gran desafío de Android si quiere finalmente triunfar en el mercado es eliminando esta idea y lograr robar clientes de Apple. Siendo socio de Android, HTC deberá contribuir con esta tarea.

Los típicos atributos que los usuarios de iPhone destacan son: Alta calidad, Mejor tecnología, Interfaz, Precisión, Elegancia, Operación fluida de la pantalla táctil, Experiencia de navegación³. Una de las grandes ventajas del iPhone reside en la respuesta que tiene su pantalla táctil y la velocidad y simpleza de su Interfaz. Basta con unos pocos segundos de uso para advertir que existe una diferencia. No es una diferencia significativa, pero existe y es notoria. En oposición, una de las principales causas de insatisfacción en el Droid Incredible es justamente la lenta interfaz. Si bien no es significativa ante la principal razón (vida útil de la batería), es comparable a la incidencia de las fallas en la figura 4.3. Por otro lado resulta ser una característica deseable para el usuario y trabajar este punto podría mejorar la posición de Android con respecto a iPhone. Esto significa que la velocidad de respuesta tal vez no sea un problema actual, pero para el consumidor estaría dentro de las características de “desempeño y deleite”. Para entender mejor estos conceptos será de gran utilidad analizar el diagrama de Kano de la figura 4.11.

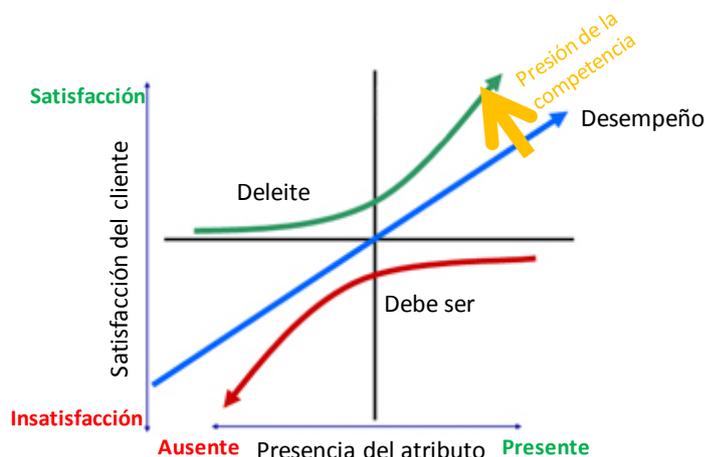


Figura 4.11. Diagrama de Kano

La curva de *deleite* representa aquellos atributos que harán muy feliz al consumidor pero si no estuvieran presentes comprarían el producto de todas formas. La curva de *desempeño* refiere al rendimiento del producto sobre esa característica, cuanto mejor sea el desempeño más contento estará el cliente. Por último, la curva del *Debe ser* representa a aquellos atributos que el consumidor da por sentado que existen en el producto, y si no estuvieran sería inaceptable para el usuario. Por ejemplo, una habitación de hotel que no tenga camas.

³ Theiphoneaddict.com. Why is the iPhone so popular?

Reiterando, la velocidad de respuesta pertenece a las curvas de *desempeño* y *deleite*. Cuanto mejor sea la velocidad más contento estará el usuario. Para el caso del iPhone la velocidad de la interfaz es un atributo distintivo que genera deleite en sus consumidores. En el diagrama se advierte que la curva de *deleite* también genera una presión por parte de la competencia. La presión se origina a partir del comportamiento dinámico de los consumidores: lo que hoy consideran como *deleite* mañana lo considerarán como *debe ser*. Por este motivo es importante tanto para HTC como para Android mejorar el desempeño de la respuesta.

4.2.2. Identificación de los síntomas

Para analizar la velocidad de respuesta desde el punto de vista del consumidor, se tomaron todos los usuarios que se quejaron sobre este tema, los cuales sumaban un total de 167. Sobre estas 167 opiniones se aplicó el mismo proceso de MT que se explicó en el capítulo 3, y de esta manera se identificaron los principales síntomas que el usuario percibe como rendimiento lento. Estos 167 usuarios representan el 1,5% del total (11022) de la muestra y 5,1% del total de usuarios insatisfechos (3264). En la figura 4.12 se presentan los síntomas encontrados.

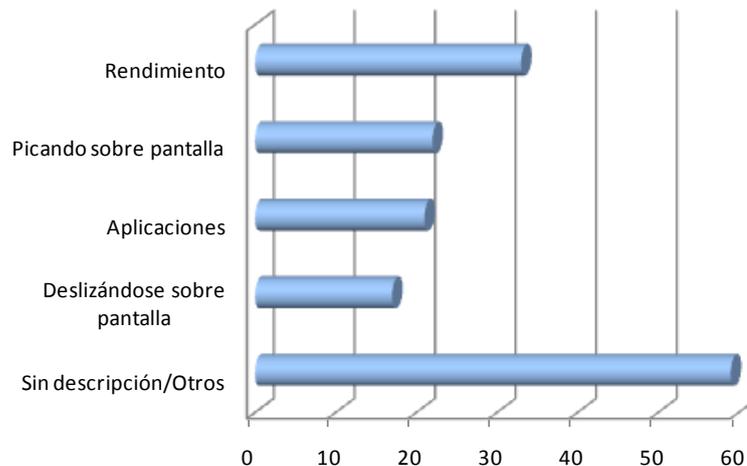


Figura 4.12. Incidencia de síntomas

A continuación se dan los detalles del significado de cada categoría.

1. **Rendimiento.** Refiere al rendimiento general del teléfono, agrupando comentarios del tipo “el teléfono es lento” o “funciona lento”.
2. **Picando sobre pantalla.** Representa la acción de picar el dedo sobre la pantalla, ya sea para escribir o para dar una orden. Luego de ejecutar esta

acción existe un retraso que no agrada al usuario. Aquí se agrupan comentarios del tipo “escribir es lento” o “el teclado es lento”

3. **Aplicaciones.** Al correr o instalar aplicaciones el rendimiento del teléfono no es el esperado. Se juntan comentarios del tipo “retraso al correr x aplicación”.
4. **Desplazándose sobre pantalla.** Representa la acción de arrastrar el dedo sobre el display, ya sea para cambiar de pantalla o para moverse hacia arriba, abajo y a los costados. Agrupa comentarios del tipo “pantalla lenta”.
5. **Sin descripción/Otros.** No hay detalles sobre qué es lo que funciona lento. Simplemente se expresa la existencia de un retraso. También se agrupan casos aislados que pertenecen a un usuario en particular.

Lamentablemente una gran cantidad de usuarios no da detalles sobre cuál es el retraso o qué funciona lento en su producto. Sin embargo 87 consumidores, el 52% de la muestra, expresaron su descontento en torno a los primeros 4 casos. De esta manera, se puede acotar el problema de la velocidad de respuesta y saber en qué puntos del producto focalizarse para mejorar.

4.2.3. Proyección de síntomas

Antes de ver cómo se comporta cada síntoma en el tiempo es oportuno recordar los primeros 2 puntos de la lista de actualizaciones que se presentó en la sección 4.1.3. En la versión Android 2.2 se agregaron las siguientes características:

1. Optimización de la velocidad del sistema operativo, la memoria y el rendimiento.
2. Mejoramiento de la velocidad de las aplicaciones gracias a implementación de JIT.

Sería interesante observar, así como se hizo para el bluetooth, si la actualización mejoró o no el problema. En la figura 4.13 se tiene la proyección de los síntomas.

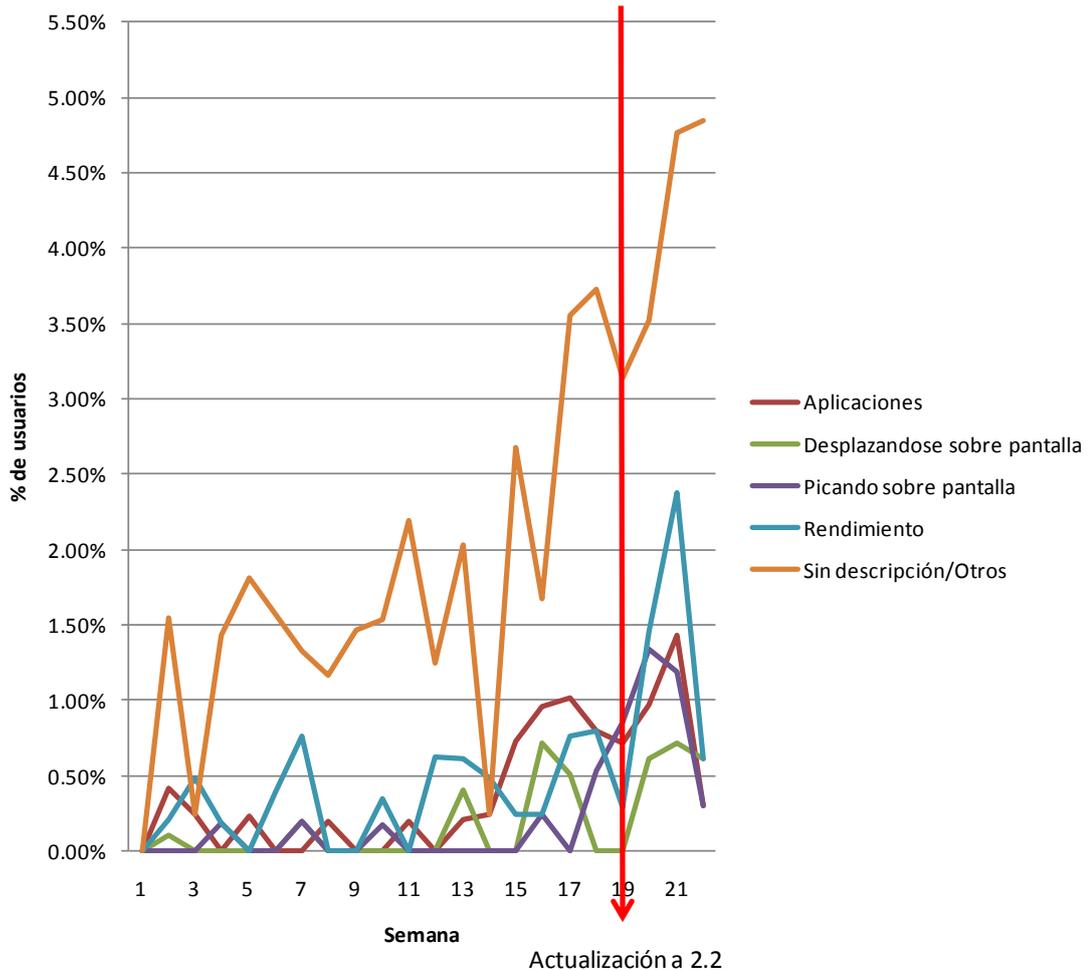


Figura 4.13. Proyección de síntomas

En este gráfico el eje vertical representa la cantidad de usuarios expresando el síntoma sobre la cantidad total de gente comentando sobre el Droid Incredible, por semana. Se observa que para los cuatro síntomas identificados, el comportamiento es parecido al caso del bluetooth que se mostró en la sección 4.1.3: un fuerte incremento luego de la actualización hasta las 2 semanas, cuando el proceso de actualización termina y cae bruscamente. Al parecer este es el típico comportamiento del usuario en vísperas de actualización. Nuevamente, faltan datos para tomar una conclusión definitiva. Sin embargo, llama la atención el caso de “Sin Descripción/Otros”, que luego de la actualización únicamente aumenta. Pareciera que HTC ha ocasionado un problema durante el proceso de actualización que resta velocidad al producto. Aunque se debe tener cuidado al analizar esta categoría, como se dijo anteriormente captura comentarios del tipo “existe un retraso”, sin dar muchos detalles. Como la MT no es una ciencia exacta, lo que sucede es que esas quejas en realidad hacen referencia a un retraso en el proceso de actualización: son consumidores impacientes que luego de la semana 19 no han recibido todavía su nueva versión de Android. Por lo tanto la categoría “Sin

Descripción/Otros” no es representativa en este análisis temporal. Además, es importante aclarar que mientras este análisis temporal toma el total de los consumidores, el análisis de la figura 4.12 se aplicó sobre una muestra de usuarios con problemas de velocidad de respuesta.

La actualización del sistema operativo será un paso clave para Android (y HTC) en la mejora de este punto, sobre todo teniendo en cuenta que el hardware del Incredible se encuentra en los mismos niveles que la competencia. En la tabla 4.1 se comparan algunas especificaciones entre los cuatros productos considerados en este trabajo.

Teléfono	Procesador	Memoria RAM	Sistema operativo
Incredible	1000 MHz	512 MB	Android 2.X
Droid X	1000 MHz	512 MB	Android 2.X
iPhone 4	1000 MHz	512 MB	iOS 4.X
Bold 9700	624 MHz	256 MB	BlackBerry 5.0.X

Tabla 4.1. Comparación de características velocidad

En esta tabla se listan las características que más se relacionan con la velocidad de respuesta, aunque cabe aclarar que por sí solas no son las que la determinan. Se ve que el Incredible cuenta con igual o mejores niveles de procesador y memoria RAM que sus competidores. Sin embargo exhibe un resultado menos satisfactorio en velocidad de respuesta, incluso peor que el Bold 9700 que cuenta con menor capacidad. Si bien existen diferencias entre los proveedores de piezas (como procesador y memoria), la mayor variabilidad se verá en el uso que se le dé a esas piezas, y ahí es donde entra el juego el sistema operativo. Por esta razón, una buena actualización podría llevar a HTC a acercarse a la curva de deleite del diagrama de Kano.

5. COMENTARIOS FINALES

En este capítulo se presentarán los comentarios finales del estudio. En primer lugar se explicará la forma en que una organización puede emplear el proceso presentado, luego un breve análisis de factibilidad y finalmente las conclusiones de la tesis.

5.1. UTILIDAD DE LA INFORMACIÓN PARA LA ORGANIZACIÓN

En la sección 1.1.2 se explicó como la voz del consumidor termina definiendo la voz del proceso, por lo que el estudio de la voz del consumidor es un procedimiento que permitirá mejorar la calidad de los productos de una empresa. La figura 5.1 muestra este proceso de mejora continua.

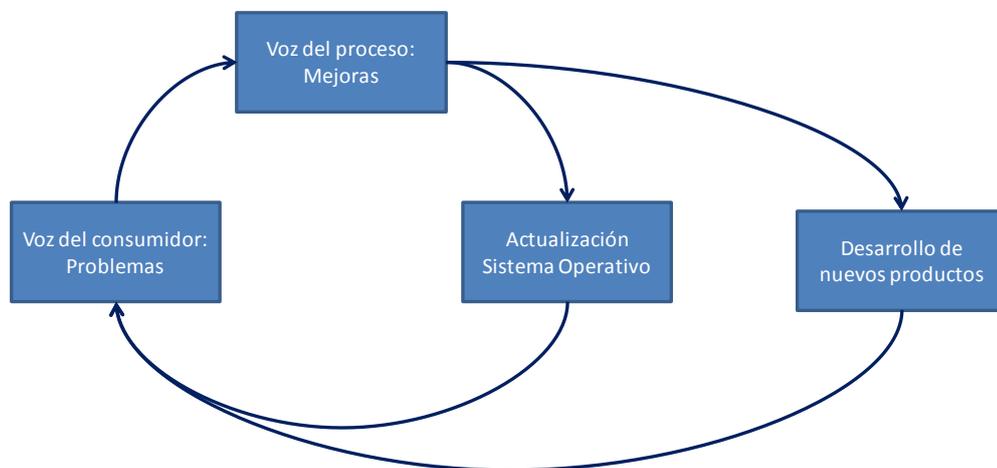


Figura 5.1. Mejoramiento de la calidad

En la figura se presentan dos situaciones. En primer lugar la actualización del sistema operativo del celular que permitirá solucionar problemas a corto plazo. En segundo lugar un caso más general, aplicable a cualquier tipo de industria, donde el estudio de la voz del consumidor ayudará al desarrollo de nuevos productos que cumplan con las expectativas del cliente.

La MT permite capturar los problemas que el consumidor enfrenta a partir de información desestructurada basada en el lenguaje natural. Esto permite analizar gran cantidad de información que a priori carece de valor al estar desorganizada. Tal es el caso de la información proveniente de la web. La MT permite introducir nuevas y valiosas fuentes de información a la organización pero esto no significa que la MT excluya a otros métodos de clasificación. Por el contrario, la MT puede procesar todas las fuentes en simultáneo bajo un mismo sistema de categorización de información: un sistema diseñado por el

analista para investigar los temas que realmente interesan a la organización. La figura 5.2 representa la idea expuesta en este párrafo.

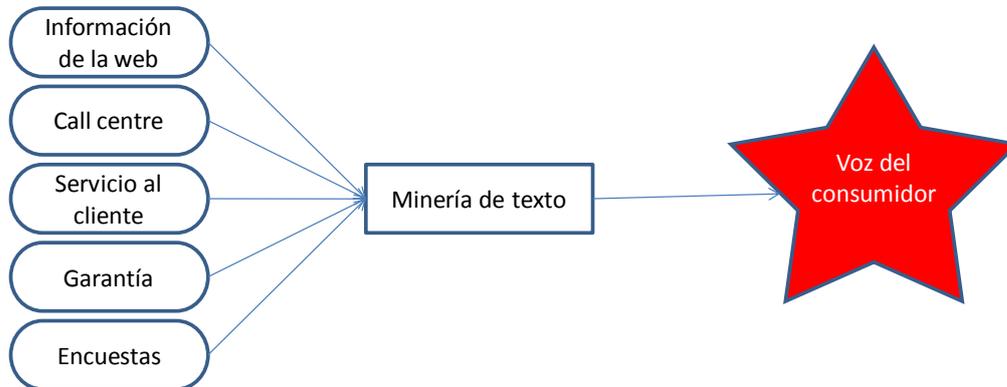


Figura 5.2 Sistema único de categorización con MT

Para el éxito del proceso, es importante que exista un importante vínculo entre el área de investigación de mercado y de la experiencia del consumidor con las áreas de procesos y desarrollo. Por ejemplo, a través de reuniones semanales de compartimiento de resultados.

5.2. FACTIBILIDAD DE LA MINERÍA DE TEXTO

En esta sección se evaluarán las posibilidades que una organización tiene para implementar la MT como herramienta para el estudio de la voz del consumidor.

5.2.1. Su fundamento: la mejora de la eficiencia

La MT permite mejorar considerablemente el proceso de categorización de la información, esta es su principal ventaja. Como se comentó en la sección 3.3.1 la operación “procesar y clasificar la información de 1000 usuarios” toma 30 minutos con el proceso propuesto: 29 minutos de extracción de la Web y 1 minuto de categorización con MT. Esto puede ser comparado con otras alternativas de clasificación, como por ejemplo, el call centre y el método de la encuesta. La tabla 5.1 compara estas posibilidades

MT	Call centre	Encuesta
1,8 segundos/usuario	6 minutos/usuario	25,2 min/usuario

Para el caso de la encuesta, el autor ha participado en procesos de encuesta a clientes. Se tomó el parámetro de que 100 encuestas pueden ser respondidas por igual cantidad de usuarios en aproximadamente 42 horas. Si bien es el método más lento es el que proporciona la información de mayor detalle para la organización, pues la encuesta apunta a responder las

preguntas que la empresa desea contestar. Los otros dos casos son de expresión más libre y el usuario dice solamente algunas cosas.

La gran eficiencia lograda por la MT tiene el costo de una peor eficacia: su precisión es del 70% al 80%. Por otro lado, otras alternativas ofrecen una precisión cercana al 100% (Siempre existe el error humano).

Otra ventaja de la MT es que permite llegar a un nivel de detalle de gran utilidad para la organización. Esto se vio en la sección 4.2, profundizando sobre la velocidad de respuesta.

5.2.2. Implementación

La implementación de la MT es básicamente una decisión de compra, pues la MT es una herramienta informática, en este caso particular el IBM SPSS Modeler. De todas formas, la adquisición de la herramienta no significa que inmediatamente se logran los resultados, existe una etapa de entrenamiento y desarrollo (capítulo 3) que toma su tiempo de trabajo. La figura 5.3 muestra las tareas a llevarse a cabo para implementar la MT en la empresa.



Figura 5.3. Implementación de MT

En la figura se observa un proceso rojo y otro azul con operaciones equivalentes: compra, entrenamiento y desarrollo. El rojo corresponde a la implementación de la MT, mientras que el azul corresponde a la implementación de Web Scraping para usar la web como fuente. En este trabajo se plantea el uso de la web como fuente de información, aunque la empresa podría tener otras fuentes de texto que desee analizar, por lo que el uso de la web puede considerarse opcional. Se observa en la figura que la implementación dura unas 11 semanas, siendo la tarea más larga el desarrollo de la MT, tema que se vio en el capítulo 3 y cuenta con una duración de 8 semanas.

El entrenamiento está incluido en la compra de las herramientas y se da a un empleado o grupo de empleados. En principio no sería necesario contratar nuevos empleados, pues la capacitación les permitiría tomar el conocimiento y aplicarlo a sus tareas diarias. Se asume que el soporte informático necesario para operar la herramienta ya existe, pues hoy en día esto es lo normal para la

mayoría de las industrias. De todas formas, a continuación se listan los requerimientos de sistema para instalar el IBM SPSS Modeler

Sistema operativo

- Microsoft® Windows® 7 (Professional y Enterprise) x32 y x64 Editions.
- Microsoft Windows Vista (Business y Enterprise) con Service Pack 1 x32 y x64 Editions
- Microsoft Windows XP Professional con Service Pack 3 x32 y x64 Editions.

Hardware

- Procesador Intel® Pentium® o mayor (para 32.bit Windows); x64 (AMD 64 y EM64T) familia de procesadores (para 64.bit Windows).
- Un monitor con resolución 1024x768 o mayor.
- Una unidad DVD.ROM drive si se instala desde el disco.
- Espacio libre en el disco. 2 GB.
- RAM. 2 GB o más
- Software. Microsoft® Internet Explorer® 6.0 o mayor para ayuda online.

El costo de la licencia es de **20000 dólares** y es permanente. Aunque es importante aclarar que existen otras soluciones de este tipo pero todas rondan este orden de magnitud. Por otro lado, si se quisiera tomar la web como fuente, adicionalmente se deberá adquirir una herramienta de web Scraping. La utilizada en esta tesis es Kapow Web Data Server, cuyo costo es de **40000 dólares**. Al igual que con el IBM SPSS Modeler, existen otras soluciones de costo similar.

5.2.3. Perfil de empresa

Para el proceso que se presenta en este trabajo el tipo de empresa a la cual se aplicaría mejor es la siguiente:

- Una empresa que cuente con una gran cantidad de discusión sobre sus productos en la web.
- Fuerte orientación al mejoramiento de la calidad de los productos.

Las organizaciones que típicamente se ajustan a esta descripción son empresas que fabrican productos electrónicos, como por ejemplo: celulares (el caso de este trabajo), cámaras digitales, GPS, computadoras, televisores, laptops, etc. A su vez, podrían considerarse algunos casos alternativos, como por ejemplo:

- Una empresa con una gran cantidad de información en base a texto que necesite procesarla de alguna manera eficiente.
- Una empresa que busque un método de investigar a la competencia.

5.3. CONCLUSIONES

5.3.1. Sobre los resultados obtenidos

Haciendo uso de la MT se investigó la voz de los consumidores del Droid Incredible, tomando 11022 usuarios reales de la web que emitieron comentarios. La MT permitió extraer la opinión de estos comentarios, una aplicación conocida como MO, y de esta manera se logró obtener información de alto valor a partir de datos desestructurados.

Este trabajo contaba inicialmente con 2 objetivos, los cuales fueron listados en la sección 1.5.1 y se reiteran a continuación:

1. A través del uso de la MT, identificar los sentimientos, actitudes, opiniones y posturas de un grupo de individuos hacia un determinado objeto de estudio.
2. Demostrar que el dominio público de la web constituye una valiosa fuente de información sobre los consumidores para la organización.

Un criterio de éxito para determinar el cumplimiento de los objetivos es que:

- La organización puede tomar acciones sobre el análisis proporcionado.
- La organización puede priorizar sobre cuáles son las acciones más importantes que debe tomar.
- El análisis es tan convincente que directamente implica las acciones que se deben tomar.

Bajo la mirada de este criterio los objetivos han sido cumplidos. En primer lugar se dio un listado de 10 fallas y 10 insatisfacciones indicando claramente cuál es el problema y con la MT que da veracidad a los resultados encontrados: el caso de la señal en el iPhone muestra la robustez de la MT. En segundo lugar, los problemas fueron ordenados por frecuencia de incidencia, indicando una prioridad. Conociendo la severidad de estos problemas, la empresa sabrá cuál es la primera falla o insatisfacción que deberá atacar. Finalmente, para el tercer punto, se puede hacer referencia al caso de la insatisfacción respecto de la vida útil del batería. La especificación es encuentra muy por debajo del nivel competitivo e indica que HTC debería usar baterías más duraderas en sus productos. Por otro lado la profundización de la velocidad de respuesta en la

sección 4.2 permite saber con precisión quirúrgica cuáles son los puntos a mejorar, basándose en la detección de los síntomas.

Habiendo aprobado el criterio de éxito, se concluye que la web cuenta con valiosa información sobre la voz del consumidor, la cual puede ser capturada a través de un proceso de MT.

5.3.2. Sobre el trabajo realizado a nivel académico

El proyecto final de Ingeniería Industrial ha permitido al alumno (al autor) relacionar conceptos y técnicas innovadoras que tomó de su experiencia profesional con conceptos y herramientas desarrolladas durante el estudio de Ingeniería Industrial en el ITBA. En particular, la principal relación es con la materia de calidad, de la cual se emplearon algunas de sus herramientas, como diagramas de Pareto, diagrama de Kano y proyecciones temporales. Si bien el trabajo estuvo mayormente enfocado hacia esta materia, el análisis también considera aspectos de Diseño de Producto, Marketing, Investigación de operaciones, Dinámica de sistemas, Programación, entre otras. Este trabajo podría ser tomado bajo cualquiera de estos enfoques y ser repetido desde aquellos puntos de vista.

Se presenta en este trabajo un proceso de mejoramiento de la calidad (Sección 5.1), el cual obtiene resultados con la ayuda de la MT. La propuesta final que el alumno hace es la posibilidad de aplicar este proceso a una infinidad de industrias, por ejemplo las listadas en la sección 5.2.3. El mejoramiento de la satisfacción del consumidor es una base para la optimización de los resultados de una organización, y sobre esta noción el alumno percibe la representación de un concepto esencial de Ingeniería Industrial: precisamente, la optimización de resultados. Por razones de confidencialidad de información, el grado en que una empresa puede mejorar los resultados no tuvo lugar en el trabajo. Sin embargo, dentro del contexto considerado, existen varios parámetros para el éxito del proceso planteado. Por ejemplo, nivel de ventas o tasa de retorno de productos.

6. ANEXOS

Debido a la extensión de los anexos, éstos se encuentran en formato digital. A continuación se detalla el contenido de los mismos.

6.1. Anexo I: Comentarios extraídos de la Web

6.1.1. Droid Incredible (HTC)

6.1.2. iPhone 4 (Apple)

6.1.3. Bold 9700 (BlackBerry)

6.1.4. Droid X (Motorola)

6.2. Anexo II: Descriptores de las principales categorías

6.2.1. Descriptores de fallas

6.2.2. Descriptores de insatisfacciones

7. BIBLIOGRAFÍA

Ah-Hwee Tan. 1999. Text Mining: Promises and Challenges. Kent Ridge Digital Labs.

Amazon. <http://www.amazon.com/>. Página vigente al 31/10/2010.

Android Community. <http://androidcommunity.com/forums/index.php>. Página vigente al 31/10/2010.

Android Forums. <http://androidforums.com/>. Página vigente al 31/10/2010.

Apple Support Forums. <http://discussions.apple.com/index.jspa>. Página vigente al 31/10/2010.

AT&T. <http://www.att.com/>. Página vigente al 31/10/2010.

Best Buy. <http://www.bestbuy.com/>. Página vigente al 31/10/2010.

BlackBerry Support Community Forums. <http://supportforums.blackberry.com/>. Página vigente al 31/10/2010.

Cellphone Forums. <http://cellphoneforums.net/>. Página vigente al 31/10/2010.

Cnet. <http://www.cnet.com/>. Página vigente al 31/10/2010.

Facebook. <http://www.facebook.com/>. Página vigente al 31/10/2010.

Gsmarena. <http://www.gsmarena.com/>. Página vigente al 31/10/2010.

Howard Forums. <http://www.howardforums.com/>. Página vigente al 31/10/2010.

HTC Forums. <http://community.htc.com/na/htc-forums/f/default.aspx?tab=support>. Página vigente al 31/10/2010.

IBM SPSS Modeler. 2009. Manual de Usuario.

Kapow Web Data Server. 2009. Manual de Usuario.

Letstalk. <http://www.letstalk.com/>. Página vigente al 31/10/2010.

Motorola Owner's Forum. <https://supportforums.motorola.com/index.jspa>. Página vigente al 31/10/2010.

Ochoa, Xavier. 2008. Quantitative analysis of user generated content on the web.

Phonearena. <http://www.phonearena.com/>. Página vigente al 31/10/2010.

Phonescoop. <http://www.phonescoop.com/>. Página vigente al 31/10/2010.

Shopping. <http://www99.shopping.com/>. Página vigente al 31/10/2010.

TalkAndroid forums. <http://www.talkandroid.com/>. Página vigente al 31/10/2010.

T-Mobile Community. <http://forums.t-mobile.com/>. Página vigente al 31/10/2010.

US Census Bureau. <http://www.census.gov/>. Página vigente al 31/10/2010.

Verizon Wireless. <http://www.verizonwireless.com/b2c/index.html>. Página vigente al 31/10/2010.

Viewpoints. <http://www.viewpoints.com/>. Página vigente al 31/10/2010.

Wikipedia. <http://www.wikipedia.org/>. Página vigente al 31/10/2010.