

**CIS1630AP06**  
Real Time Social Data Mining

Daniel Alejandro Calambás Marín  
Jaime Andrés Mendoza Mendoza

PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
CARRERA DE INGENIERIA DE SISTEMAS  
BOGOTÁ, D.C.  
2016



CIS1630AP06  
Real Time Social Data Mining

**Autor(es):**

Daniel Alejandro Calambás Marín  
Jaime Andrés Mendoza Mendoza

MEMORIA DEL TRABAJO DE GRADO REALIZADO PARA CUMPLIR UNO  
DE LOS REQUISITOS PARA OPTAR AL TITULO DE INGENIERO DE  
SISTEMAS

**Director**

Ing. Alexandra Pomares Quimbaya Ph.D.

**Jurados del Trabajo de Grado**

Juan Carlos García Díaz

Rafael Andrés Gonzales Rivera

**Página web del Trabajo de Grado**

<http://pegasus.javeriana.edu.co/~CIS1630AP06>

PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
CARRERA DE INGENIERIA DE SISTEMAS  
BOGOTÁ, D.C.  
Noviembre, 2016

**PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE INGENIERIA  
CARRERA DE INGENIERIA DE SISTEMAS**

**Rector Magnífico**

Jorge Humberto Peláez Piedrahita, S.J.

**Decano Facultad de Ingeniería**

Ingeniero Jorge Luis Sánchez Téllez

**Director de la Carrera de Ingeniería de Sistemas**

Ingeniera Mariela Josefina Curiel Huérfano

**Director Departamento de Ingeniería de Sistemas**

Ingeniero Efraín Ortiz Pabón

**Artículo 23 de la Resolución No. 1 de Junio de 1946**

*“La Universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus proyectos de grado. Sólo velará porque no se publique nada contrario al dogma y la moral católica y porque no contengan ataques o polémicas puramente personales. Antes bien, que se vean en ellos el anhelo de buscar la verdad y la Justicia”*

## **AGRADECIMIENTOS**

Agradecemos primeramente a Dios por habernos permitido estudiar Ingeniería de Sistemas en la Pontificia Universidad Javeriana y por habernos dado la inteligencia y perseverancia para alcanzar esta meta.

A nuestras familias por todo el apoyo tanto moral como económico que nos han dado no solo a lo largo del desarrollo de este trabajo de grado sino también a lo largo de todos estos años de estudio.

A nuestras novias, Estefanía y Diana, por apoyarnos, subir nuestros ánimos en los momentos más críticos y por estar siempre pendientes de nosotros.

A nuestra directora de trabajo de grado, Ingeniera Alexandra Pomares por guiarnos e instruirnos en todo el desarrollo de este trabajo de grado ya que sin ella este proyecto no hubiese sido posible.

## CONTENIDO

<b>CONTENIDO .....</b>	<b>5</b>
<b>LISTA DE ILUSTRACIONES.....</b>	<b>7</b>
<b>LISTA DE TABLAS .....</b>	<b>9</b>
<b>I - INTRODUCCIÓN.....</b>	<b>12</b>
<b>II - DESCRIPCIÓN GENERAL .....</b>	<b>14</b>
1. OPORTUNIDAD, PROBLEMÁTICA, ANTECEDENTES .....	14
1.1. <i>Formulación del problema que se resolvió .....</i>	<i>15</i>
1.2. <i>Justificación del problema.....</i>	<i>15</i>
1.3. <i>Impacto Esperado .....</i>	<i>16</i>
2. DESCRIPCIÓN DEL PROYECTO.....	16
2.1. <i>Objetivo general.....</i>	<i>16</i>
2.2. <i>Objetivos específicos.....</i>	<i>16</i>
3. METODOLOGÍA .....	17
<i>CRISP-DM: Entendimiento del problema.....</i>	<i>17</i>
<i>Personal Extreme Programming: Desarrollo de Software .....</i>	<i>19</i>
<b>III – MARCO TEÓRICO .....</b>	<b>21</b>
1. MARCO CONTEXTUAL.....	21
5. MARCO CONCEPTUAL .....	23
<b>IV – DESARROLLO DE TRABAJO .....</b>	<b>25</b>
1. VISTA GENERAL DEL PROYECTO .....	25
<i>¿Qué es RART? .....</i>	<i>25</i>
<i>¿Por qué un Framework? .....</i>	<i>25</i>
<i>¿Qué tipo de información se extrae de la red social Facebook? .....</i>	<i>25</i>
<i>¿Cómo funciona?.....</i>	<i>26</i>
2. DESARROLLO DEL FRAMEWORK RART .....	27
<i>Definición y levantamiento de requerimientos.....</i>	<i>27</i>
<i>Definición de casos de uso.....</i>	<i>36</i>
<i>Definición de las tecnologías a utilizar .....</i>	<i>39</i>
<i>Diseño del sistema.....</i>	<i>40</i>

---

<b>V – RESULTADOS Y VALIDACIÓN .....</b>	<b>61</b>
<i>Resultados .....</i>	<i>61</i>
<b>VI – CONCLUSIONES, IMPACTOS Y TRABAJO FUTURO .....</b>	<b>77</b>
1.    CONCLUSIONES.....	77
<i>Pregunta de Investigación Respondida.....</i>	<i>77</i>
<i>Objetivos Cumplidos.....</i>	<i>77</i>
2.    ANÁLISIS DE IMPACTO DEL DESARROLLO .....	77
<i>Impacto Disciplinar .....</i>	<i>77</i>
<i>Impacto Social.....</i>	<i>78</i>
<i>Impacto Económico .....</i>	<i>78</i>
3.    TRABAJO FUTURO.....	78
<b>VII- REFERENCIAS Y BIBLIOGRAFÍA .....</b>	<b>79</b>
<b>VIII – ANEXOS .....</b>	<b>81</b>



**LISTA DE ILUSTRACIONES**

Ilustración 1 - Metodología .....	17
Ilustración 2 - Ciclo de Vida de Extreme Programming (Extreme Programming: A gentle introduction, 2016).....	19
Ilustración 3 - Funcionamiento de RART .....	26
Ilustración 4 - Proceso de Ingeniería de requerimientos .....	27
Ilustración 5 - Plantilla Especificación de Requerimientos .....	28
Ilustración 6 - Diagrama Casos de Uso .....	37
Ilustración 7 - Diagrama de Arquitectura y Componentes de RART .....	40
Ilustración 8 - Estructura de los Archivos y Base de Datos .....	44
Ilustración 9 - Diagrama de Despliegue (1).....	46
Ilustración 10 - Diagrama de Despliegue (2).....	48
Ilustración 11 - Diagrama de Secuencia Ingresar al Sistema.....	49
Ilustración 12 - Diagrama de Secuencia Ingresar Página (id) .....	50
Ilustración 13 - Diagrama de Secuencia Ingresar Página (nombre).....	51
Ilustración 14 - Diagrama de Secuencia Buscar Publicación .....	52
Ilustración 15 - Diagrama de Secuencia Ver Detalle Publicación .....	53
Ilustración 16 - Diagrama de Secuencia Buscar Página en la Base de Datos.....	54
Ilustración 17 - Visualizar publicaciones página .....	55
Ilustración 18 - Diagrama de Secuencia Filtrar publicación .....	55
Ilustración 19 - Diagrama de Secuencia Graficar Reacciones .....	56
Ilustración 20 - Proceso de Extracción.....	57
Ilustración 21 - Proceso de Consulta.....	58

---

Ilustración 22 - Diagrama de Clases Componente Extracción de Datos .....	59
Ilustración 23 - Diagrama de Clases Componente Carga y Transformación de Datos .....	60
Ilustración 24 - Cumplimiento de Objetivos Específicos .....	62
Ilustración 25 - Página Principal RART .....	62
Ilustración 26 - Menú Extraer Información .....	63
Ilustración 27 - Menú Buscar Información .....	64
Ilustración 28 - Actualizar Información .....	65
Ilustración 29 - Menú Configuraciones .....	66
Ilustración 30 - Cantidad publicaciones vs. tiempo .....	69
Ilustración 31 - Cantidad reacciones vs. tiempo .....	69
Ilustración 32 - Cantidad comentarios vs. tiempo .....	70
Ilustración 33 - Cantidad publicaciones vs. tiempo (3 en simultaneo) .....	71
Ilustración 34 - Cantidad de comentarios vs. tiempo (3 en simultaneo) .....	71
Ilustración 35 - Cantidad de reacciones vs. tiempo (3 en simultaneo) .....	72
Ilustración 36 - Tiempo de consulta .....	73
Ilustración 37 - Resultados Pruebas de Usuario .....	76

## LISTA DE TABLAS

Tabla 1 - Análisis de Tecnologías.....	18
Tabla 2 - Priorización de Requerimientos .....	30
Tabla 3 - Validación de Requerimientos .....	31
Tabla 4 - Descripción de Requerimientos .....	36
Tabla 5 - Casos de Prueba .....	68
Tabla 6 - Escala de Calificación TAM.....	74
Tabla 7 - Preguntas Pruebas de Usuario.....	75

## RESUMEN

Una de las tareas más complejas que realizan las empresas es recolectar información necesaria y suficiente para lograr conocer las necesidades de sus consumidores, con el fin de ofrecerles productos y/o servicios que satisfagan plenamente sus necesidades. Por otra parte, la red social Facebook cuenta con la mayor cantidad de usuarios registrados y activos en el mundo (Stone, 2016), por ende, esto la convierte en una rica fuente de información, la cual al organizarla correctamente puede brindar solución a la problemática mencionada. Lo que se desarrolla en este proyecto es un componente de software el cual extrae información de las publicaciones realizadas por las empresas dentro de la red social Facebook con el fin de generar un repositorio de información apto para aplicar procesos de *Data Analytics*, con el objetivo de perfilar los consumidores digitales de dichas empresas.

## ABSTRACT

One of the most complex tasks that companies do is to collect enough information to know the needs of its consumers in order to offer products and/or services to satisfy their needs. In the order hand, the social network Facebook has the largest number of registered and active users around the globe, therefore, this makes Facebook a rich source of information, which properly organized can provide solution to the mentioned problematic. This project developed a software component that extracts information from the social network Facebook in order to generate an information repository suitable to apply *Data Analytics* processes for the purpose of profiling the consumers of mentioned enterprises.

## RESUMEN EJECUTIVO

El problema abordado durante el desarrollo de este proyecto fue la necesidad de extraer información proveniente de la red social Facebook, con propósito de tener una base de datos a la cual aplicar procesos de *Data Analytics*, que genere conocimiento para apoyar la toma de decisiones. Facebook genera una gran cantidad de información cada minuto, lo que convierte a esta problemática un problema a abordar desde la perspectiva de Big Data.

La solución desarrollada consiste en un *Framework* cuyo propósito es extraer información de la red social Facebook. Esta información es producto de las publicaciones que realiza una *fan page* de la red social. Las publicaciones obtienen comentarios y reacciones de los usuarios que las visualizan y esta información es extraída a una base de datos por medio del *Framework* desarrollado conocido como RART.

RART utiliza un API de la red social Facebook, llamado RestFb, el cual permite realizar el proceso de consulta y extracción de la información allí existente. RART fue desarrollado como una aplicación web en el lenguaje de programación Java.

Dado el volumen, variedad y velocidad con la que se genera información dentro de la red social, fue necesario implementar tecnologías de almacenamiento con alcances de Big Data. RART utiliza dos medios de almacenamiento. El primero de ellos es Mongo DB, una base de datos no relacional en donde se almacena la información semiestructurada extraída de Facebook. La segunda es Apache Solr, un motor de búsqueda el cual permite almacenar información no estructurada y brindando un alto rendimiento a la hora de consultar la información que allí se almacena.

Una vez desarrollado el *Framework*, se sometió a pruebas con futuros usuarios del mismo. Los resultados mostraron una gran acogida por parte de ellos, los cuales consideran que RART es una herramienta de gran impacto a la hora de analizar los comportamientos de los usuarios de la red social Facebook.

## I - INTRODUCCIÓN

Con la creación de la Alianza CAOBA (Centro de Excelencia y Apropiación en Big Data y *Data Analytics*) (TIC, 2016) liderada por la Pontificia Universidad Javeriana, las empresas que hacen parte de esta alianza dieron a conocer sus necesidades tecnológicas. Una de estas empresas, la cual se dedica a la producción de alimentos, identificó la necesidad de perfilar sus consumidores digitales con base en la información obtenida en las publicaciones realizadas dentro de la red social Facebook.

Una vez identificada esta necesidad, se propuso construir un *Framework* el cual extraiga las reacciones y comentarios obtenidos por las publicaciones realizadas por las distintas *fan page*, que hacen parte de la empresa de alimentos. Esto, con el fin de organizarla dentro de un repositorio de datos al cual se le podrá aplicar procesos de *Data Analytics*, y así poder generar conocimiento acerca de los consumidores de sus productos.

Este documento tiene como propósito presentar el proceso y los resultados llevados a cabo en este trabajo de grado, requisito de la carrera de Ingeniería de Sistemas de la Pontificia Universidad Javeriana, donde se desarrolló el *Framework* anteriormente mencionado, llamado RART. RART cumple la funcionalidad de extraer información proveniente de publicaciones realizadas por *las fan page* de la red social Facebook con el fin de crear un repositorio organizado con dicha información.

Este documento consta de las siguientes secciones:

**Descripción General:** describe la formulación del problema identificado y los objetivos que se deben cumplir para solucionar la problemática, además de la metodología implementada para la identificación y desarrollo de la solución.

**Marco Teórico:** explica los conceptos teóricos necesarios para entender la problemática y desarrollar la solución.

**Desarrollo de la Solución:** describe todas las actividades realizadas como parte del desarrollo del *Framework* RART el cual satisface la necesidad identificada.

**Resultados:** describe los resultados obtenidos como parte del desarrollo de la solución y muestra cómo se dio cumplimiento a cada objetivo planteado.

**Conclusiones:** presenta las conclusiones obtenidas como parte del desarrollo de este trabajo de grado, el impacto esperado del mismo y las actividades que pueden llegar a ser parte del trabajo futuro a realizar.

## II - DESCRIPCIÓN GENERAL

### 1. Oportunidad, Problemática, Antecedentes

CAOBA es una Alianza conformada en diciembre de 2015 por once entidades del sector público privado y de la academia. Esta alianza liderada por la Pontificia Universidad Javeriana busca fomentar el desarrollo de capacidades en Big Data y *Data Analytics* con el fin de favorecer la innovación y la investigación en dicho tema para promover la competitividad tanto de las empresas como de la academia (TIC, 2016).

Dentro de las empresas que integran el sector privado de la Alianza CAOBA se encuentra una empresa del sector de los alimentos. Uno de los principales intereses de esta empresa es conocer más de cerca a sus consumidores digitales a partir de sus hábitos o estilo de vida, para ofrecerles nuevas experiencias que satisfagan plenamente sus aspiraciones o deseos (Pomares, 2016).

Dado que dicha empresa compite en mercados altamente cambiantes y el éxito de la misma radica, principalmente, en satisfacer las necesidades de los consumidores (Villota, 2016), es necesario conocer a qué tipo de consumidores se está enfrentando y dado el hecho que todos los consumidores tienen distintos hábitos de consumo, es imprescindible determinar dichos comportamientos (Franco & Libreros, 2013).

El comportamiento de cada individuo dentro del mercado es diferente uno de otro. Por mucho que estos difieran es posible analizar las tendencias y características con el fin de determinar unos grupos o segmentos específicos del mercado. La importancia de determinar dichos comportamientos radica en conocer el nicho de mercado a profundidad para entender, con exactitud, qué ofrecer y cómo ofrecer un producto o servicio al tener en cuenta los gustos específicos de cada segmento (Pérez & Martínez, 2006).

Las redes sociales se alimentan diariamente de una inmensa cantidad de comentarios y reacciones de sus usuarios frente a algún evento o publicación, reflejando en gran medida el comportamiento de las personas. A su vez, son medios en donde los consumidores expresan sus gustos acerca de las experiencias de su diario vivir. Por ende, esto las convierte en una rica fuente de información no solo para la empresa anteriormente mencionada sino para que las



empresas interesadas en determinar los comportamientos de consumo de su nicho de mercado generen una ventaja competitiva (García, 2010).

Las empresas publican a través de su cuenta de la red social Facebook publicidad y campañas promocionales e informativas relacionadas a las diferentes marcas y líneas de productos o servicios que estas poseen. De estas publicaciones se pueden extraer datos como los comentarios que hacen los usuarios y las reacciones de estos, ante dichas publicaciones. Estos datos y metadatos permiten determinar los comportamientos y tipos de consumidores de acuerdo con sus reacciones, ya sean positivas o negativas.

Extraer toda esta gran cantidad de información y procesarla de tal manera que se estructure para facilitar su posterior análisis, permite generar conocimiento para mejorar la toma de decisiones empresariales. Este proceso es conocido como *Data Analytics* (SAS Institute, 2003) y es llevado a cabo con una serie de técnicas y algoritmos los cuales organizan la información de tal manera que es posible hallar un nuevo conjunto de información el cual puede responder de una mejor manera una pregunta de negocio planteada.

### **1.1. Formulación del problema que se resolvió**

¿Cómo extraer y organizar datos en tiempo real, que brinden información de la interacción usuario-publicación en Facebook, para apoyar el proceso de aplicación de *Data Analytics*?

### **1.2. Justificación del problema**

Analizar la información proveniente de redes sociales es indispensable para realizar los procesos de *Data Analytics* con el fin de determinar los comportamientos (Sanz, 2003), reacciones y gustos de las personas frente a una situación determinada. Para esto, es necesario extraer dicha información, clasificarla según su utilidad y luego organizarla, de tal manera que sea fácil de consultar cuando se esté realizando algún proceso de *Data Analytics*, generando conocimiento que soporte el proceso de toma de decisiones dentro de las empresas (Schroeck *et al*, 2012).

Esto es posible realizarlo desarrollando un marco de trabajo (*Framework*) que permita extraer y organizar información en tiempo real proveniente de la red social Facebook. Su creación fue

realizada por medio del lenguaje de programación Java que facilita la extracción de la información desde Facebook con su API escrito en el mismo lenguaje. Además, la información extraída es posible almacenarla en una base de datos Apache Solr (Apache, 2016) la cual brinda un excelente rendimiento a la hora de almacenar y consultar información no estructurada, como lo son los comentarios realizados por un usuario en una publicación.

### **1.3. Impacto Esperado**

Este trabajo de grado es parte de la primera fase de un proyecto de CAOBA el cual tiene como objetivo lograr segmentar los consumidores digitales de una empresa de alimentos con información proveniente de publicaciones realizadas dentro de la red social Facebook. Se espera, que una vez entregado el software, producto de este trabajo de grado, se brinde a CAOBA una herramienta la cual apoye los procesos de *Data Analytics* a ser desarrollados con el fin de generar nuevo conocimiento con base en la información extraída desde la red social.

## **2. Descripción del Proyecto**

Esta sección describe el objetivo general y los objetivos específicos del trabajo de grado y, además, describe las fases metodológicas utilizadas para culminar con éxito el desarrollo del proyecto.

### **2.1. Objetivo general**

Desarrollar un marco de trabajo (*Framework*) para recolectar información de publicaciones realizadas por una empresa de alimentos dentro de la red social Facebook, con el fin de generar un repositorio de datos apto para soportar procesos de *Data Analytics*.

### **2.2. Objetivos específicos**

1. Desarrollar un componente de software que extraiga información de usuarios y publicaciones de la red social Facebook.
2. Desarrollar un componente de software que organice la información extraída dentro de un repositorio de datos.
3. Validar el funcionamiento de los componentes desarrollados y la integración entre los mismos.

### 3. Metodología



Ilustración 1 - Metodología

La metodología utilizada para el desarrollo de este trabajo se compone de dos fases: la primera fase de Entendimiento del Problema donde se levantó la información necesaria para tener claridad del alcance del proyecto y la segunda fase correspondiente al Desarrollo del Software donde se aplicó un proceso de Ingeniería de Software para cumplir de manera exitosa con el desarrollo del producto. Para mayor información acerca de la metodología y ciclo de vida del proyecto ver el Anexo [RART Software Project Management Plan](#).

#### CRISP-DM: Entendimiento del problema

Teniendo en cuenta que la finalidad del Software desarrollado es brindar un repositorio apto para aplicar procesos de *Data Analytics*, durante esta etapa se aplicaron las primeras tres fases de la metodología CRISP-DM (Azevedo, 2008) con el fin de garantizar la definición de unos objetivos claros y un alcance que cumpla dicha finalidad. A continuación, se presentan estas tres etapas, que hacen parte de CRISP-DM, y las actividades y resultados obtenidos en cada una de ellas:

##### 1. Entendimiento del Negocio

Esta etapa consistió en entender las necesidades principales que el cliente desea satisfacer con el software resultado de este trabajo de grado. El cliente, la alianza CAOBA, desea poder tener una fuente de información limpia y organizada de los consumidores digitales con información proveniente de sus publicaciones realizadas dentro de la red social Facebook.

##### 2. Entendimiento de los Datos

En esta etapa se indagó acerca de las tecnologías existentes que permiten extraer información de la red social Facebook.

Para ello se descargaron los principales APIs existentes y se realizaron pruebas de extracción de información con cada uno de ellos. Como se muestra en la tabla 1, se calificaron los siguientes aspectos:

1. **Soporte por parte de Facebook:** se determinó si las tecnologías seleccionadas están soportadas por parte de Facebook con el fin de seleccionar una que posea dicho soporte.
2. **Consistencia de la información:** con cada una de las tecnologías se realizaron procesos de extracción y se compararon los resultados para determinar qué tan consistente fue la información resultante de la extracción.
3. **Facilidad de uso:** así mismo se determinó que tan fáciles de utilizar son las tecnologías de aprendizaje de acuerdo al conocimiento de los desarrolladores del proyecto y así escoger la que tenga una curva de aprendizaje más conveniente.





Tecnología	Soporte por parte de Facebook	Consistencia de la información	Facilidad de uso
 spring <small>by Pivotal.</small>	Si	7/10	6/10
 restfb	Si	8/10	9/10
 kinvey	Si	5/10	4/10
 facebook4j	No	3/10	9/10

Tabla 1 - Análisis de Tecnologías

Como resultado del análisis llevado a cabo, se seleccionó el API RestFb dado que está soportado por la red social Facebook. Adicionalmente, brinda la mejor consistencia en la información extraída y es fácil de utilizar de acuerdo con los conocimientos de los desarrolladores del proyecto. (Facebook, 2016).

### 3. Preparación de los Datos

Una vez teniendo claro el alcance, objetivos y restricciones de las tecnologías se procedió a determinar qué información extraer. En esta etapa de la metodología se obtuvo como resultado el listado de los datos a extraer de las publicaciones realizadas dentro de la red social. Esta información se puede ver en la sección [2. Desarrollo del Framework RART](#) en la sección de [Persistencia](#).

### Personal Extreme Programming: Desarrollo de Software

Dado que el equipo de trabajo disponible para este trabajo de grado constó únicamente de dos personas, se hizo uso de la metodología ágil de desarrollo de software conocida como *Personal Extreme Programming* (Extreme Programming: A gentle introduction, 2016). A continuación, se presentan las distintas etapas, que se muestran en la ilustración 2 que hacen parte de dicha metodología y cómo se llevó a cabo en el desarrollo del trabajo de grado.

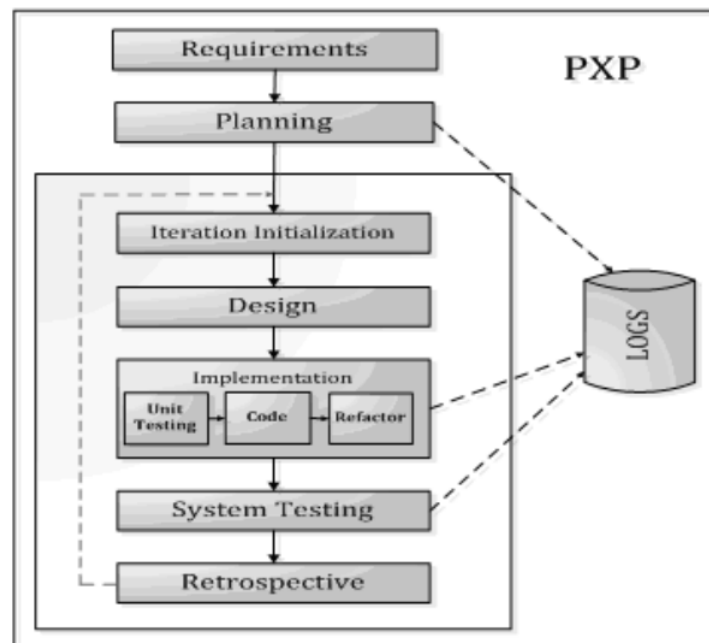


Ilustración 2 - Ciclo de Vida de Extreme Programming (Extreme Programming: A gentle introduction, 2016)

**1. Requerimientos**

Esta etapa tiene como objetivo determinar los requerimientos que debe cumplir el software. Para documentar los requerimientos del sistema, se hizo uso del documento conocido como SRS (por sus siglas en inglés, Software Requirements Specification) en donde se plasmó la metodología de recolección de los requerimientos y la documentación detallada de los mismos con el fin de tener una fuente de consulta a la hora de su implementación. Esta etapa se realizó con base en la información recolectada en la fase de Entendimiento del Problema.

**2. Planeación**

Esta etapa tiene como objetivo realizar un plan de trabajo para el desarrollo del software. En esta etapa se realizó el documento conocido como SPMP (por sus siglas en inglés, Software Project Management Plan) en el cual se plasmó la metodología de desarrollo de este trabajo de grado.

**3. Inicio de Iteración**

En esta etapa se dio inicio formal al desarrollo del software.

**4. Diseño**

Una vez inició, formalmente, la fase de desarrollo de software, la primera actividad realizada fue determinar el mejor diseño a ser implementado. Para ello, se utilizó el documento conocido como SDD (por sus siglas en inglés, Software Design Document) en donde se plasmó todo lo relacionado con el diseño de los componentes que hacen parte de RART.

**5. Implementación**

En esta fase se desarrollaron los requerimientos plasmados en el SRS según el diseño detallado en el SDD y de acuerdo al plan de desarrollo definido en el SPMP.

**6. Pruebas**

En esta etapa se realizaron las pruebas del software desarrollado con el fin de validar su correcto funcionamiento.

- a. Pruebas de casos de uso: estas pruebas consistieron en validar el funcionamiento de cada uno de los casos de uso definidos como parte del desarrollo del SRS.

- b. Pruebas de rendimiento: estas pruebas se realizaron cuantificando el tiempo en que demora el software en ejecutar sus procesos ya sea de extracción o de consulta.
- c. Pruebas de usuario: las pruebas de usuario consistieron en reunir un equipo de personas a las cuales se les enseñó a utilizar el software, para luego determinar la aceptación de estos a la hora de usarlo.

## 7. Retrospectiva

Dada la naturaleza iterativa de la metodología, en esta etapa se regresa a la fase 3 con el fin de complementar los productos resultado de cada etapa. El desarrollo del software RART, producto de este trabajo de grado, fue realizado en tres iteraciones.

Una vez se realizaron con éxito las etapas anteriores se procedió a realizar este documento conocido como las Memorias del trabajo de grado en donde se detalla cómo se realizaron todas las actividades las cuales llevaron a culminar exitosamente del desarrollo del *Framework* RART.

# III – MARCO TEÓRICO

## 1. Marco Contextual

*Data Analytics* es el proceso de examinar los datos con el fin de hallar hechos, relaciones, patrones, revelaciones y/o tendencias. El propósito del *Data Analytics* es apoyar los procesos de toma de decisiones (Arcitura., 2016).

La Minería de datos es un proceso de *Data Analytics* por medio del cual los datos almacenados se convierten en información y finalmente, con la aplicación de algunas técnicas específicas, se extrae conocimiento relevante y no fácilmente detectable de esa información. Existen cuatro técnicas principales las cuales son las más usadas en la Minería de datos (Arcitura., 2016).

### 1. Descriptivas

Las tecinas descriptivas se encargan de responder preguntas acerca de eventos que ya han ocurrido con base en los datos analizados (Arcitura., 2016).

### 2. Diagnósticas

Las técnicas diagnósticas tienen como objetivo determinar la causa de un fenómeno que ocurrió en el pasado, usando preguntas que se enfocan en la razón detrás del evento.

### 3. Predictivas

Las técnicas predictivas tienen como objetivo determinar el resultado de un evento que puede ocurrir en el futuro (Arcitura., 2016). Las predicciones son realizadas con base en patrones, tendencias y excepciones encontradas en los datos actuales e históricos.

### 4. Prescriptivas

Las técnicas prescriptivas están basadas en los resultados de las técnicas predictivas, al indicar acciones que se deberían realizar. Estas técnicas se enfocan en que opción indicada se debe seguir y en por qué y cuándo se debería seguir, con el fin de obtener una ventaja o mitigar un riesgo (Arcitura., 2016).

Con el paso del tiempo la cantidad de datos que deben ser procesados con las distintas técnicas existentes ha aumentado exponencialmente y así mismo se espera que el tiempo de procesamiento sea mínimo, estas exigencias han dado origen a lo que se denomina Big Data (Schroeck *et al*, 2012).

La razón de ser del Big Data es realizar procesos de *Data Analytics* de una forma más rápida, eficiente y a bajo coste teniendo en cuenta las siguientes características (SAS Institute, 2003):

- ✓ **Variedad:** la variedad hace referencia a los diferentes tipos de datos que existen ya sean estructurados, semiestructurados y no estructurados, los cuales se pueden analizar de distintas maneras.
- ✓ **Volumen:** el volumen hace referencia a la gran cantidad de datos que se generan cada día más, lo cual afecta la forma en la que se analiza la información y por ende su rendimiento.
- ✓ **Velocidad:** dada la gran cantidad de datos, estos se deben procesar y analizar de una forma rápida para agilizar la toma de decisiones.
- ✓ **Veracidad:** la veracidad hace referencia a la relevancia de los datos a la hora de realizar un análisis sobre los mismos.



- ✓ **Valor:** hace referencia a la usabilidad que se le puede llegar a dar a la información, entre más veraz sea, mayor valor tendrá.

Para poder aplicar procesos de Big Data es imprescindible contar con información almacenada dentro de un repositorio. Un repositorio es un sitio el cual almacena y permite guardar nueva información.

Por otra parte, Facebook es una de las principales redes sociales existentes y cada instante de tiempo se publica una gran cantidad de información. Una red social es un sitio web donde sus usuarios intercambian comentarios, gustos y/u opiniones respecto a cualquier tema por medio de una publicación (Alsubagh, 2015). La gran cantidad de publicaciones que se generan son una rica fuente de información la cual podría determinar las tendencias y gustos de un segmento de usuarios (Sanz, 2003).

Dentro de las empresas del sector privado que hacen parte de la Alianza CAOBA, se encuentra una empresa productora de alimentos. Esta empresa quiere realizar una segmentación de sus clientes digitales con información proveniente de la red social Facebook. Su interés particular es aplicar técnicas de *Data Analytics* para así determinar los distintos nichos de mercado los cuales dan a conocer sus comportamientos de consumo dentro de dicha red social. Para poder aplicar procesos de *Data Analytics* se debe contar con un repositorio de datos organizado el cual contenga información proveniente de las publicaciones (Pomares, 2016).

Dada la gran cantidad de publicaciones que se generan cada instante de tiempo en Facebook, crear un repositorio no es un proceso trivial. Para ello es necesario extraer la información relacionada a las publicaciones, para luego hacer una limpieza de esta misma y finalmente almacenarla dentro de un repositorio de forma tal que se facilite la aplicación de procesos de *Data Analytics* (Sanz, 2003).

## 5. Marco Conceptual

Dentro de esta sección se presentan los conceptos más importantes a tener presentes para entender este documento y los procesos y actividades llevados a cabo

---

✓ **Red Social**

Una red social es un sitio web donde sus usuarios intercambian comentarios, gustos y/u opiniones respecto a cualquier tema por medio de una publicación realizada ya sea por una persona o una empresa (Comunitarios, 2014).

✓ **Fan Page**

Página web que hace parte de una red social la cual es administrada por una persona o una empresa (Comunitarios, 2014).

✓ **Publicación**

Una publicación es un contenido multimedia compartido por una *fan page* dentro de la red social Facebook la cual puede ser visualizada por los usuarios que siguen la fan page.

✓ **Reacción**

Expresión de un sentimiento percibido por un usuario de la red social Facebook ante una publicación visualizada (Comunitarios, 2014). Facebook permite a sus usuarios dar a las publicaciones los siguientes tipos de reacciones

- **Like:** cuando a un usuario le gusta una publicación.
- **Love:** cuando a un usuario le encanta una publicación.
- **Haha:** cuando a un usuario le divierte una publicación.
- **Wow:** cuando a un usuario le sorprende una publicación.
- **Sad:** cuando a un usuario le entristece una publicación,
- **Angry:** cuando a un usuario le enoja una publicación.

✓ **Comentario**

Opinión dada por un usuario la cual es expresada por medio de un texto en una publicación realizada por una fan page (Comunitarios, 2014).

✓ **API**

Abreviación de Application Programming Interface: conjunto de funcionalidades que ofrece un software para un fin específico. Para el caso de este proyecto, el API utilizado se llama **RestFb** y es el que permite consultar y extraer información existente dentro de la red social Facebook.

## IV – DESARROLLO DE TRABAJO

Esta sección explica qué es RART, el producto de software desarrollado en este trabajo de grado, y cómo fue el proceso de análisis, diseño y desarrollo llevados a cabo para su construcción.

### 1. Vista general del proyecto

#### ¿Qué es RART?

RART es un *Framework* desarrollado en Java el cual tiene como propósito extraer información proveniente de la red social Facebook con el fin de generar un repositorio de datos al cual se le pueda aplicar procesos de *Data Analytics* que permitan hallar nueva información valiosa para los interesados.

#### ¿Por qué un *Framework*?

Un *Framework* o marco de trabajo, es un software desarrollado por componentes que se pueden reemplazar, actualizar, configurar o modificar independientemente según las necesidades específicas sin tener que rehacer todo, modificando lo nuevo y reutilizando los componentes previamente desarrollados (Gutierrez, 2016). Por lo tanto, RART es un software que se desarrolló por componentes, donde cada uno tiene una función específica e independiente de las demás, de tal manera que si se desea modificar un componente se puede realizar sin tener que modificar la totalidad del software.

#### ¿Qué tipo de información se extrae de la red social Facebook?

RART extrae información proveniente de publicaciones realizadas por una o un conjunto de *fan page* existentes dentro de la red social Facebook. Esta información contempla lo siguiente:

1. El id y nombre de la *fan page* que realiza una publicación de interés.
2. El id y mensaje de la publicación de interés.
3. La imagen publicada, en caso de tenerla.
4. Los id y nombres de los usuarios de la red social que comentaron la publicación con sus respectivos mensajes.

5. Los id y nombres de los usuarios de la red social que reaccionaron ante la publicación con su respectivo tipo de reacción.

### ¿Cómo funciona?

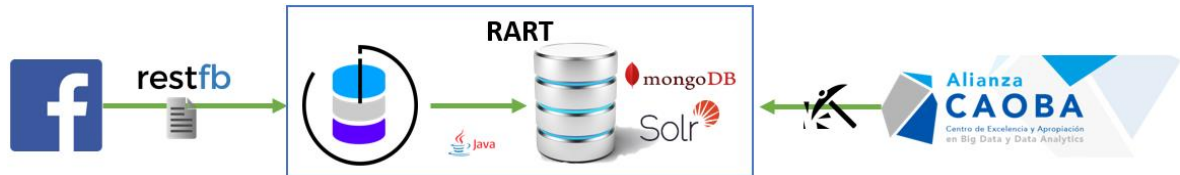


Ilustración 3 - Funcionamiento de RART

Como se muestra en la ilustración 3, RART utiliza un API de Facebook llamado **RestFb** con el fin de extraer información y procesarla para almacenarla dentro de un repositorio de datos el cual permite consultarla de manera fácil y rápida. Esto facilitará los procesos de *Data Analytics* que se vayan a realizar con la información existente dentro del repositorio.

El repositorio de datos consta de dos partes. La primera de una base de datos Mongo DB (Mongo DB, 2016) donde se almacena la información semiestructurada y la segunda de un motor de búsqueda conocido como Apache Solr (Apache, 2016) en donde se almacena la información no estructurada.

## 2. Desarrollo del *Framework* RART

### Definición y levantamiento de requerimientos

El proceso utilizado para definir los requerimientos que RART debe cumplir está descrito en el proceso que se muestra a continuación. Para mayor detalle acerca del proceso ver el anexo [RART Software Requirements Specification](#).

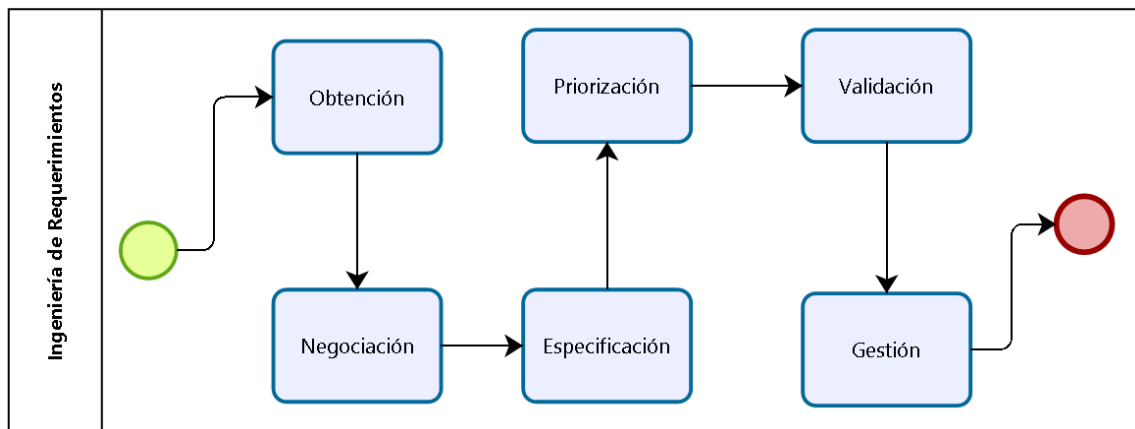


Ilustración 4 - Proceso de Ingeniería de requerimientos

### Obtención de Requerimientos

El proceso de obtención de requerimientos se llevó a cabo luego de haber realizado una entrevista con la directora de este trabajo de grado, quien hace parte de CAOBA, con el fin de especificar las principales funcionalidades del *Framework*. Luego de este primer acercamiento, se realizó la primera especificación de requerimientos para presentarlo a los interesados y obtener una retroalimentación con el propósito de realizar las debidas correcciones. Para este proceso adaptamos algunas de las actividades que propone Sommerville (Kotonya, 1998):

#### Descubrimiento de los requerimientos:

Consiste en interactuar con los *stakeholders* para recopilar sus requerimientos y los requerimientos del modelo de dominio. Aparte de los *stakeholders*, para el equipo de trabajo también se hace necesario estudiar el API de Facebook para determinar la viabilidad de los requerimientos.

### Clasificación y organización de los requerimientos:

Consiste en organizar por grupos todos los requerimientos que se descubrieron en la actividad anterior, los cuales no tienen estructura alguna.

### Negociación de Requerimientos

El proceso de negociación consiste en someter los requerimientos a una prueba de aceptación por parte de los clientes, aquellos requerimientos que no pasan dicha prueba, son redefinidos o en su defecto eliminados del proyecto.

### Especificación de Requerimientos

La plantilla por medio de la cual se realizó la documentación correspondiente a la especificación de requerimientos contiene la siguiente información.

ID Requerimiento		Fecha de Creación	
Versión		Fecha de Modificación	
Tipo de Requerimiento		Casos de Uso Asociados (ID's)	
Requerimientos Predecesores (ID's)		Estado del Rquerimiento	
Autores		Prioridad	
Descripción			
Observaciones	Fecha	Descripción	

Ilustración 5 - Plantilla Especificación de Requerimientos

Esta plantilla permitió tener un desglose de cada requerimiento y así tener una base de conocimiento para consultar a lo largo del desarrollo del proyecto.

Esta especificación también permitió determinar las distintas funcionalidades en las que se agrupan los distintos requerimientos. A continuación, se presentan las categorías identificadas.

#### 1. Ingresar al sistema

Esta categoría cuenta únicamente con un requerimiento asociado y su finalidad es permitir al usuario ingresar al sistema.

## 2. **Buscar dentro de Facebook**

Esta categoría consta de cuatro requerimientos los cuales tienen como finalidad permitirle al usuario buscar información existente dentro de la red social Facebook.

## 3. **Extraer información encontrada en Facebook**

Esta categoría consta de 13 requerimientos cuyo fin es extraer la información de interés del usuario desde la red social Facebook.

## 4. **Visualizar con información proveniente de Facebook**

Esta categoría consta de 13 requerimientos cuyo fin es mostrarle al usuario la información encontrada en Facebook con los criterios que este ingresa.

## 5. **Estructurar**

Esta categoría consta de cuatro requerimientos los cuales tienen como fin estructurar la información extraída desde la red social Facebook de acuerdo con el modelo de la base de datos de RART.

## 6. **Guardar**

Esta categoría consta de cuatro requerimientos cuyo fin es guardar la información extraída y estructurada dentro de la base de datos de RART.

## 7. **Buscar en la base de datos**

Esta categoría consta de 11 requerimientos los cuales tienen como fin brindarle al usuario la funcionalidad de buscar información dentro de la base de datos de RART.

## 8. **Visualizar con información de la base de datos**

Esta categoría consta de 13 requerimientos los cuales tiene como fin mostrarle al usuario la información existente dentro de la base de datos de RART de acuerdo con lo encontrado usando los parámetros ingresados.

### **Priorización de Requerimientos**

De acuerdo con las funcionalidades mencionadas anteriormente su prioridad de implementación se determinó de la siguiente manera:

<b>Funcionalidad</b>	<b>Prioridad</b>
----------------------	------------------

Ingresar al sistema	Alta
Buscar dentro de Facebook	Alta
Extraer información encontrada en Facebook	Alta
Visualizar con información proveniente de Facebook	Media
Estructurar	Alta
Guardar	Alta
Buscar en la base de datos	Media
Visualizar con información de la base de datos	Media

Tabla 2 - Priorización de Requerimientos

### Validación de Requerimientos

El proceso de validación consistió en asegurar que los requerimientos definidos hayan sido precisos, además se examinaron en detalle para detectar inconsistencias, omisiones, errores y finalmente verificar que cumplen con los estándares definidos por el equipo de trabajo.

Para este proceso se utilizó una plantilla la cual contiene 8 preguntas que permitieron garantizar la calidad de la especificación de los requerimientos planteados.

No.	Pregunta
1	¿El requerimiento está establecidos de manera clara? ¿Puede malinterpretarse?
2	¿La fuente del requerimiento está identificada? ¿El requerimiento ha sido revisado por dicha fuente?
3	¿El requerimiento está cuantitativamente restringido?



4	¿En la matriz de trazabilidad están ubicados los requerimientos asociados a él?
5	¿El requisito incumple con alguna de las restricciones del modelo de dominio del sistema?
6	¿El requerimiento se puede probar? ¿Se pueden especificar las pruebas?
7	¿El requerimiento es rastreable en cualquier modelo del sistema que haya sido generado?
8	¿La especificación está estructurada de tal manera que se pueda traducir en productos de trabajo más técnicos?

Tabla 3 - Validación de Requerimientos

### Gestión de Requerimientos

La gestión del cambio de requerimientos ayuda a que los cambios propuestos sean tratados de forma consistente y que los cambios en el documento se realicen de forma controlada. Las principales actividades realizadas fueron:

1. Análisis del problema y especificación del cambio: esto comienza cuando se identifica algún problema en los requerimientos. El problema se analiza para validar su argumento y finalmente los resultados se pasan al solicitante con las observaciones para que realice las correcciones y ahí sí formalice su propuesta.
2. Análisis del cambio y cálculo de costos: mediante el rastreo del requerimiento se puede analizar el efecto del cambio. Dicho cambio se valora en función de las modificaciones al documento de requerimientos, al diseño e implementación. Si el cambio es aprobado es necesario realizar las modificaciones en los artefactos mencionados.
3. Implementación del cambio: se aplica el cambio al documento, diseño e implementación. Se debe procurar no tener que realizar grandes cambios en la organización del documento o tener que redactar una gran cantidad del mismo.

Una vez se aplicó el proceso de levantamiento y definición de requerimientos se obtuvo la especificación detallada de cada uno de ellos.

Funcionalidad	Descripción del Requerimiento
<b>Ingresar al sistema</b>	El <i>Framework</i> permite al usuario ingresar al sistema
<b>Buscar dentro de Facebook</b>	El <i>Framework</i> permite al usuario buscar una página de Facebook por su nombre
	El <i>Framework</i> permite al usuario buscar una página de Facebook por su ID

	El <i>Framework</i> permite al usuario buscar las publicaciones realizadas por una página de Facebook, dado su nombre o ID
	El <i>Framework</i> permite al usuario buscar las publicaciones realizadas por un conjunto de páginas de Facebook, dado sus nombres y/o IDs
<b>Extraer la Información Encontrada en Facebook</b>	El <i>Framework</i> permite extraer el nombre de una o muchas páginas de Facebook
	El <i>Framework</i> permite extraer el ID de una o muchas páginas de Facebook
	El <i>Framework</i> permite extraer el ID de una o muchas publicaciones de Facebook realizadas por una página
	El <i>Framework</i> permite extraer el mensaje de una o muchas publicaciones de Facebook realizadas por una página
	El <i>Framework</i> permite extraer la imagen de una o muchas publicaciones de Facebook realizadas por una página
	El <i>Framework</i> permite extraer los IDs de los comentarios realizados en un o muchas publicaciones
	El <i>Framework</i> permite extraer los mensajes de los comentarios realizados en una publicación
	El <i>Framework</i> permite extraer los IDs de los usuarios de Facebook que comentaron una publicación
	El <i>Framework</i> permite extraer los nombres de los usuarios que comentaron una publicación
	El <i>Framework</i> permite extraer la cantidad de reacciones obtenidas por una publicación
	El <i>Framework</i> permite extraer el tipo de reacciones obtenidas por una publicación
	El <i>Framework</i> permite extraer el ID del usuario de Facebook que reaccionó ante una publicación
	El <i>Framework</i> permite extraer el nombre del usuario de Facebook que reaccionó ante una publicación
<b>Visualizar con Información</b>	El <i>Framework</i> permite al usuario visualizar el nombre de la(s) página(s) buscada(s) en Facebook dado el id o nombre de una página

<b>Proveniente de Facebook</b>	El <i>Framework</i> permite al usuario visualizar el(los) ID(s) de la(s) página(s) buscada(s) en Facebook dado un nombre de página
	El <i>Framework</i> permite al usuario visualizar el(los) mensaje(s) de la(s) publicación(s) de Facebook dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar el ID de la(s) publicación(s) de Facebook dado el id o nombre de una página o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar la(s) imágenes de la(s) publicación(s) de Facebook dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar los mensajes de los comentarios realizados en una publicación de Facebook dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar los IDs de los comentarios realizados en una publicación de Facebook dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar los nombres de los usuarios de Facebook que realizaron un comentario en una publicación dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar los IDs de los usuarios de Facebook que realizaron un comentario en una publicación dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar la cantidad de reacciones obtenidas por una publicación de Facebook dado el id de una publicación o una palabra contenida en el mensaje de la publicación
	El <i>Framework</i> permite al usuario visualizar el tipo de reacciones obtenidas por una publicación de Facebook dado el id de una publicación o una palabra contenida en el mensaje de la publicación

	<p>El <i>Framework</i> permite al usuario visualizar el ID del usuario de Facebook que reaccionó a una publicación dado el id de una publicación o una palabra contenida en el mensaje de la publicación</p> <p>El <i>Framework</i> permite al usuario visualizar el nombre del usuario de Facebook que reaccionó a una publicación dado el id de una publicación o una palabra contenida en el mensaje de la publicación</p>
<b>Estructurar</b>	<p>El <i>Framework</i> permite estructurar la información de Nombre y ID de una página de Facebook de acuerdo a la estructura establecida para almacenar dentro de la base de datos</p> <p>El <i>Framework</i> permite estructurar la información de Nombre, ID, mensaje e imagen de una publicación de Facebook de acuerdo a la estructura establecida para almacenar dentro de la base de datos</p> <p>El <i>Framework</i> permite estructurar la información de ID, mensaje, ID de usuario y nombre de usuario de un comentario realizado en una publicación por un usuario de Facebook de acuerdo a la estructura establecida para almacenar dentro de la base de datos</p> <p>El <i>Framework</i> permite estructurar la información de tipo de reacción, ID de usuario y nombre de usuario de una reacción realizada en una publicación por un usuario de Facebook de acuerdo a la estructura establecida para almacenar dentro de la base de datos</p>
<b>Guardar</b>	<p>El <i>Framework</i> permite guardar la información de Nombre y ID de una página de Facebook dentro de la base de datos</p> <p>El <i>Framework</i> permite guardar la información de Nombre, ID, mensaje e imagen de una publicación de Facebook dentro de la base de datos</p> <p>El <i>Framework</i> permite guardar la información de ID, mensaje, ID de usuario y nombre de usuario de un comentario realizado en una publicación por un usuario de Facebook dentro de la base de datos</p> <p>El <i>Framework</i> permite guardar la información de tipo de reacción, ID de usuario y nombre de usuario de una reacción realizada en una publicación por un usuario de <i>Facebook</i> dentro de la base de datos</p>

<b>Consultar la Base de Datos</b>	El <i>Framework</i> permite buscar una página por su nombre
	El <i>Framework</i> permite buscar una página por su ID
	El <i>Framework</i> permite buscar un conjunto de páginas dada una palabra
	El <i>Framework</i> permite buscar publicaciones dada una palabra
	El <i>Framework</i> permite buscar publicaciones dada una palabra y periodo de tiempo
	El <i>Framework</i> permite buscar publicaciones de una página
	El <i>Framework</i> permite buscar publicaciones de una página dada una palabra
	El <i>Framework</i> permite buscar publicaciones de una página dada una palabra y un periodo de tiempo
	El <i>Framework</i> permite buscar publicaciones de un conjunto de páginas dados sus IDs
	El <i>Framework</i> permite buscar publicaciones de un conjunto de páginas dada una palabra y los IDs de las páginas
El <i>Framework</i> permite buscar publicaciones de un conjunto de páginas dada una palabra, un periodo de tiempo y los IDs de las páginas	
<b>Visualizar con información de la Base de Datos</b>	El <i>Framework</i> permite al usuario visualizar el nombre de la(s) página(s) buscada(s) dentro de la base de datos
	El <i>Framework</i> permite al usuario visualizar el(los) ID(s) de la(s) página(s) buscada(s) dentro de la base de datos
	El <i>Framework</i> permite al usuario visualizar el(los) mensaje(s) de la(s) publicación(s) existentes dentro de la base de datos
	El <i>Framework</i> permite al usuario visualizar el ID de la(s) publicación(s) de Facebook
	El <i>Framework</i> permite al usuario visualizar la(s) imágenes de la(s) publicación(s) existentes dentro de la base de datos
	El <i>Framework</i> permite al usuario visualizar los mensajes de los comentarios existentes dentro de la base de datos
	El <i>Framework</i> permite al usuario visualizar los IDs de los comentarios existentes dentro de la base de datos

El <i>Framework</i> permite al usuario visualizar los nombres de los usuarios de Facebook que realizaron un comentario en una publicación existente dentro de la base de datos
El <i>Framework</i> permite al usuario visualizar los IDs de los usuarios de Facebook que realizaron un comentario en una publicación existente dentro de la base de datos
El <i>Framework</i> permite al usuario visualizar la cantidad de reacciones obtenidas por una publicación existente dentro de la base de datos
El <i>Framework</i> permite al usuario visualizar el tipo de reacciones obtenidas por una publicación de Facebook existente dentro de la base de datos
El <i>Framework</i> permite al usuario visualizar el ID del usuario de Facebook que reaccionó a una publicación existente dentro de la base de datos
El <i>Framework</i> permite al usuario visualizar el nombre del usuario de Facebook que reaccionó a una publicación existente dentro de la base de datos

Tabla 4 - Descripción de Requerimientos

### Definición de casos de uso

Una vez documentados los requerimientos del software de definieron los casos de uso. En la ilustración 6 se muestra el diagrama UML correspondiente a los casos de uso definidos para RART y una descripción de cada uno. La definición de casos de uso de encuentra en el anexo [RART Software Requirements Specification](#).

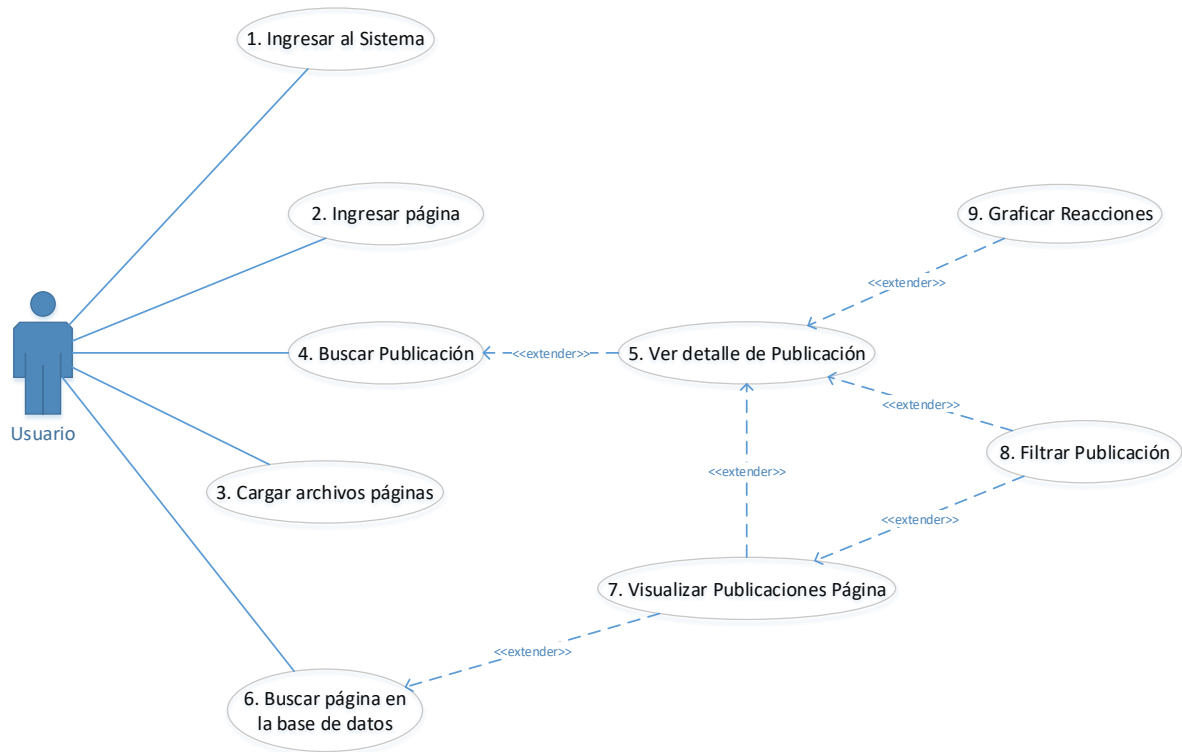


Ilustración 6 - Diagrama Casos de Uso

### 1. Ingresar al sistema

Esta funcionalidad permite al usuario ingresar al sistema con el fin de realizar tareas de extracción y consulta de información.

### 2. Ingresar página

Esta funcionalidad permite al usuario ingresar una página o el id de una página existente dentro de la red social Facebook con el fin de extraer la información correspondiente a la página y sus publicaciones.

### 3. Cargar archivos página

Esta funcionalidad permite al usuario ingresar al sistema un archivo formato .TXT con una lista de nombres o ID's de páginas existentes dentro de la red social Facebook con el fin de extraer la información correspondiente a la página y sus publicaciones.

#### **4. Buscar publicación**

Esta funcionalidad permite al usuario buscar una publicación dentro de la base de datos del sistema.

#### **5. Ver detalle de publicación**

Esta funcionalidad permite al usuario ver el detalle de una publicación previamente buscada y seleccionada por el usuario.

#### **6. Buscar página en la base de datos**

Esta funcionalidad permite al usuario buscar una página dentro de la base de datos, la cual haya sido guardada previamente.

#### **7. Visualizar publicaciones página**

Esta funcionalidad permite al usuario visualizar las publicaciones realizadas por una página, luego de haberla buscado dentro de la base de datos del sistema.

#### **8. Filtrar publicaciones**

Esta funcionalidad permite al usuario filtrar las publicaciones arrojadas por el sistema de acuerdo con los criterios ingresados por el usuario.

#### **9. Graficar Reacciones**

Esta funcionalidad permite al usuario visualizar gráficamente el número de reacciones obtenidas por una publicación a través del tiempo.



## **Definición de las tecnologías a utilizar**

### **1. RestFb API**

Después de haber analizado distintos APIs existentes los cuales permiten extraer información de la red social Facebook, se seleccionó RestFb dado que se acopló al alcance definido, es decir, permite extraer la información determinada. Además, este API está soportado directamente por la red social.

### **2. Java**

Se seleccionó Java como el lenguaje de desarrollo de RART dado que el API elegido está escrito en este mismo lenguaje. Además, Java es el lenguaje con el cual los autores de este trabajo de grado están más familiarizados y en el que tienen más experiencia.

### **3. Mongo DB**

En el análisis realizado en la etapa de entendimiento de los datos se determinó que la información a extraer será de dos tipos. El primero de ellos es información semiestructurada la cual se almacena dentro de Mongo DB.

### **4. Apache Solr**

El otro tipo de información a guardar es no estructurado. Se seleccionó Apache Solr con el fin de guardar este tipo de información dado que este motor de búsqueda realiza un índice con la información almacenada por palabras lo cual facilita la búsqueda de información.

## Diseño del sistema

### Arquitectura del sistema

Una vez definidos los requerimientos de RART y casos de uso a implementar, se realizó el diseño arquitectural de la solución. En la ilustración 7 se muestra la arquitectura desarrollada la cual se basa en un modelo arquitectural por capas, en donde cada capa le brinda servicios a la capa que está inmediatamente a su derecha. El proceso desarrollado para el diseño de la arquitectura y diseño de RART se encuentra en el Anexo RART Software Design Document.

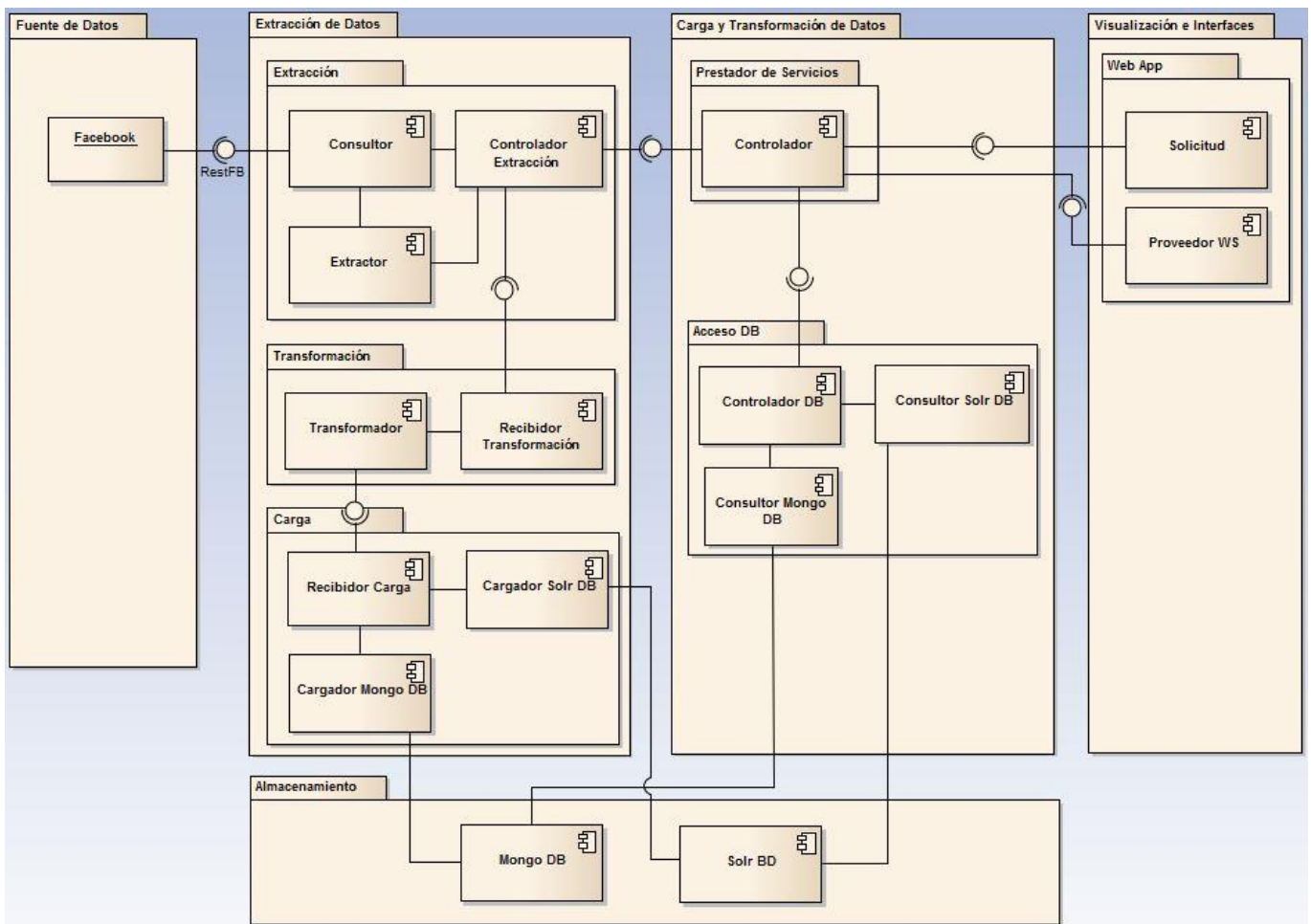


Ilustración 7 - Diagrama de Arquitectura y Componentes de RART

En la ilustración 7 también se muestran los componentes existentes dentro de cada capa de la arquitectura de RART y las conexiones existentes entre ellos. Este diagrama fue elaborado según los estándares de UML.

A continuación, se describen las capas que hacen parte de la arquitectura de RART y sus funcionalidades.

## 1. Fuente de Datos

Esta es la capa que corresponde a la fuente de datos de donde se va a adquirir información, para el caso de RART, Facebook es la única fuente de datos de la cual se alimentará el sistema.

## 2. Extracción de Datos

Dentro de esta capa se realiza un proceso ETL el cual consiste en extraer la información de la fuente, transformarla de acuerdo con el modelo de la base de datos y realizar el debido proceso de persistencia. Dentro de esta capa se encuentran los siguientes componentes:

- a. Extracción: este componente se encarga de realizar el proceso de extracción de información de Facebook por medio del API. Consta de los siguientes componentes:
  - I. Controlador extracción: componente encargado de controlar los procesos de búsqueda y extracción de información de Facebook.
  - II. Consultor: componente encargado de hacer consultas y búsquedas de información dentro de Facebook.
  - III. Extractor: componente encargado de realizar extracción masiva de información de Facebook. Además, este componente proporciona los datos en bruto al componente de transformación.
- b. Transformación: este componente se encarga de transformar la información extraída de acuerdo con el modelo de la base de datos para guardarla dentro de la base de datos.
  - I. Recibidor Transformación: este componente es el encargado de recibir la información extraída de Facebook en bruto para luego transformarla.
  - II. Transformador: este componente recibe la información organizada, realiza el proceso de transformación y la envía al componente encargado de la carga.
- c. Carga: esta capa se encarga de realizar el proceso de guardar la información dentro de las dos bases de datos existentes. Consta de los siguientes componentes:
  - I. Recibidor Carga: componente encargado de recibir la información transformada de acuerdo con el modelo de la base de datos.

- II. Cargador Mongo DB: componente encargado de realizar la persistencia de la información semiestructurada dentro de la base de datos Mongo.
- III. Cargador Solr DB: componente encargado de realizar la persistencia de la información NO estructurada dentro de la base de datos Solr.

### 3. Carga y Transformación de Datos

Esta capa tiene dos servicios principales, el primero es brindar y la segunda es acceder a la base de datos para realizar tareas de consulta de información dentro de la misma. Dentro de esta capa se encuentran los siguientes componentes:

- a. Prestador de Servicios: esta capa es la encargada de brindar los servicios tanto de extracción como de consulta de información. Consta del siguiente componente:
  - I. Controlador: este componente es el encargado de recibir las peticiones de extracción de consulta y dirigir las al componente encargado de cada tarea.
- b. Acceso DB: este componente es el encargado de consultar información dentro de la base de datos. Consta de los siguientes componentes:
  - I. Controlador DB: este componente es el encargado de recibir las peticiones de consulta y además realizar el proceso de orquestar la extracción de información en las bases de datos.
  - II. Consultor Mongo DB: este componente es el encargado de extraer información dentro de la base de datos Mongo de acuerdo con los parámetros indicados por el componente controlador DB.
  - III. Consultor Solr DB: este componente es el encargado de extraer información dentro de la base de datos Solr de acuerdo con los parámetros indicados por el componente controlador DB.

### 4. Almacenamiento

Esta capa brinda el almacenamiento de la información extraída por el *Framework* RART y es transversal a las capas de Extracción de Datos y Visualización e Interfaces debido a que estas guardan y consultan información en la base de datos respectivamente.

- a. Mongo DB: dentro de esta base de datos se almacena la información semiestructurada previamente extraída.
- b. SOLR DB: dentro de este motor de búsqueda se realiza el almacenamiento de la información NO estructurada que tiene relación con la información semiestructurada de la base de datos Mongo.

## 5. Visualización e Interfaces

Esta capa brinda dos servicios. El primero son los servicios que se le brindan al usuario por medio de la interfaz gráfica para que éste pueda realizar tanto procesos de extracción como procesos de consulta de información. Dentro de esta capa se encuentran los siguientes componentes:

- a. Web App: componente encargado de proporcionar las interfaces necesarias para que el usuario pueda realizar los procesos de consulta y extracción de información. Consta del siguiente componente:
  - I. Solicitud: este componente recibe las peticiones de extracción y consulta por parte del usuario y es quien finalmente le proporciona al usuario una respuesta de su solicitud.
  - II. JSON WS: este componente proporciona a otros sistemas una interfaz la cual permite acceder a las funcionalidades, de extracción y consulta de RART, por medio de servicios web.

## Persistencia

La persistencia de la información se realizó en dos bases de datos como se muestra en la ilustración 8.

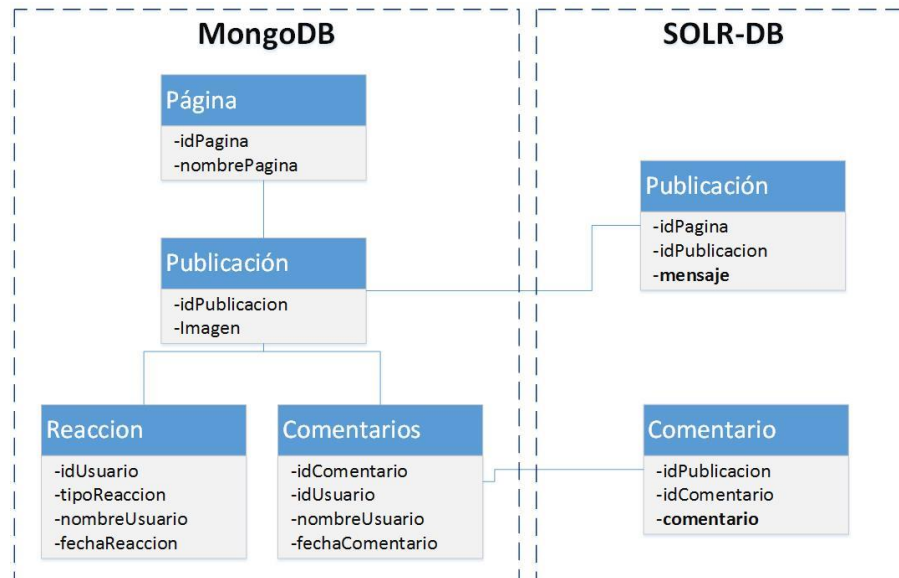


Ilustración 8 - Estructura de los Archivos y Base de Datos

### 1. Mongo DB

Dentro de esta base de datos se almacenan los archivos que contienen información semiestructurada como se describe a continuación:

✓ **Página:**

1. idPagina: identificador de la *fan page* dentro de la red social Facebook.
2. nombrePagina: nombre con el que se conoce la *fan page* dentro de la red social Facebook.

✓ **Publicación:** este archivo contiene la información correspondiente a una publicación realizada por una *fan page* de la red social Facebook

1. idPublicacion: identificador de la publicación dentro de la red social Facebook.
2. Imagen: imagen adjunta en la publicación (en caso de tenerla).

✓ **Reacción**

1. idUsuario: identificador del usuario dentro de la red social Facebook.

2. tipoReacción: tipo de la reacción obtenida en la publicación: Like, Sad, Wow, Angry, Love, Haha.
3. nombreUsuario: nombre del usuario que reaccionó ante la publicación.
4. fechaReacción: fecha en la que se generó la reacción.

✓ **Comentarios**

1. idComentario: identificador correspondiente al comentario realizado en una publicación.
2. idUsuario: identificador del usuario que realizó el comentario en una publicación.
3. nombreUsuario: nombre del usuario que realizó un comentario en la publicación.
4. fechaComentario: fecha en la que el usuario realizó el comentario en la publicación.

## 2. Solr

Dentro de esta base de datos se almacena la información no estructurada que corresponde al mensaje escrito por una *fan page* en una publicación realizada y a los comentarios realizados por los usuarios en las publicaciones:

✓ **Publicación**

1. idPagina: identificador de la *fan page* dentro de la red social Facebook.
2. idPublicación: identificador de la publicación dentro de la red social Facebook.
3. mensaje: texto que hace parte de la publicación realizada por la *fan page*.

✓ **Comentario**

1. idPublicacion: identificador de la publicación dentro de la red social Facebook.
2. idComentario: identificador correspondiente al comentario realizado en una publicación.
3. comentario: texto que corresponde a un comentario realizado por un usuario en una publicación.

## Diagrama de Despliegue

Para la arquitectura planteada para RART, se propone realizar el despliegue en una arquitectura que consta de 4 nodos como se muestra en la ilustración 9.

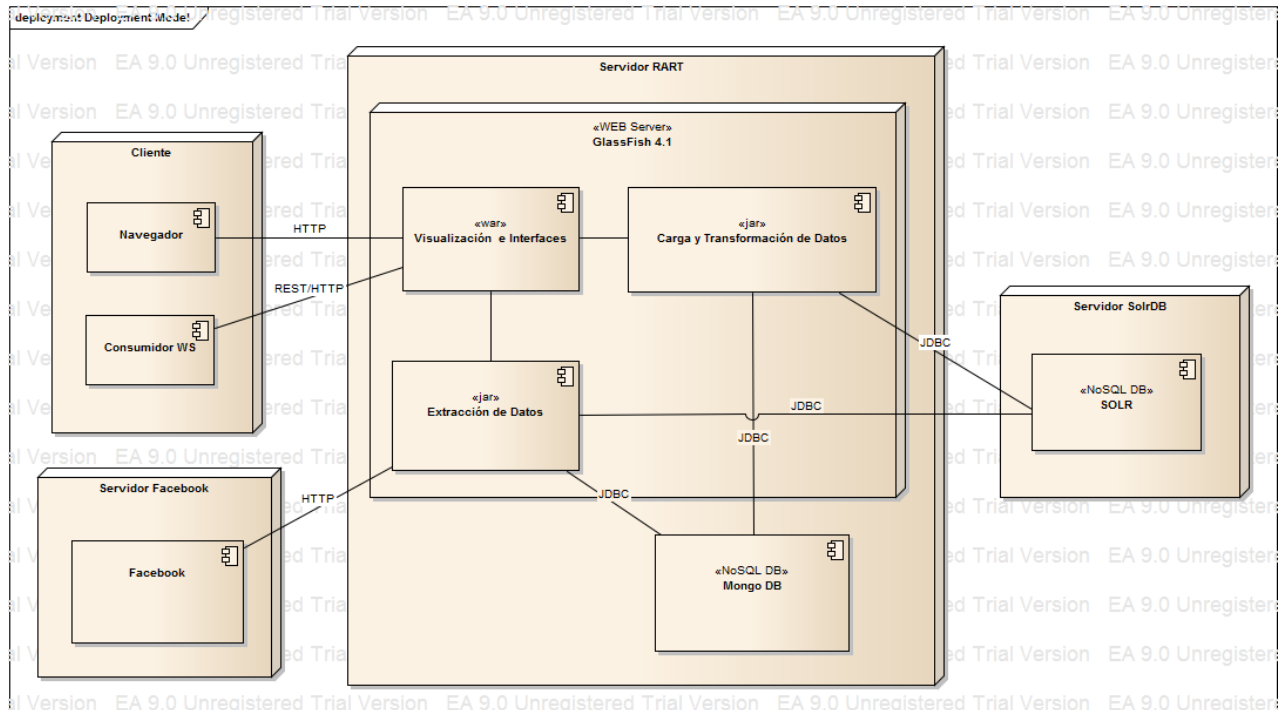


Ilustración 9 - Diagrama de Despliegue (1)

Este modelo de despliegue consta de los siguientes nodos:

1. **Servidor Facebook:** este nodo corresponde al servidor donde se encuentra alojado Facebook al cual se tiene acceso por el API llamado RestFb y contiene el siguiente componente el cual corresponde a la capa de Fuente de Datos dentro de la arquitectura:
  - a. **Facebook:** Facebook es la aplicación correspondiente a la red social desde la cual se extrae información.
2. **Cliente:** este nodo corresponde al computador desde el cual el usuario del sistema ingresa para realizar procesos de extracción y consulta de información. Contiene los siguientes componentes:
  - a. **Navegador:** corresponde a la aplicación en la cual se mostrará al usuario la interfaz gráfica del *Framework*.



- b. **Consumidor WS:** corresponde a la interfaz de un sistema el cual consume, por medio de servicios web, los servicios que ofrece RART para extraer y consultar información.
- 3. **Servidor RART:** corresponde al servidor de producción en el cual se ejecuta RART. Consta de los siguientes componentes:
  - a. **GlassFish 4.1:** este nodo corresponde al servidor de aplicaciones dentro del cual se realizará el despliegue de los siguientes componentes de la aplicación. Estos corresponden a las capas de la arquitectura lógica anteriormente mencionadas:
    - i. Visualización e Interfaces
    - ii. Extracción de Datos
    - iii. Carga y Transformación de Datos
  - b. **Mongo DB:** corresponde a la base de datos encargada de almacenar la información semiestructurada.
- 4. **Servidor Solr DB:** este es un nodo dedicado el cual contiene la base de datos SOLR y contiene el siguiente componente:
  - a. **SOLR:** corresponde a la base de datos encargada de almacenar la información NO estructurada.

Otra posible arquitectura propuesta para realizar el despliegue de RART se muestra en la ilustración 10 y varía en que el motor de búsqueda SolrDB está inmerso dentro del nodo **Servidor RART**.

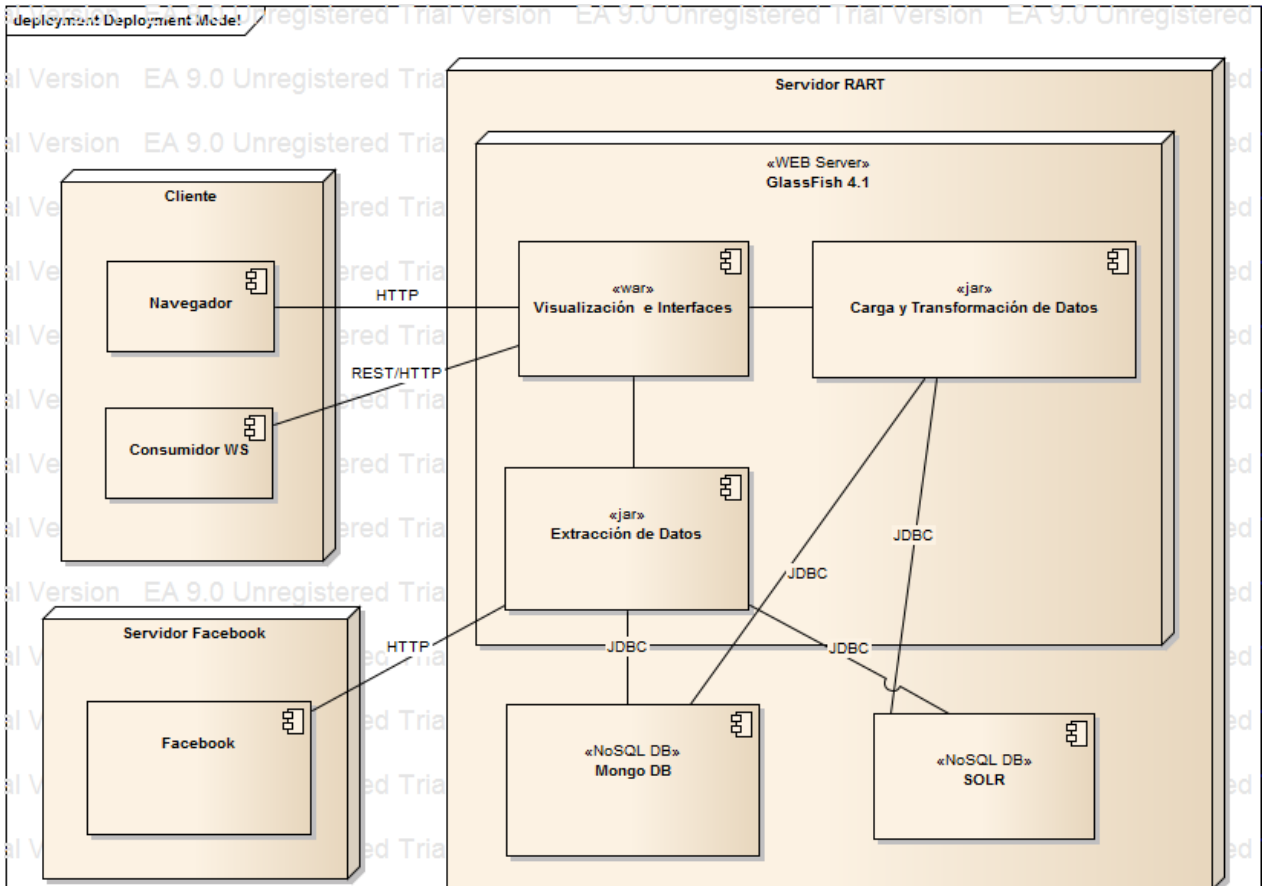


Ilustración 10 - Diagrama de Despliegue (2)

## Diagramas de Secuencia

Una vez definida la arquitectura del sistema y el modelo de persistencia de la información se modelaron, con estándares UML, los diagramas de secuencia correspondientes a los casos de uso definidos. La información detallada de los diagramas se encuentra en el Anexo RART Software Design Document.

### 1. Ingresar al sistema

Este diagrama describe el componente principal con el que usuario interactúa para ingresar al sistema de manera exitosa. Este diagrama se muestra en la ilustración 11.

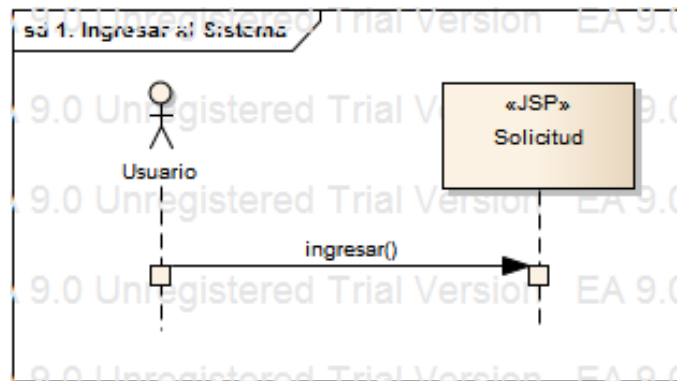


Ilustración 11 - Diagrama de Secuencia Ingresar al Sistema

### 2. Ingresar página

Este diagrama describe cómo interactúan los componentes del sistema cuando el usuario ingresa el nombre o id de una página para extraer la información respectiva de Facebook. Este caso de uso tiene dos escenarios:

- ✓ Cuando el usuario ingresa el id de una página para buscarla y el sistema extrae la información correspondiente a la página del id ingresado. Este diagrama se muestra en la ilustración 12. Para mayor precisión de este diagrama consulte el anexo [RART Software Design Document](#).

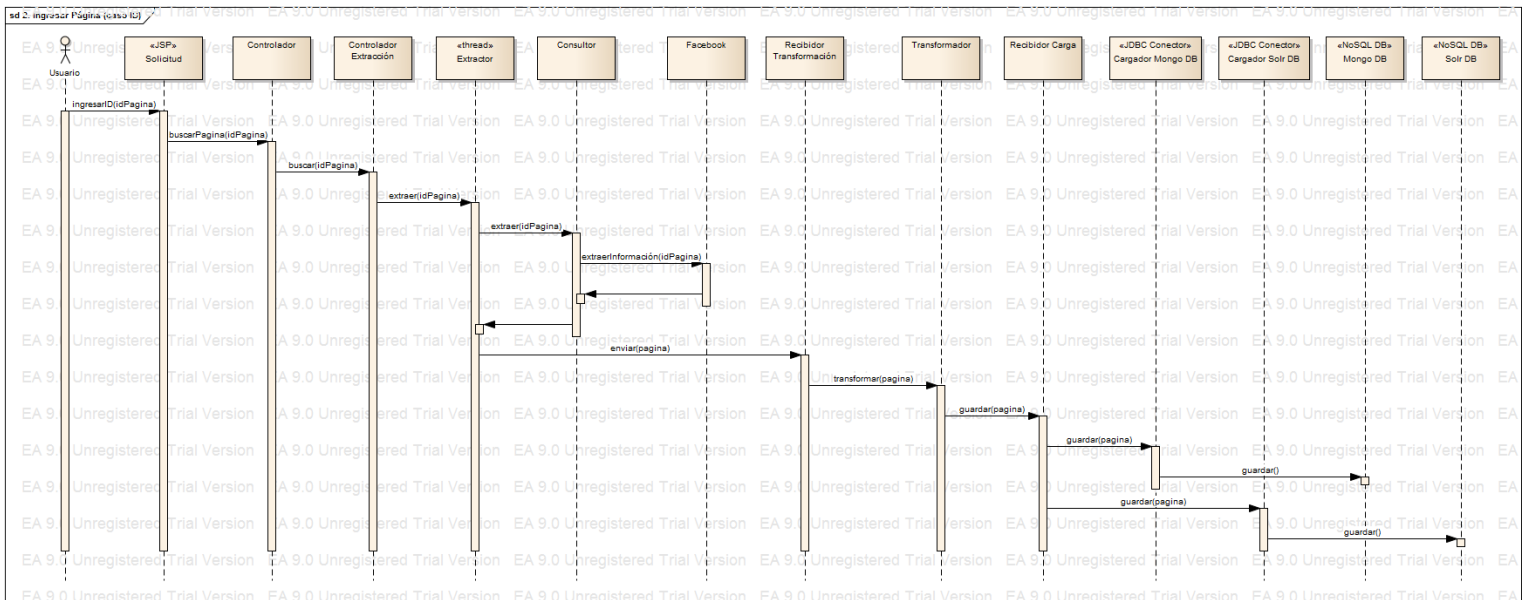


Ilustración 12 - Diagrama de Secuencia Ingresar Página (id).

- ✓ Cuando el usuario ingresa una palabra que corresponde o está contenida en el nombre de la página. En este caso el sistema busca dentro de Facebook las páginas en cuyo nombre se encuentra la palabra ingresada. Este diagrama se muestra en la ilustración 13. Para mayor precisión de este diagrama consulte el anexo [RART Software Design Document](#).

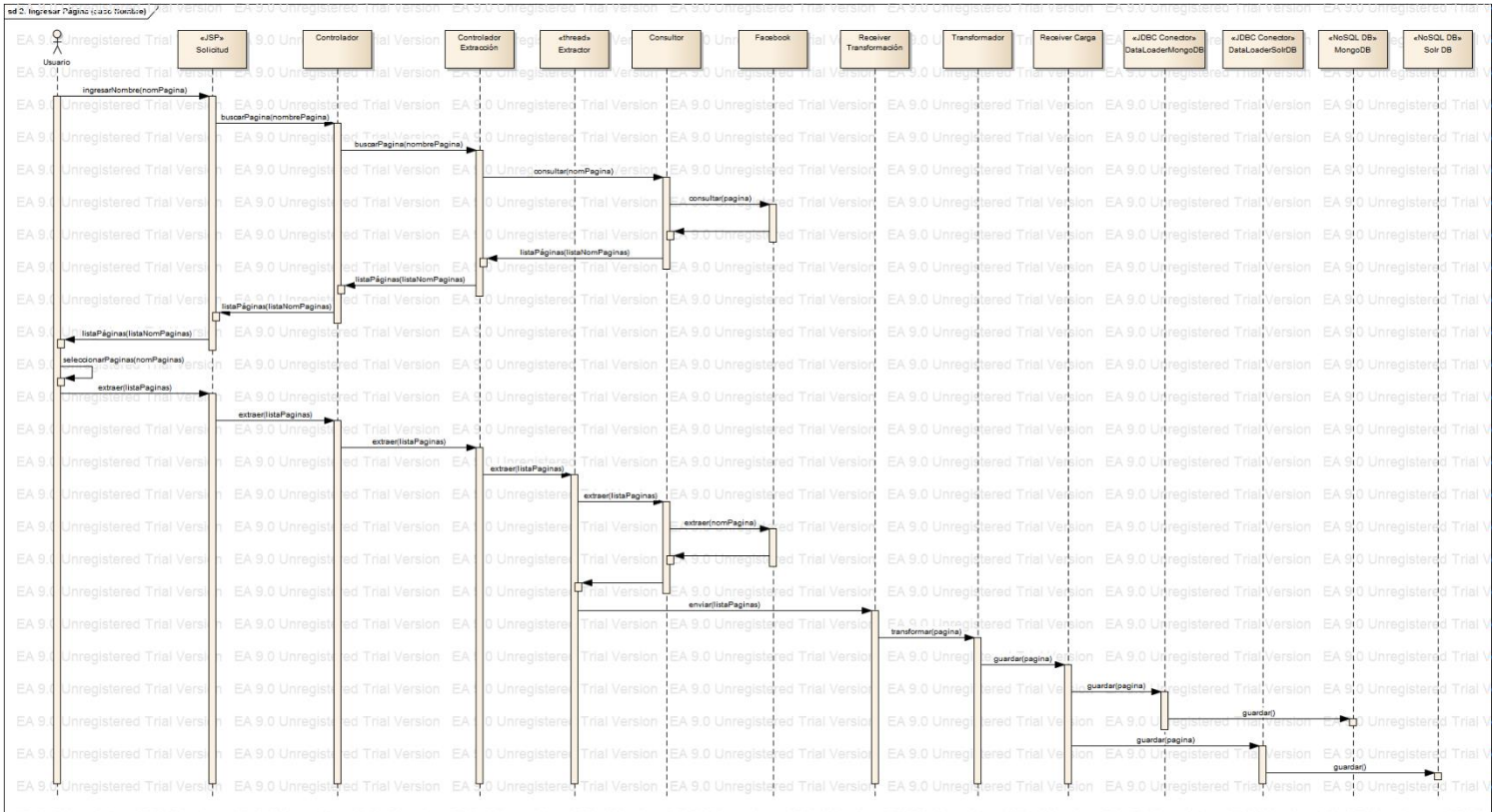


Ilustración 13 - Diagrama de Secuencia Ingresar Página (nombre)

### 3. Cargar archivo páginas

Este diagrama es muy similar al del caso de uso anterior descrito en el diagrama de la ilustración 13. La diferencia es que el usuario no ingresa la información de una sola *fan page*, sino que ingresa un archivo TXT con la información de un conjunto de *fan page*, ya sean ids o nombres.

### 4. Buscar publicación

Este diagrama muestra la interacción de los componentes cuando el usuario busca información acerca de una *fan page* dentro de la base de datos del sistema. Este diagrama se muestra en la ilustración 14.

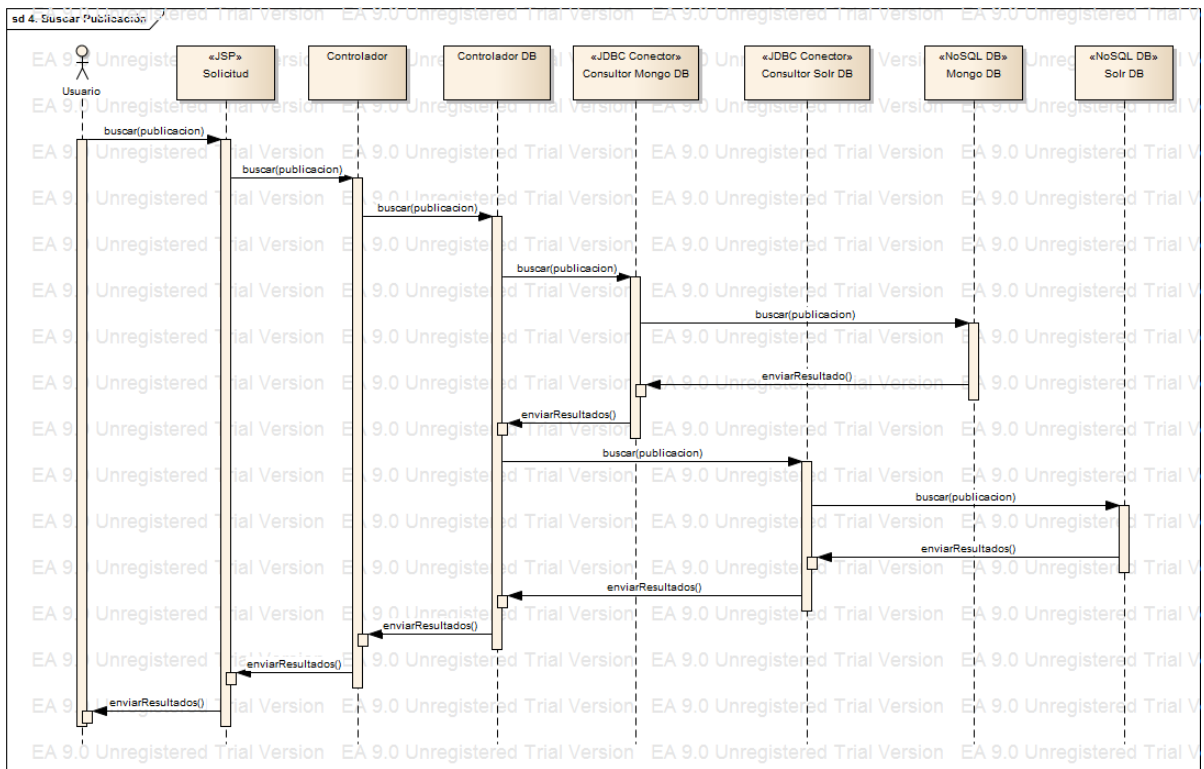


Ilustración 14 - Diagrama de Secuencia Buscar Publicación

### 5. Ver detalle publicación

Este diagrama muestra cómo interactúan los componentes del sistema involucrados en el caso de uso cuando un usuario quiere visualizar el detalle de una publicación. Este diagrama se muestra en la ilustración 15.

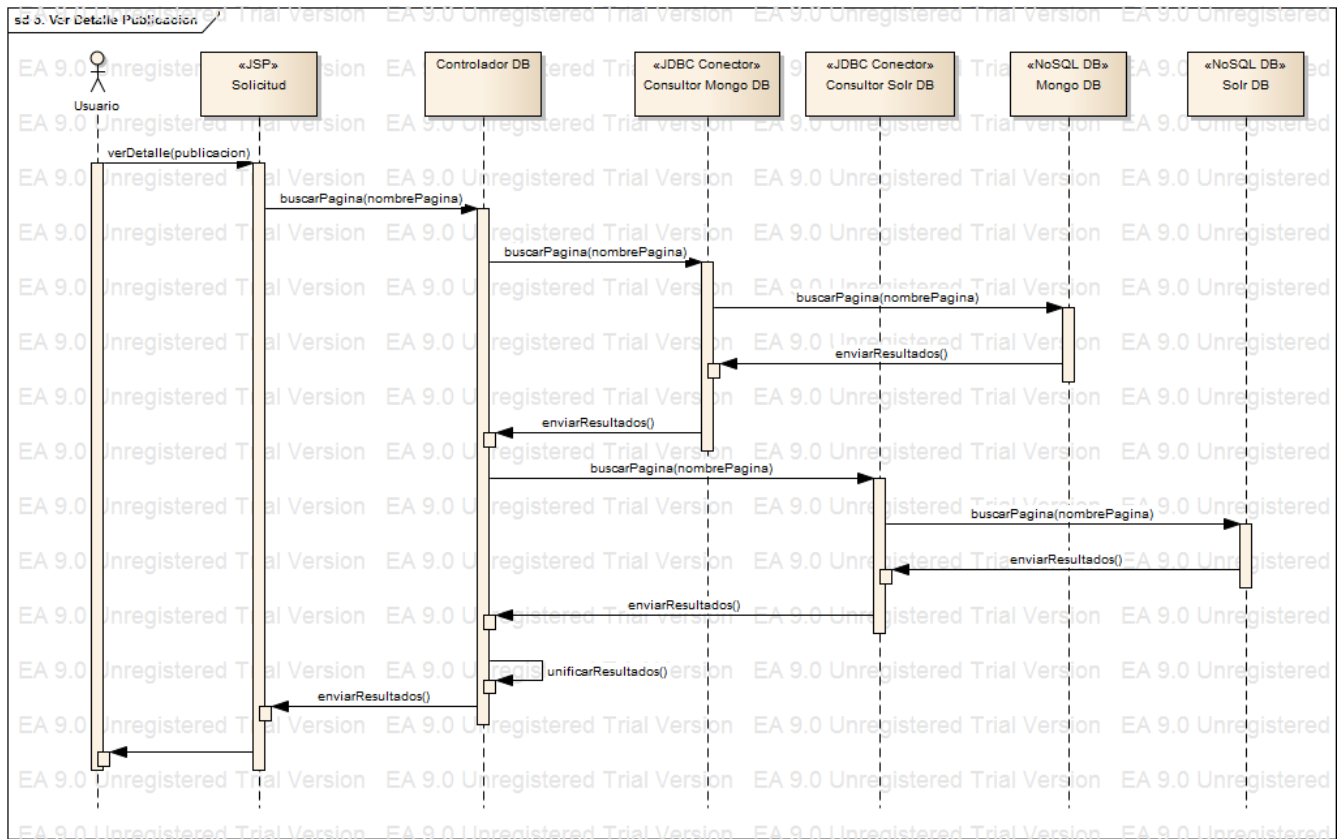


Ilustración 15 - Diagrama de Secuencia Ver Detalle Publicación

## 6. Buscar página en la base de datos

Este diagrama muestra la interacción de los componentes encargados de consultar información dentro de la base de datos cuando el usuario desea ver las publicaciones existentes dentro de la misma. Este diagrama se muestra en la ilustración 16.

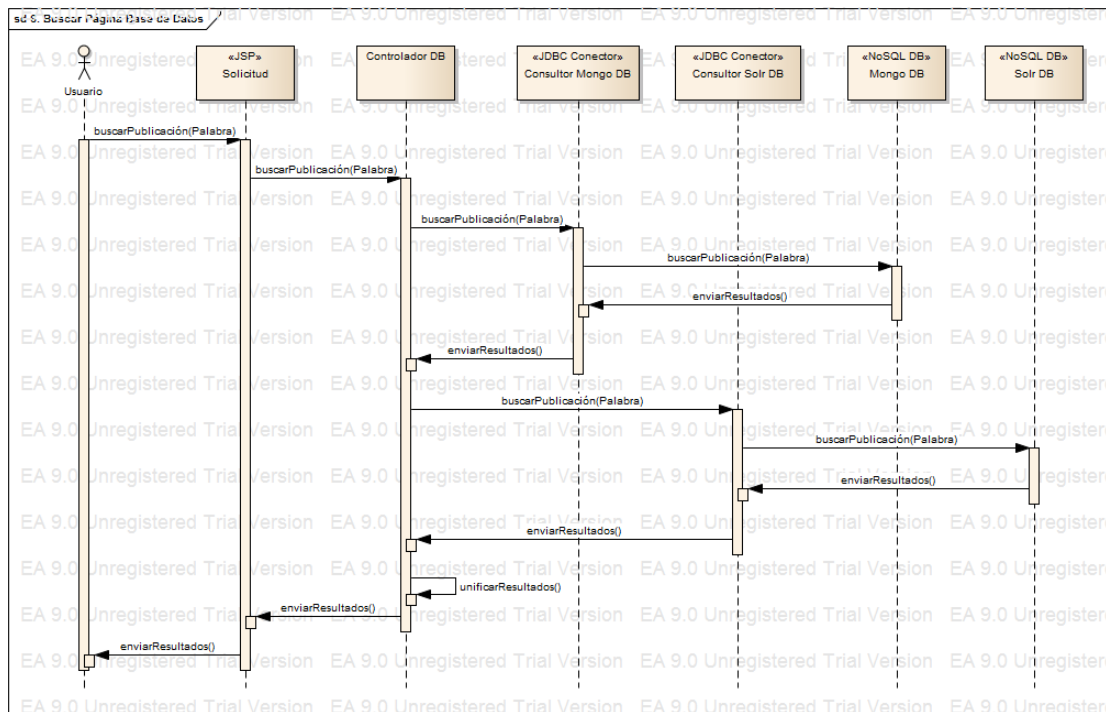


Ilustración 16 - Diagrama de Secuencia Buscar Página en la Base de Datos

## 7. Visualizar publicaciones página

Este diagrama muestra la interacción de los componentes cuando el usuario desea visualizar las publicaciones de una página previamente buscada. Este diagrama se muestra en la ilustración 17.



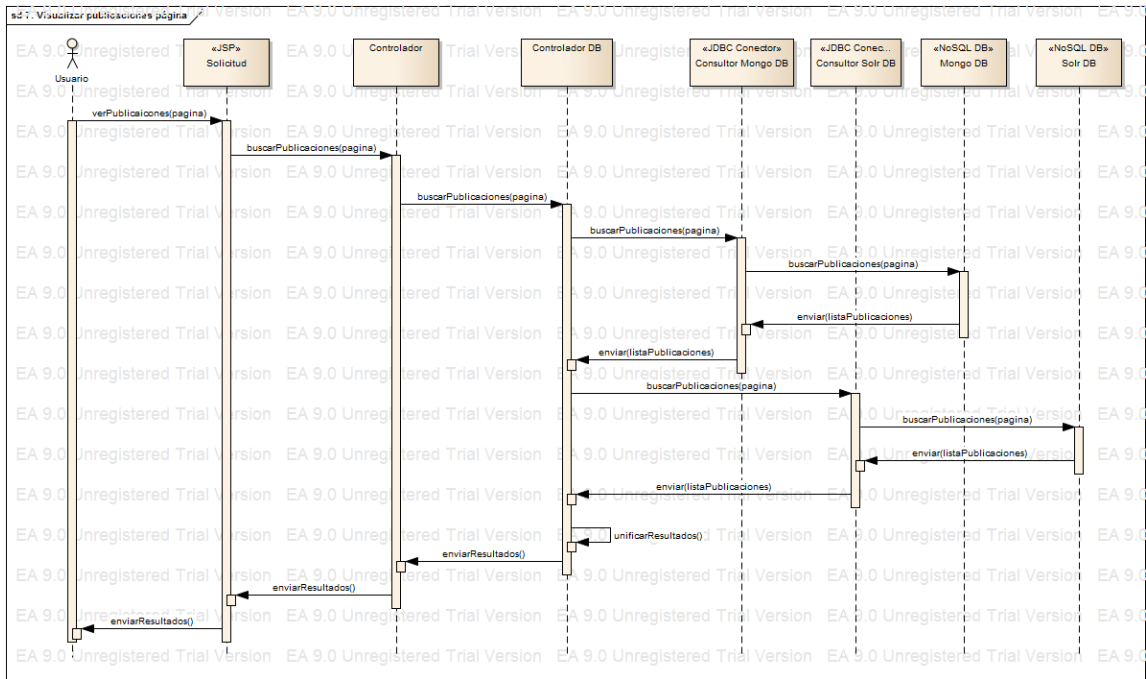


Ilustración 17 - Visualizar publicaciones página

### 8. Filtrar publicación

Este diagrama muestra los componentes que interactúan cuando el usuario desea filtrar, por ciertos parámetros, la información de un conjunto de publicaciones previamente buscadas. Este diagrama se muestra en la ilustración 18.

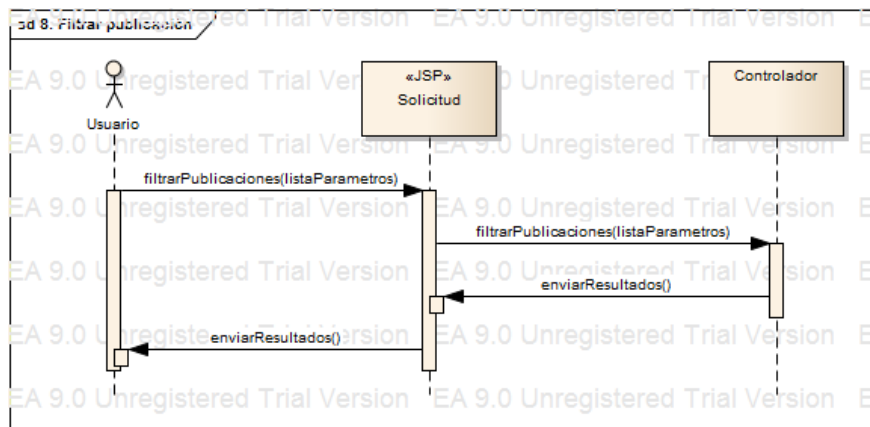


Ilustración 18 - Diagrama de Secuencia Filtrar publicación

## 9. Graficar reacciones

Este diagrama da a conocer los componentes necesarios para graficar las reacciones de una publicación escogida por el usuario. Este diagrama se muestra en la ilustración 19.

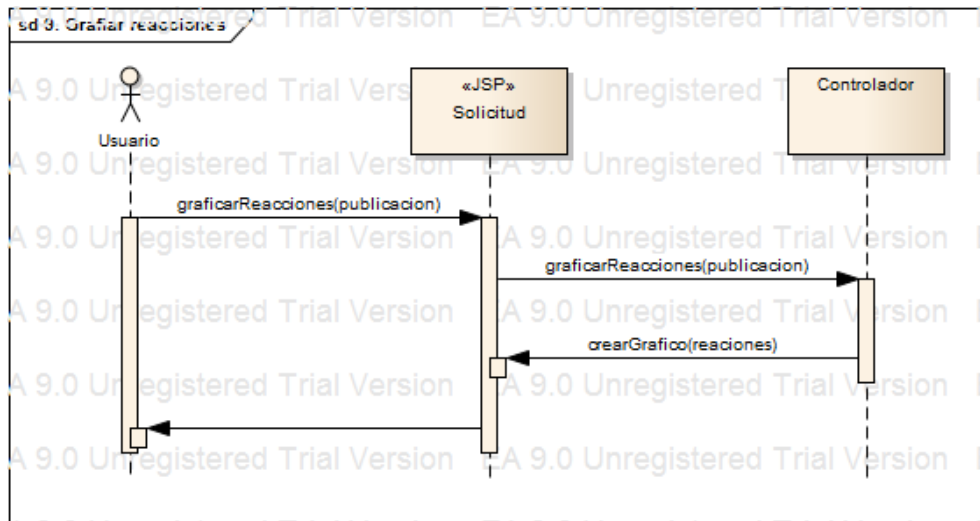


Ilustración 19 - Diagrama de Secuencia Graficar Reacciones

## Procesos del Sistema

A continuación, se presentan los procesos llevados a cabo por RART con el objetivo de llevar a cabo sus funcionalidades.

### 1. Proceso de Extracción

En la ilustración 20 se puede apreciar el proceso de extracción de la información. Este proceso se lleva a cabo cuando el usuario del sistema desea extraer información de Facebook.

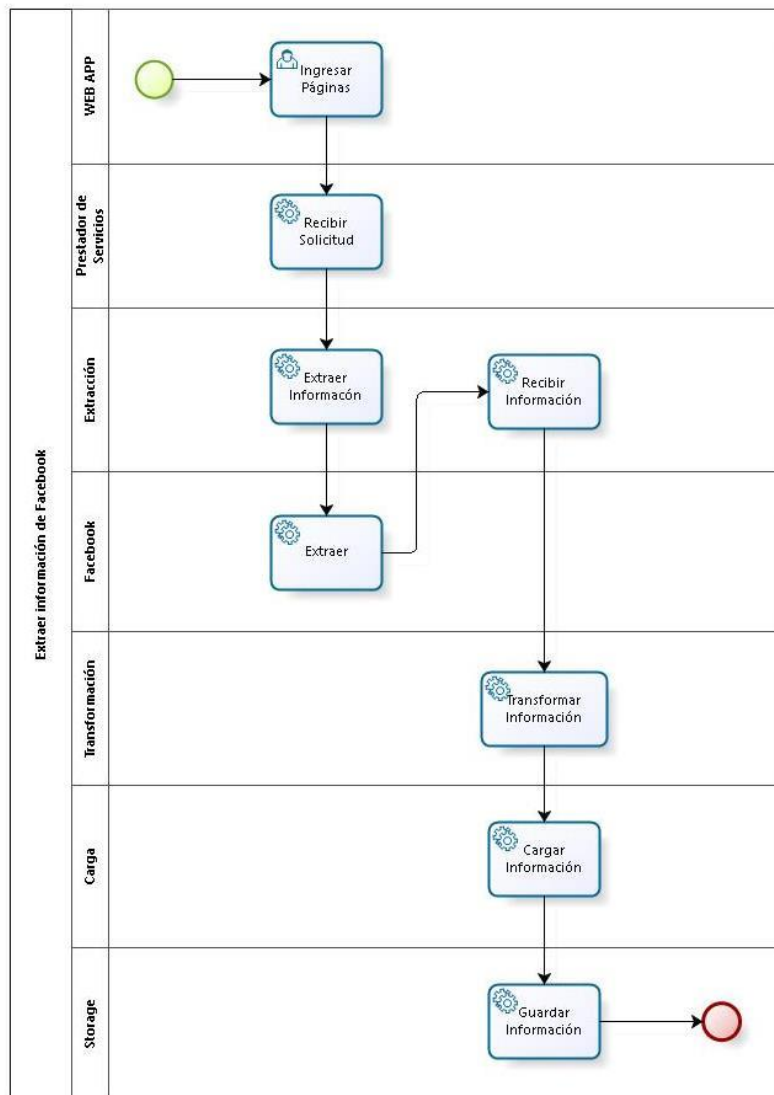


Ilustración 20 - Proceso de Extracción

## 2. Proceso de Consulta de Información

Cuando se ha realizado el proceso de extracción al menos una vez, el usuario podrá consultar la información existente dentro de la base de datos. Este proceso se puede apreciar en la ilustración 21.

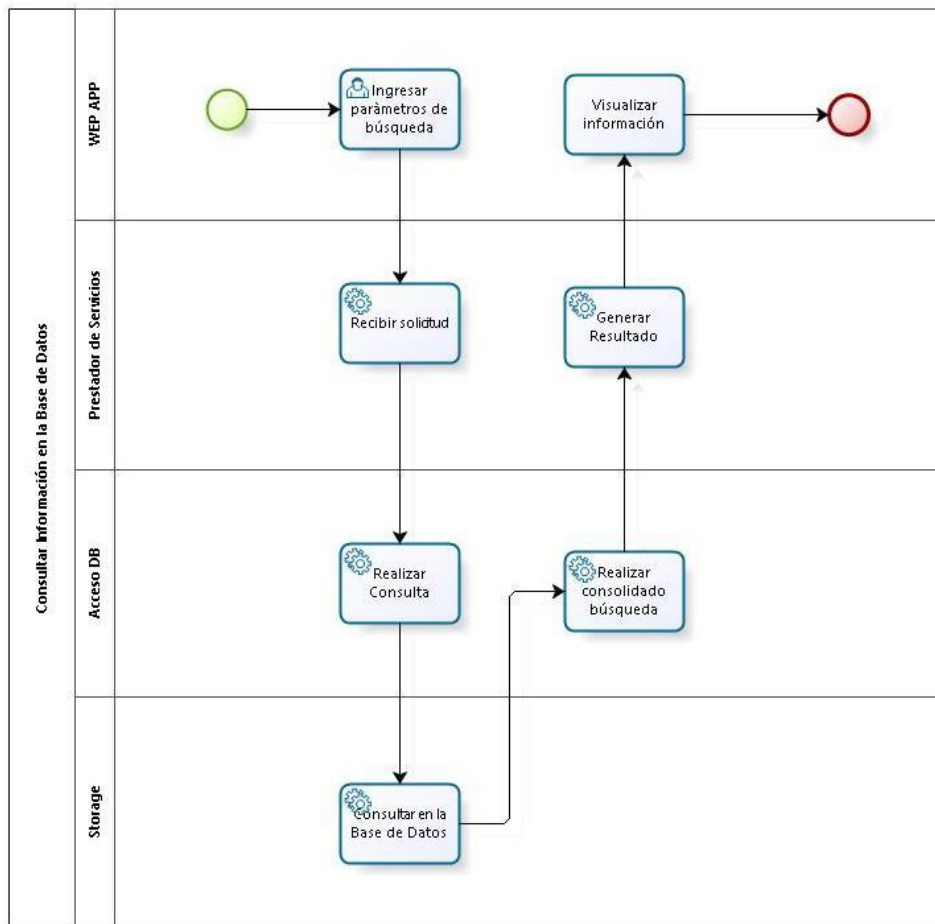


Ilustración 21 - Proceso de Consulta

### Diagrama de Clases Componente Extracción de Datos

El diagrama de la ilustración 22, muestra las clases que hacen parte del componente encargado de extraer información de la red social Facebook y de almacenar dicha información dentro de la base de datos del sistema. Este componente está conformado por ocho clases las cuales corresponden a los ocho componentes que hacen parte de la arquitectura de la capa de Extracción de Datos como se muestra en la sección de [Arquitectura del Sistema](#) donde se detallan las funciones de cada uno.

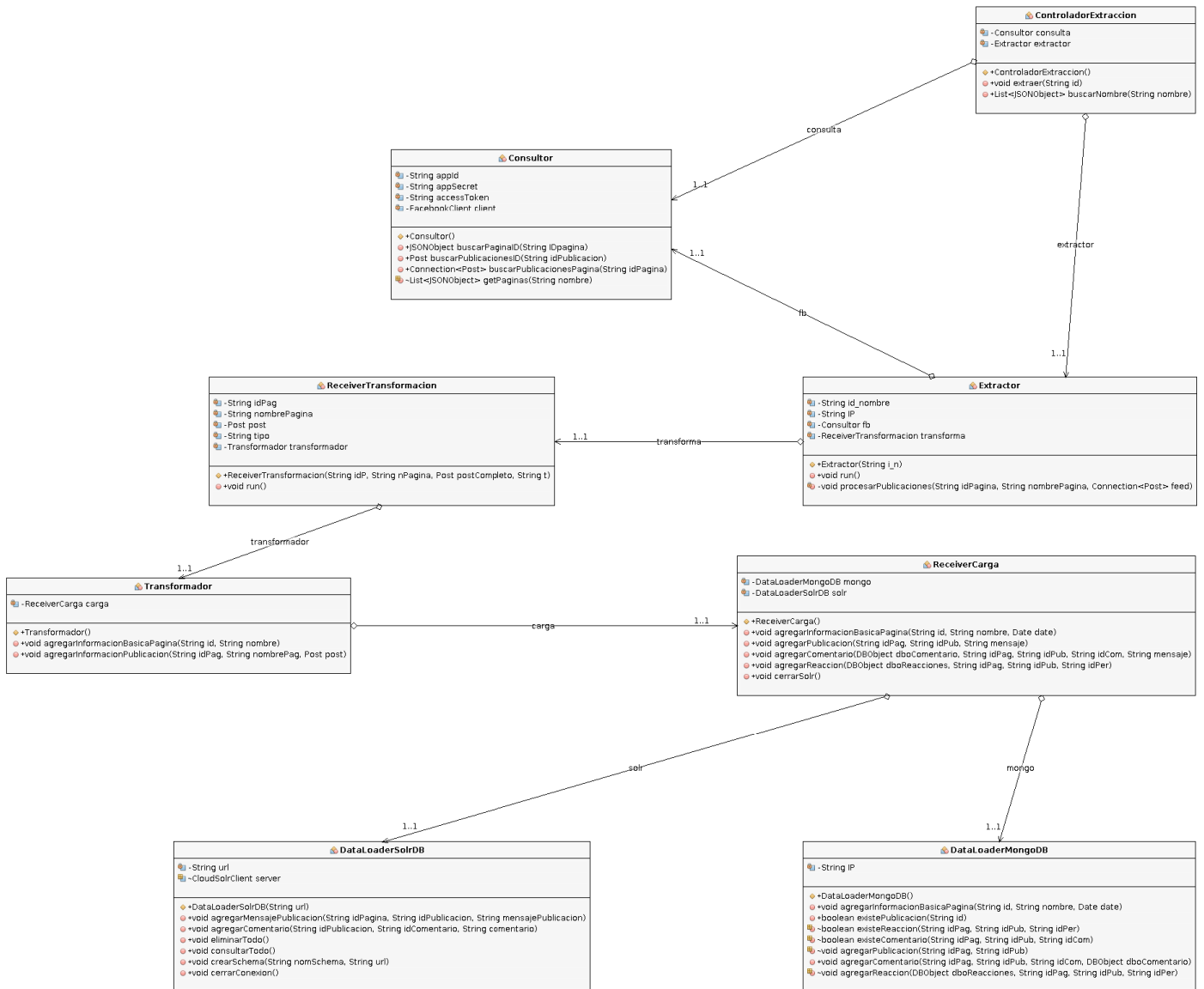


Ilustración 22 - Diagrama de Clases Componente Extracción de Datos

El patrón utilizado en el diseño de RART fue el patrón **singleton**. Según Larman, (Larman, 1999) el patrón **singleton** se utiliza para tener una instancia de una clase con el objetivo de tener solo un punto de acceso global. Este patrón fue utilizado en el siguiente contexto:

✓ **Carga información**

Para realizar el proceso de guardar la información dentro de la base de datos, se utilizó el patrón **singleton** con el fin de garantizar una instancia única de carga para cada base de datos de RART. El patrón se aplicó en las siguientes clases:

- ✓ Data Loader Mongo DB
- ✓ Data Loader Solr DB

**Diagrama de Clases Componente Carga y Transformación de Datos**

Este diagrama muestra las clases que hacen parte del componente encargado de extraer información de la base de datos del Sistema cuando el usuario desea consultarla y se puede apreciar en la ilustración 23. Las clases que se pueden apreciar en este diagrama corresponden a los 4 componentes que hacen parte de la capa Carga y Transformación de Datos como se describe en la sección [Arquitectura del Sistema](#) y cumplen con las funcionalidades allí mencionadas.

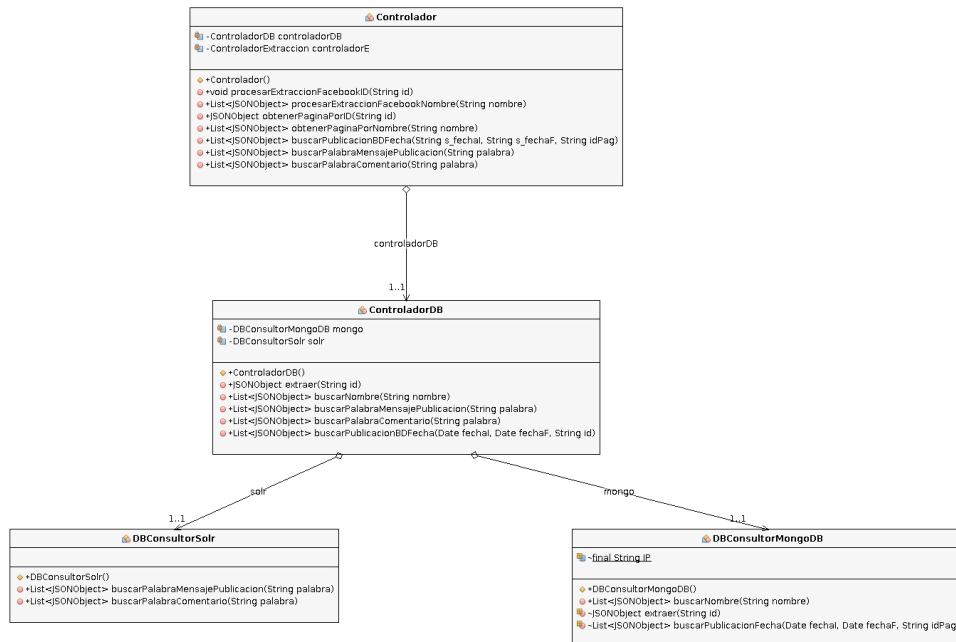


Ilustración 23 - Diagrama de Clases Componente Carga y Transformación de Datos

## V – RESULTADOS Y VALIDACIÓN

### Resultados

#### Cumplimiento de Objetivo General

El objetivo general definido para este proyecto fue el siguiente:

“Desarrollar un marco de trabajo (*Framework*) para recolectar información de publicaciones realizadas por una empresa de alimentos dentro de la red social Facebook, con el fin de generar un repositorio de datos apto para soportar procesos de *Data Analytics*.”

El objetivo principal se cumplió satisfactoriamente dado que el producto resultado de este proyecto cumple con la tarea de extraer la información desde la red social Facebook, la estructura y la guarda dentro de una base de datos la cual consta de dos partes, una en donde se almacenan los datos semiestructurados y la otra donde se almacenan los datos no estructurados.

#### Cumplimiento de Objetivos Específicos

Objetivo	Estado
1. Desarrollar un componente de software que extraiga información de usuarios y publicaciones de la red social Facebook.	<b>Cumplido.</b> Se desarrolló el componente de Software llamado “Extracción de Datos” el cual realiza el proceso de Extracción, Transformación y Carga de la información proveniente de la red social Facebook. Además, se desarrolló el componente de Software llamado “Carga y Transformación de Datos” el cual se encarga de extraer la información existente dentro de la base de datos de RART con el fin de mostrarla al usuario.
2. Desarrollar un componente de software que organice la información extraída dentro de un repositorio de datos.	
3. Validar el funcionamiento de los componentes desarrollados y la integración entre los mismos.	Cumplido. Por medio de los casos de prueba se validó el correcto funcionamiento de los

	componentes los cuales interactúan entre sí a la hora de generar los resultados esperados.
--	--

Ilustración 24 - Cumplimiento de Objetivos Específicos

## Interfaz Gráfica de RART

Las interfaces gráficas diseñadas, brindan a los usuarios de RART acceder a las funcionalidades que el *Framework* ofrece, las cuales se describen en la sección [Definición de Casos de Uso](#).

El funcionamiento detallado del sistema se encuentra en el anexo Manual de Usuario.

### 1. Página principal

La ilustración 25 muestra la página principal de RART. En ella se encuentra un menú en la parte izquierda el cual proporciona al usuario la facilidad de acceder a las principales funcionalidades del *Framework*: Extraer Información, Buscar Información, Actualizar Información y Configuraciones.

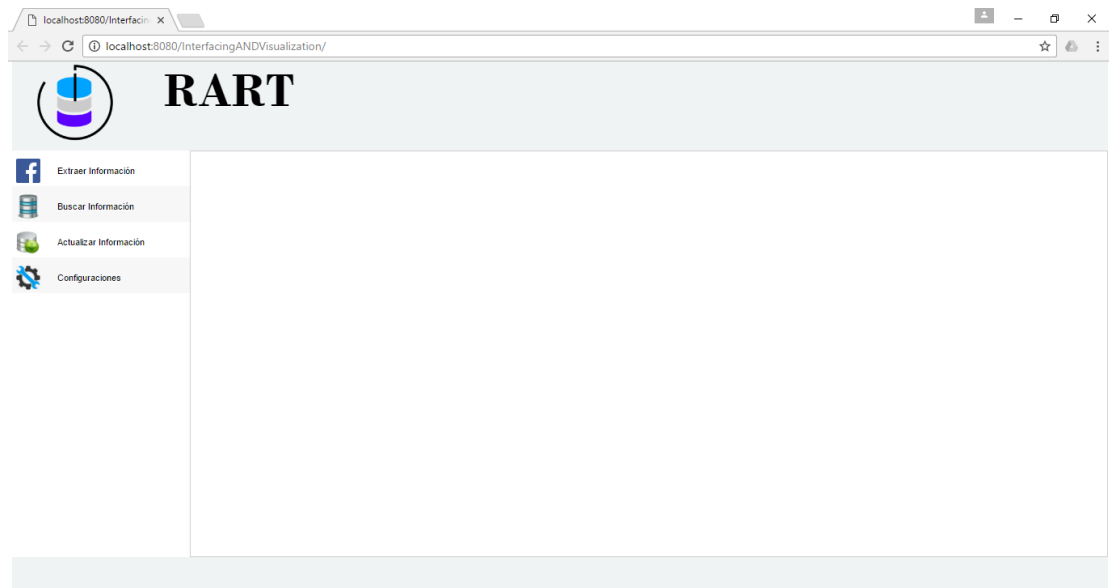


Ilustración 25 - Página Principal RART



## 2. Menú Extraer Información

Este menú, visualizado en la ilustración 26, permite al usuario realizar las tareas de extracción de información de la red social Facebook. Dentro de este menú se encuentran tres botones:

1. **Buscar página por ID:** este botón permite al usuario ingresar a la funcionalidad de ingresar un ID de una página existente dentro de la red social Facebook con el fin de extraer la información de sus publicaciones.
2. **Buscar Página por Nombre:** este botón permite al usuario ingresar a la funcionalidad de ingresar una palabra para que el sistema busque dentro de la red social Facebook las páginas existentes que contengan la palabra ingresada dentro de su nombre.
3. **Buscar Página en Archivo:** este botón permite acceder a la funcionalidad donde el usuario ingresa al sistema un archivo formato .TXT el cual contiene nombre y/o IDs de páginas de Facebook.



Ilustración 26 - Menú Extraer Información

### 3. Menú Buscar Información

En la ilustración 27 se presenta el menú correspondiente a la funcionalidad que permite al usuario buscar información existente dentro de la base de datos de RART. Dentro de este menú se encuentran tres botones:

1. **Buscar Página:** este botón permite al usuario buscar, por medio de una palabra, las páginas existentes dentro de la base de datos las cuales hayan sido extraídas previamente.
2. **Buscar Publicación:** este botón permite al usuario acceder a la funcionalidad de buscar una publicación dentro de la red social Facebook, por medio de una palabra y un periodo de tiempo.
3. **Listar Todas las Publicaciones:** este botón permite al usuario visualizar todas las páginas existentes dentro de la base de datos de RART las cuales han sido extraídas previamente.



Ilustración 27 - Menú Buscar Información

#### 4. Menú Actualizar Información

Esta funcionalidad permite al usuario actualizar toda la información existente dentro de la base de datos. Como se muestra en la ilustración 28, cuando el usuario oprime el botón correspondiente, el sistema empieza a actualizar toda la información de las páginas y sus respectivas publicaciones existentes dentro de la base de datos con la información de Facebook.



Ilustración 28 - Actualizar Información

## 5. Menú Configuraciones

Como se muestra en ilustración 29, esta funcionalidad permite al usuario configurar en el sistema el token de acceso a Facebook, y las IDs y puertos correspondientes a Mongo DB y Solr DB. Esta información se encuentra detallada en el anexo Manual de Instalación y Configuración.



Ilustración 29 - Menú Configuraciones

## Validación

### 1. Casos de Prueba

Este proceso de validación se realizó verificando el funcionamiento de cada caso de uso y validando que las salidas, en cada prueba, fueran consistentes. Los resultados obtenidos en este proceso fueron los siguientes:

Caso de Uso	Entrada	Resultado Esperado	Resultado	Conclusión
Ingresar al Sistema	Ingresar al sistema por medio de la URL proporcionada	Ingreso al sistema correctamente	Se ingresó al sistema correctamente	Correcto funcionamiento

Ingresar Página	Ingresar id de pagina	El sistema extrae la información correspondiente a la página del id ingresado	El sistema extrajo la información correctamente	Correcto funcionamiento
	Ingresar una palabra correspondiente al nombre de una página	El sistema busca y muestra las páginas correspondientes a ese nombre	El sistema mostró las páginas que existen con el nombre ingresado	Correcto funcionamiento
Cargar archivos página	Archivo de texto con ids y nombre de páginas	El sistema muestra las páginas existentes correspondientes a los ids y nombres ingresados en el archivo de texto	El sistema mostró las paginas encontradas con los ids y nombres ingresados en el archivo	Correcto funcionamiento
Buscar publicación	Buscar publicación que contenga una palabra dada por el usuario	El sistema muestra las publicaciones que contengan la palabra ingresada	El sistema mostró las publicaciones que contienen la palabra ingresada	Correcto funcionamiento
	Buscar una publicación que contenga una palabra dada por el usuario y además en un periodo de fecha	El sistema muestra las publicaciones que contengan la palabra ingresada	El sistema mostró las publicaciones que contenías la palabra ingresada en el periodo de tiempo dado.	Correcto funcionamiento
Ver detalle de publicación	El usuario selecciona ver el detalle de la publicación	El sistema muestra el detalle de la publicación seleccionada	El sistema mostró el detalle de la página seleccionada	Correcto funcionamiento
Buscar página en la base de datos	El usuario ingresa el id de una página	El sistema muestra la información correspondiente a la página buscada	El sistema mostró la página buscada	Correcto funcionamiento
	El usuario ingresa una palabra contenida en el nombre de una página	El sistema muestra la información correspondiente a la página buscada	El sistema mostró las páginas encontradas que contienen la palabra ingresada en su nombre	Correcto funcionamiento

Visualizar publicaciones página	El usuario selecciona ver las publicaciones de la página seleccionada	El sistema muestra las publicaciones de la página seleccionada	El sistema mostró las publicaciones correspondientes a la página seleccionada	Correcto funcionamiento
Filtrar publicaciones	El usuario ingresa una palabra para buscar dentro de las publicaciones de la página seleccionada	El sistema muestra las publicaciones que contienen la palabra ingresada	El sistema mostró las publicaciones que contienen la palabra buscada dentro de su nombre	Correcto funcionamiento
	El usuario ingresa una fecha inicial y una fecha final	El sistema muestra las publicaciones existentes que se encuentran dentro de ese rango de fechas	El sistema mostró las publicaciones que se encuentran dentro del rango de tiempo ingresado	Correcto funcionamiento
Graficar Reacciones	El usuario selecciona graficar las reacciones de una publicación seleccionada	El sistema muestra la gráfica correspondiente	El sistema mostró la gráfica de reacciones correspondiente.	Correcto funcionamiento

Tabla 5 - Casos de Prueba

## 2. Pruebas de Desempeño

Con el fin de medir el desempeño del RART se realizaron 4 pruebas de extracción de información de publicaciones, de 3 páginas de sectores económicos diferentes.

1. Textiles: Arturo Calle.
2. Tecnología: Xbox Colombia.
3. Alimentos: Hamburguesas El Corral.

Por cada una de estas páginas se realizó la extracción de publicaciones durante una ventana de tiempo de 30 minutos. Adicionalmente, cada minuto se consultó dentro de la base de datos de RART la cantidad de publicaciones, reacciones y comentarios.

Las ilustraciones 30, 31 y 32 muestran la cantidad de publicaciones, reacciones y comentarios, extraídas en cada una de las páginas anteriormente mencionadas respectivamente.

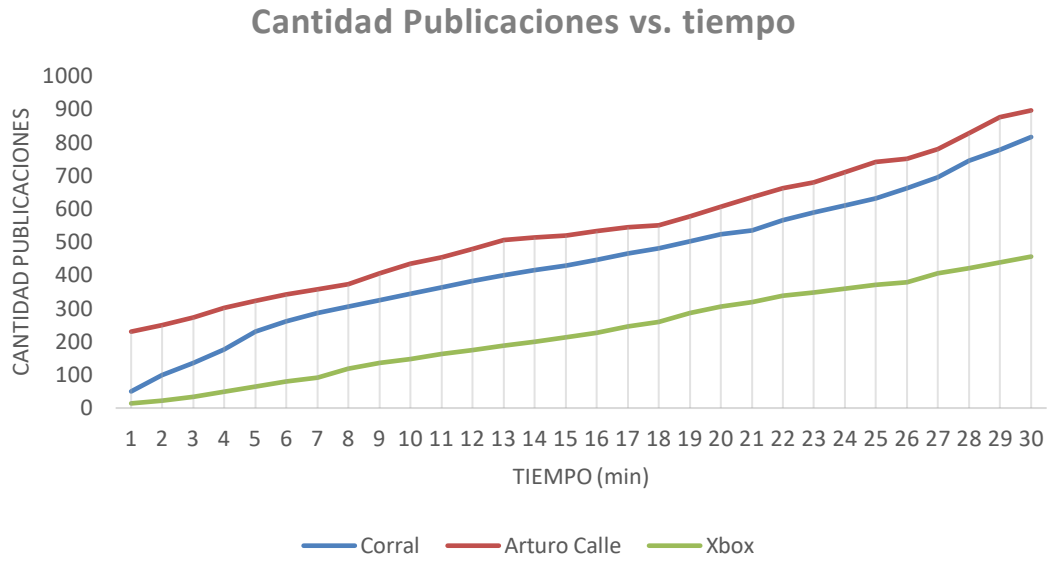


Ilustración 30 - Cantidad publicaciones vs. tiempo

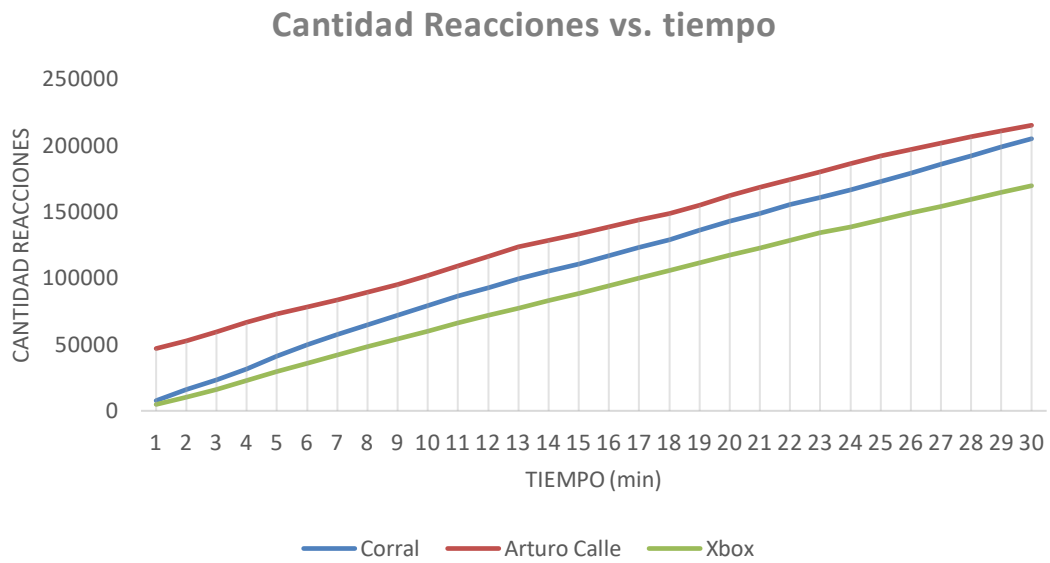


Ilustración 31 - Cantidad reacciones vs. tiempo

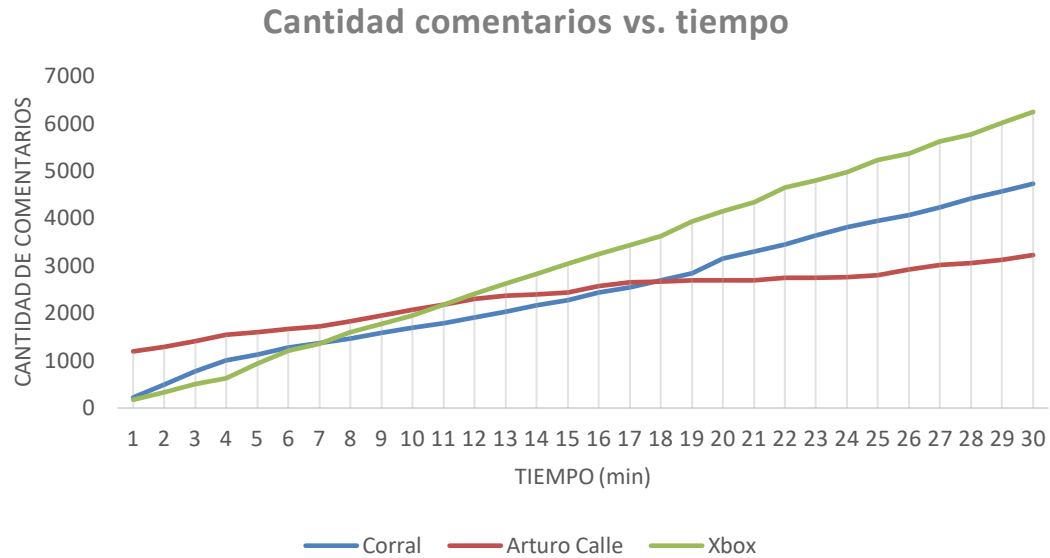


Ilustración 32 - Cantidad cometarios vs. Tiempo

Una vez realizadas las pruebas de extracción de publicaciones, reacciones y comentarios de las páginas mencionadas de forma individual, se procedió a realizar el proceso de extracción de dicha información en forma paralela, es decir de las tres páginas al mismo tiempo.

Las ilustraciones 33, 34 y 35 muestran los resultados de las pruebas de extracción de las tres páginas en simultaneo y se comparan con el resultado esperado el cuales determinó por medio de la sumatoria de las pruebas realizadas de forma individual.

En un principio se puede apreciar que la cantidad de publicaciones no se acerca a los valores esperados, pero al pasar el tiempo se va acercando, hasta que llega un momento en el que los valores obtenido al paralelizar el proceso obtienen mejores resultados que los que se esperaban en la sumatoria.



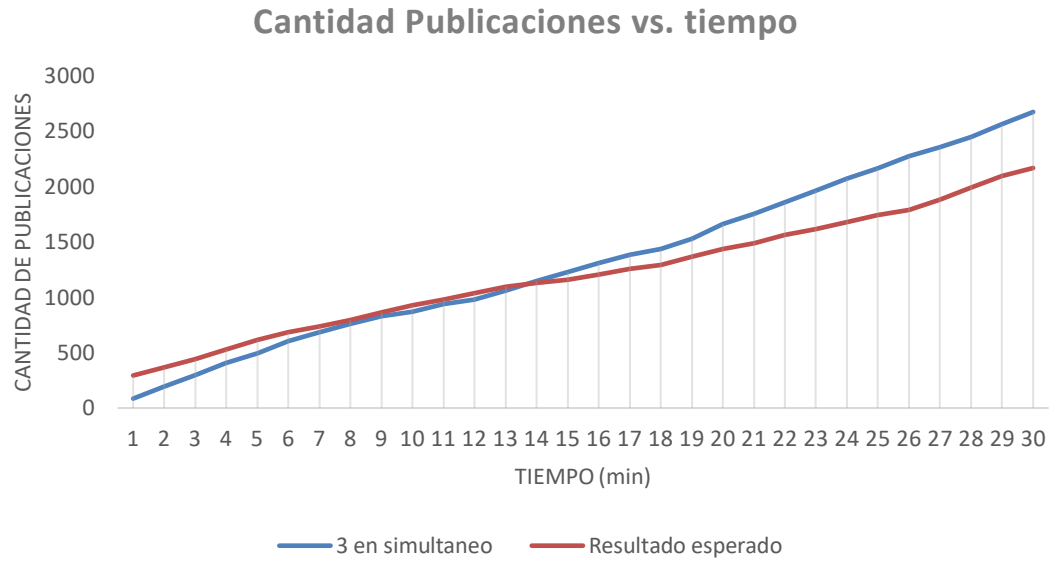


Ilustración 33 - Cantidad publicaciones vs. tiempo (3 en simultaneo)

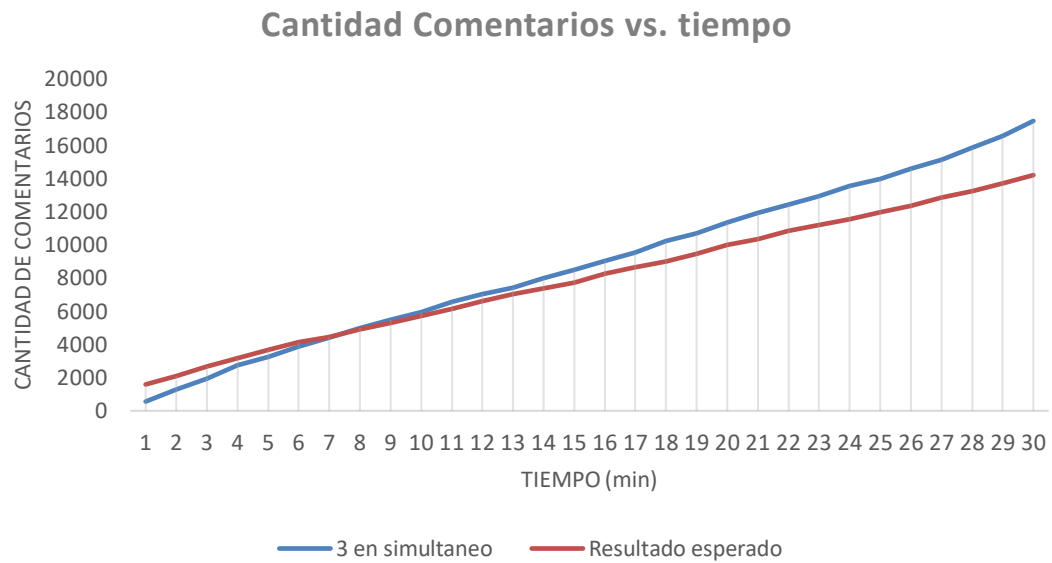


Ilustración 34 - Cantidad de comentarios vs. tiempo (3 en simultaneo)

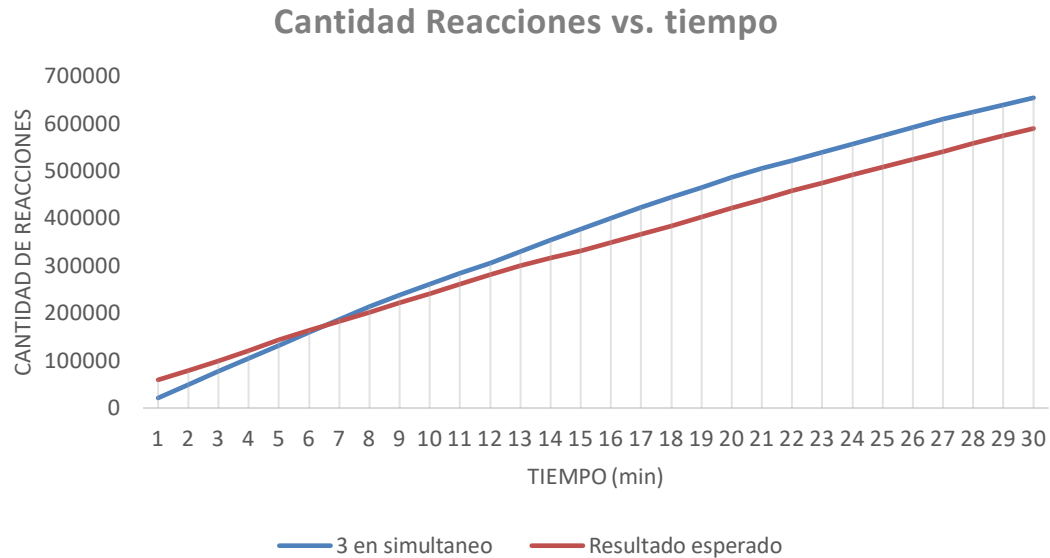


Ilustración 35 - Cantidad de reacciones vs. tiempo (3 en simultaneo)

Finalmente, en las pruebas de desempeño evaluamos la utilidad de RART, tomando los tiempos de consulta de toda la información existente en la base de datos. Este proceso se realizó mientras se almacenaba información en las bases de datos consultando cada minuto, durante 30 minutos, toda la información existente en la base de datos y registrando el tiempo que tardaba RART en arrojar el resultado esperado. En la ilustración 36 se aprecia el resultado que demuestra la utilidad de RART, pues mientras que 30 minutos después se siguen extrayendo datos de Facebook, la consulta de todos esos datos desde las bases de datos, toma menos de 1 segundo.

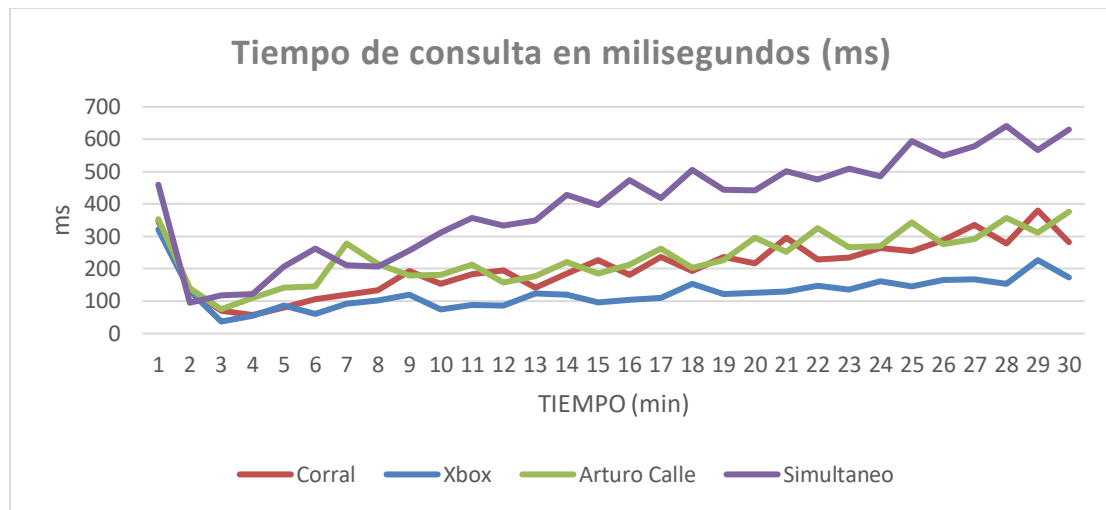


Ilustración 36 - Tiempo de consulta

### Conclusiones de las pruebas de desempeño

1. El desempeño demostrado durante la extracción de datos, es estable y muestra un ritmo constante de extracción de datos de la red social Facebook, que se ve directamente influido por la cantidad de comentarios y reacciones con los que cuenta cada publicación.
2. Los tiempos de consulta en los repositorios son mínimos en comparación a los tiempos de extracción directa desde la red social Facebook, justificando el trabajo desarrollado a lo largo del semestre.
3. El proceso de extracción de cada página mediante un proceso independiente, paralelizando la extracción de todas las paginas, mejora los tiempos de extracción de datos de las redes sociales.
4. Dado que es un problema con un gran volumen de datos, que se generan a gran velocidad y que cuentan con gran variedad, hacen de este un problema de Big Data.

### 3. Pruebas de Usuario

Las pruebas de usuario se realizaron por medio de una validación de pruebas TAM (Venkatesh & Hillol, 2008) las cuales consisten en realizar una serie de preguntas a usuarios de prueba

luego de conocer e interactuar con el Software. En este proceso participaron diez miembros de CAOBA los cuales demostraron una excelente aceptación del usuario. A los participantes se les expuso una presentación (ver [Anexo Presentación Pruebas TAM](#)) en donde se les contextualizó el funcionamiento y la arquitectura de RART. Además, se les realizó una demostración de RART en vivo para que conocieran las funcionalidades que este ofrece.

### Escala de calificación de pruebas

Las preguntas realizadas a los usuarios fueron calificadas de acuerdo con la escala de la tabla 6.

<b>1</b>	Muy en desacuerdo
<b>2</b>	Moderadamente en desacuerdo
<b>3</b>	Algo en desacuerdo
<b>4</b>	Neutral
<b>5</b>	Algo de acuerdo
<b>6</b>	Moderadamente de acuerdo
<b>7</b>	Totalmente de acuerdo

Tabla 6 - Escala de Calificación TAM

Las preguntas que se encuentran en la tabla 7 son las que se aplicaron a los usuarios con el fin de determinar el nivel de aceptación del mismo

<b>Categoría</b>	<b>Pregunta</b>
<b>Utilidad Percibida</b>	El uso del sistema mejora el rendimiento en mi trabajo.
	Usar el sistema en mi trabajo aumenta mi productividad.
	El uso del sistema mejora mi efectividad en mi trabajo.

	Encuentro que el sistema es útil en mi trabajo.
<b>Facilidad de Uso Percibida</b>	Mi interacción con el sistema es clara y comprensible.
	Interactuar con el sistema no requiere mucho de mi esfuerzo mental.
	Creo que el sistema es fácil de usar.
	Me resulta fácil hacer que el sistema haga lo que yo quiero que haga.
<b>Percepciones de control externo</b>	Tengo control sobre el uso del sistema.
	Tengo los recursos necesarios para usar el sistema.
	Dado los recursos, las oportunidades y los conocimientos necesarios para utilizar el sistema, sería fácil para mí utilizar el sistema.
	El sistema no es compatible con otros sistemas que uso.
<b>Placer percibido</b>	Me parece que usar el sistema puede ser agradable.
	El proceso real de usar el sistema es agradable.
	Me divierto usando el sistema.
<b>Usabilidad objetiva</b>	No se utilizaron elementos específicos. Se midió como una proporción de tiempo dedicado por el sujeto al tiempo empleado por un experto en el mismo conjunto de tareas.
<b>voluntariedad</b>	Mi uso del sistema es voluntario.
	Mi supervisor no requiere que yo use el sistema.
	Aunque podría ser útil, el uso del sistema no es ciertamente obligatorio en mi trabajo.
<b>Relevancia del trabajo</b>	En mi trabajo, el uso del sistema es importante.
	En mi trabajo, el uso del sistema es relevante.
	El uso del sistema es pertinente a mis diversas tareas relacionadas con el trabajo.
<b>Calidad de salida</b>	La calidad de la salida que recibo del sistema es alta.
	No tengo ningún problema con la calidad de la salida del sistema.
	Calculo los resultados del sistema como excelentes.
<b>Demostrabilidad del resultado</b>	No tengo dificultad para decirles a otros acerca de los resultados del uso del sistema.
	Creo que podría comunicar a otros las consecuencias de usar el sistema.
	Los resultados del uso del sistema son evidentes para mí.
	Tendría dificultades para explicar por qué el uso del sistema puede o no ser beneficioso.
<b>Intención Conductual</b>	Suponiendo que tuviera acceso al sistema, pienso usarlo.
	Dado que tenía acceso al sistema, predigo que lo usaría.
	Planeo usar el sistema en los próximos ____ meses.
<b>Utilizar</b>	¿En promedio, cuánto tiempo usted pasa en el sistema cada día?

Tabla 7 - Preguntas Pruebas de Usuario

### Resultados de pruebas de usuario

Luego de aplicar las preguntas mencionadas a los usuarios participantes, se obtuvieron los resultados que se muestran en la ilustración 37.

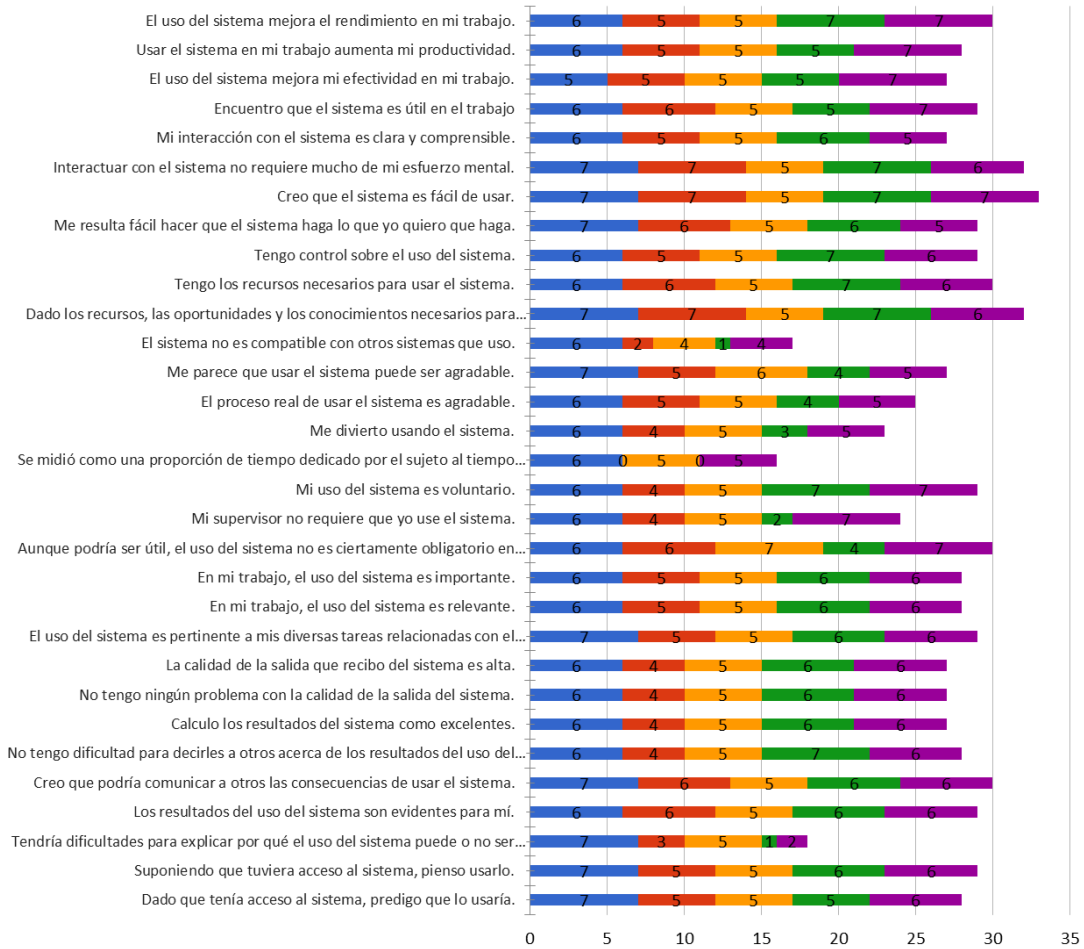


Ilustración 37 - Resultados Pruebas de Usuario

### Conclusiones pruebas es usuario

Luego de analizar los resultados se realizaron las siguientes conclusiones:

1. El *Framework* cumple con lo esperado por los usuarios del centro CAOBA.
2. El *Framework* desarrollado será muy útil para apoyar los proyectos del centro CAOBA.

3. Al usuario se le facilita desplazarse por los diferentes menús, entendiendo la razón de ser de cada uno y sacando el máximo provecho de cada uno de estos.
4. Los usuarios perciben el beneficio de usar el *Framework* y encuentran en este una buena herramienta trabajo.
5. Los usuarios definitivamente usarán el *Framework* desarrollado.

## VI – CONCLUSIONES, IMPACTOS Y TRABAJO FUTURO

### 1. Conclusiones

#### Pregunta de Investigación Respondida

Por medio del *Framework* desarrollado, es posible extraer información de publicaciones realizadas dentro de la red social Facebook. Además, RART almacena la información extraída dentro de un repositorio de datos apto para aplicar procesos de *Data Analytics* con el fin de encontrar nuevo conocimiento como tendencias y patrones.

#### Objetivos Cumplidos

Al finalizar este trabajo de grado se cumplieron los objetivos planteados como propuesta, tanto el general como los específicos. Estos resultados se encuentran descritos en las secciones [Cumplimiento de Objetivo General](#) y [Cumplimiento de Objetivos Específicos](#).

### 2. Análisis de Impacto del Desarrollo

Este proyecto fue desarrollado según las necesidades presentadas por una empresa del sector de los alimentos la cual hace parte de la Alianza CAOBA. Los impactos obtenidos durante el desarrollo de este proyecto y los cuales se esperan obtener son los siguientes:

#### Impacto Disciplinar

Como se mencionó al principio de este documento, este proyecto se abordó desde la perspectiva de Big Data. Dado que este tema está en auge, fue necesario que el equipo de trabajo se sentara a investigar la mejor forma en la que se debía abordar el problema para desarrollar la mejor solución a la problemática. Además, este proyecto es de gran valor para CAOBA ya que se

presenta como una base de investigación para continuar con el desarrollo de problemas que aborden información de la red social Facebook.

### **Impacto Social**

Este proyecto es parte de la primera etapa de un proyecto de mayor alcance y el impacto social que se toma en consideración es lograr conocer con mayor precisión las necesidades de los consumidores con el fin de poder satisfacer sus necesidades. Además, se prevé usar RART para analizar no solo consumidores, sino las necesidades o reacciones de ciudadanos ante eventos dados a conocer por medio de la red social Facebook y de esta forma obtener conocimiento que ayude a generar mejores políticas públicas.

### **Impacto Económico**

Lograr conocer con mayor precisión las necesidades de los consumidores hace que las empresas puedan ofrecer en el mercado productos más acordes a dichas necesidades lo cual se traduce en mayor ventas e incremento de rentabilidad. Se espera que con la continuidad de este proyecto por parte de CAOBA se logre este impacto planteado.

## **3. Trabajo Futuro**

De acuerdo con los resultados obtenidos y al alcance definido para éste, se espera que como trabajo futuro se realicen las siguientes actividades:

- ✓ Según las pruebas de rendimiento se espera optimizar el proceso de extracción de información con el fin de obtener un mejor rendimiento a la hora de poblar la base de datos.
- ✓ Aplicar procesos de *Data Analytics* en el repositorio de datos poblado por RART con el fin de generar conocimiento acerca del comportamiento de consumo de los usuarios de Facebook, como lo son patrones de consumo y segmentación de mercados.



## VII- REFERENCIAS Y BIBLIOGRAFÍA

- Abbasimehr, H. S. (2013). A framework for identification of high-value customers by including social network based variables for churn prediction using neuro-fuzzy techniques. *International Journal of Production Research*, 51(4), (1279-1294).
- Alsubagh, H. (2015). The Impact of Social Networks on Consumers' Behaviors. *International Journal of Business and Social Science*.
- Apache. (2016). *Apache Solr*. Obtenido de <http://lucene.apache.org/solr/>
- Arcitura., E. I. (2016). *Fundamental Big Data*.
- Azevedo, A. I. (2008). Kdd, semma crisp-dm: a parallel overviee. IADS-DM.
- Calambás, D., & Mendoza, J. (2016). RART Software Design Document.
- Calambás, D., & Mendoza, J. (2016). RART Software Project management Plan.
- Calambás, D., & Mendoza, J. (2016). RART Software Requirements Specification.
- Comunitarios, A. F. (2014). Glosario de términos utilizados en redes sociales.
- Dataprix. (11 de Abril de 2016). *Dataprix*. Obtenido de Dataprix: <http://www.dataprix.com/es/el-modelo-referencia-crisp-dm>
- Dzhurov, Y. K. (2009). Personal Extreme Programming–An Agile Process for Autonomous Developers.
- Extreme Programming: A gentle introduction*. (2016). Obtenido de <http://www.extremeprogramming.org/>
- Facebook. (2016). *DEvelopers., Facebook*. Obtenido de Facebook for Developers: <https://developers.facebook.com/docs/apis-and-sdks#third-party-sdks>
- Forouzandeh, S. S. (2014). Content marketing through data mining on Facebook social network. *Webology*, 11(1), (1).
- Franco, A. M., & Libreros, V. (Mayo de 2013). *Biblioteca Digital* . Obtenido de ICESI: [https://bibliotecadigital.icesi.edu.co/biblioteca\\_digital/bitstream/10906/76578/1/redes\\_sociales\\_marcas.pdf](https://bibliotecadigital.icesi.edu.co/biblioteca_digital/bitstream/10906/76578/1/redes_sociales_marcas.pdf)
- García, M. (2010). El impacto de las redes sociales. *Revista Dintel*.
- Gutierrez, J. (8 de Diciembre de 2016). *¿Qué es un framework?* Obtenido de [http://www.lsi.us.es/~javierj/investigacion\\_ficheros/Framework.pdf](http://www.lsi.us.es/~javierj/investigacion_ficheros/Framework.pdf)

- IEEE. (2001). Standard for Software Life Cycle Processes—Risk Management (IEEE STD 1540-2001).
- Kotonya, G. &. (1998). *Requirements engineering: processes and techniques*. Wiley Publishing.
- Larman, C. (1999). *UML y Patrones*. Pearson.
- Mongo DB. (2016). Obtenido de <https://www.mongodb.com/es>
- Nohuddin, P. N. (2014). Trend mining in social networks: from trend identification to visualization. *Expert Systems*, 31(5), (457-468).
- Perez, D., & Martinez, I. (2006). *EOI*. Obtenido de Escuela de Negocios: [http://api.eoi.es/api\\_v1\\_dev.php/fedora/asset/eoi:45089/componente45087.pdf](http://api.eoi.es/api_v1_dev.php/fedora/asset/eoi:45089/componente45087.pdf)
- Pomares, A. (28 de 04 de 2016). Entrevista trabajo de grado. (M. J. Calambás Daniel, Entrevistador)
- Sanz, L. (2003). Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes.
- SAS Institute, i. (2003). *Basic Concepts in Research and Data Analytics*.
- Schroeck, M., Shockley, R., Dolores, J., & Tufano, P. (2012). *Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de datos inciertos*.
- Spark, A. (9 de marzo de 2016). *Spark*. Obtenido de <http://spark.apache.org/>
- Stone, B. (10 de 11 de 2016). *SocialTimes*. Obtenido de Here's How Many People Are on Facebook, Instagram, Twitter and Other Big Social Networks: <http://www.adweek.com/socialtimes/heres-how-many-people-are-on-facebook-instagram-twitter-other-big-social-networks/637205>
- TIC, M. d. (1 de marzo de 2016). *Ministerio de TIC*. Obtenido de Nace CAOBA, primera alianza público privada en Big Data: <http://www.mintic.gov.co/portal/604/w3-article-14678.html>
- Venkatesh, V., & Hillol, B. (2008). Technology Acceptance Model 3 and Research Agenda on Interventions. *Decision sciences*, 39(2), 273-315.
- Villota, S. (16 de Marzo de 2016). *Universidad Autonoma de Occidente*. Obtenido de <http://red.uao.edu.co/bitstream/10614/8450/1/C0163.pdf>
- Yang, W. S. (2006). Mining social networks for targeted advertising. *In System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, (Vol. 6, pp. 137a-137a).

Zhang, J. T. (2007). Expert finding in a social network. *In Advances in Databases: Concepts, Systems and Applications*, (pp. 1066-1069).

## VIII – ANEXOS

1. [RART Software Project Management Plan](#)
2. [RART Software Requirements Specification](#)
3. [RART Software Design Document](#)
4. [RART Manual de Instalación y Configuración](#)
5. [RART Manual de Usuario](#)
6. [Componente Extracción de Datos .jar](#)
7. [Componente Carga y Transformación de Datos .jar](#)
8. [Componente Visualización e Interfaces .war](#)