

**DESARROLLO DE UNA HERRAMIENTA PARA VALORACIÓN DE
INFORMACIÓN POR PARTE DE EXPERTOS Y PUBLICO GENERAL**

JEHUDI CASTRO SIERRA
JUAN DAVID ZAPATA

PROYECTO DE GRADO INGENIERIA DE SISTEMAS

Asesor:
JORGE HERNAN ABAD LONDOÑO

**UNIVERSIDAD EAFIT
MEDELLIN
2008**

TABLA DE CONTENIDOS

1. Definición del problema	5
2. Justificación.....	7
3. Objetivo general	8
3.1. Objetivos específicos	8
4. Alcance.....	9
5. Beneficios.....	10
6. Metodología sugerida.....	11
7. Arquitectura de la Aplicación	12
7.2. Propósito.....	12
7.3. Alcance	12
7.4. Audiencia	13
7.5. Lineamientos de diseño	13
7.6. Premisas de la arquitectura	14
7.7. Vista conceptual de la arquitectura	15
7.7.1. Document Search.....	16
7.7.2 GUI (User Interface) Layer.....	17
7.7.2.1 Usuarios	17
Usuarios Registrados:.....	17
Usuarios Publicos	17
7.7.3 FrontEnd Document Search.....	18
7.7.4 Business Layer	18
7.7.4.1 Objetos y entidades de negocio	18
7.7.4.2 Flujos del proceso de negocio.....	18
7.7.5 Data Layer	19
7.7.5.1 Data Access Layer:	19
Google search engine service.....	20
8. Cronograma	21

9. INTEGRANTES.....	22
9.1. Ejecutores del proyecto	22
9.2. Asesor del proyecto	22
10. BIBLIOGRAFIA	23

INDICE DE ILUSTRACIONES

Ilustración 1: Arquitectura de la app	16
Ilustración 2: Cronograma - Actividades.....	21
Ilustración 3: Cronograma - Tiempos	21

1. Definición del problema

Más que una simple innovación tecnológica, *Internet* se ha convertido en un cambio cultural que sigue transformando la forma de comunicarnos y de acceder a la información, brindándonos una cantidad inimaginable de esta a tan solo un clic de distancia, ya que su finalidad (objetivo para el cual fue diseñado Internet) fue permitir el intercambio ilimitado de información, por lo cual las páginas WEB han registrado un crecimiento acelerado en un tiempo de vida relativamente corto. Por consecuencia cada vez son más las personas que publican páginas electrónicas de diversos tópicos sin ser sometidas a una evaluación previa por parte de editores o profesionales en la materia.

Lo que se ve reflejado en un crecimiento de la información digital de manera desordenada, carente de regulación y estándares de control de calidad. Problemática que se ve reflejada directamente en la comunidad académica, pues complica el determinar si las fuentes de datos son confiables y de calidad, sumado a esto, estudiantes e investigadores pasan horas incontables buscando información de calidad en la WEB.

La búsqueda de calidad ha hecho necesario formular propuestas para evaluar la información que se distribuyen vía Internet, que incluyen, además de los criterios existentes utilizados en medios impresos: contenido, antigüedad, periodicidad, profundidad del tema, etcétera, elementos para evaluar algunas particularidades de la información en línea, tales como: acceso, contenido, navegación y organización del sitio.

Bajo esta perspectiva, pretendemos elaborar una herramienta que permita a la comunidad académica colombiana tener las bases para enfrentar la problemática de ACCEDER A INFORMACION CONFIABLE Y DE CALIDAD A TRAVES DE INTERNET, mediante una metodología de valoración de la información en línea,

explorando una perspectiva poco desarrollada como son los antecedentes de la persona que provee la información, entre los que se incluyen la hoja de vida, y la valoración histórica de sus aportes al sistema.

2. Justificación.

El volumen de documentos al que se avoca una persona cuando esta en busca de información es desconcertante (Google indexa más de 8 mil millones de documentos), esta situación muestra un potencial problema que se constata cuando desde la academia y el ámbito investigativo se manifiesta la necesidad de encontrar material de calidad en tiempos relativamente cortos, lo que se traduce en la necesidad de la existencia de una metodología que permita la implementación de una herramienta de las características que se proponen.

Al recoger diversas opiniones en el medio y tras una amplia investigación se concluye que en este momento no existe una herramienta (ni una metodología que permita su implementación) que permita destacar la información de calidad sobre los recursos inservibles de una manera efectiva para el sector académico colombiano.

Una de las áreas importantes de la ingeniería de sistemas es el manejo de la información de manera óptima y ágil, por lo que el problema se ajusta a nuestra área de conocimiento como ingenieros de sistemas de la Universidad EAFIT.

3. Objetivo general

Desarrollar una herramienta que permita el valorar información en línea por parte de expertos y público general, que sea aplicable al sector académico colombiano.

3.1. *Objetivos específicos*

- Implementar una herramienta que permita valorar información por expertos y público general.
- Permitir la calificación de artículos a través de la herramienta que permita clasificar dicha información bajo los conceptos de la metodología de valoración que se empleara.
- Permitir la calificación de autores como proceso de apoyo en la valoración y clasificación de la información contenida en la herramienta.
- Permitir la valoración de la experiencia de los expertos (o autores de los artículos) como proceso de apoyo en la valoración y clasificación de la información contenida en la herramienta.
- Valorar la información almacenada en la herramienta basado en los conceptos de calificación de artículos, autores y valoración de experiencia de los autores.

4. Alcance

Con la realización de este proyecto se pretende realizar la implementación de un sistema web basado en una metodología específica que permita valorar documentos digitales.

5. Beneficios

Gracias a la realización de este proyecto, estudiantes, docentes e investigadores y público general tendrán acceso a una herramienta que le permitirá acceder a información especializada confiable de una manera ágil y efectiva, lo cual permitirá disminuir el tiempo que se pasa buscando dicha información en la red y emplear dicho ahorro en labores más relevantes como construcción de nuevo conocimiento a partir de lo encontrado y proyectos de investigación más profundos y completos.

Además, la Universidad EAFIT contará con una herramienta de este tipo lo que le permitirá brindar a los estudiantes una alternativa más para acrecentar sus conocimientos en áreas de estudio específicas.

6. Metodología sugerida.

Se realizarán las actividades establecidas como metodología de investigación para un proyecto investigativo.

- Realizar una investigación tanto en Internet, libros y demás herramientas que nos permitan tener un conocimiento global del tema
- Investigar casos de éxito de las empresas que implementen la gestión cuantitativa del proceso de software, ya sea en el medio colombiano ó extranjero.
- Construir una metodología que permita solucionar el problema de la valoración de la información digital para el ámbito académico colombiano.
- Construir una herramienta que aplique y plasme la metodología desarrollada basada en las etapas de construcción de un sistema de la ingeniería del software.
- Analizar las oportunidades y estrategias a seguir para la aplicación de esta metodología en Colombia en otros sectores diferentes al académico.

7. Arquitectura de la Aplicación

7.2. Propósito

La arquitectura base para la implementación de DOCUMENT SEARCH permite conocer de forma global y estandarizada los diferentes componentes de la solución. Esta arquitectura define las características más generales de la implementación del sistema y hace las veces de punto de referencia para futuros desarrollos.

Esta arquitectura está definida sobre los lineamientos arquitectónicos actuales del Modelo Vista Controlador (MVC) que es un patrón de arquitectura de software que separa los datos de una aplicación, la interfaz de usuario, y la lógica de control en tres componentes distintos.

Los lineamientos arquitectónicos aquí definidos corresponden a las necesidades específicas identificadas en el análisis del proceso y que deben ser implementadas para facilitar el manejo de la herramienta o ampliar la definición de la misma.

7.3. Alcance

La arquitectura está planteada para que sea acorde a las definiciones arquitectónicas definidas para aplicaciones J2EE.

El alcance específico de las definiciones arquitectónicas a las cuales se hace referencia en este documento pertenecen específicamente a los procesos soportados por los siguientes módulos de DOCUMENT SEARCH:

- Administración.

- Búsquedas.
- Gestión documental.

7.4. Audiencia

Este documento se plantea en una terminología técnica y va dirigido principalmente al grupo personas involucradas en las fases de construcción, pruebas, implantación y soporte de la solución.

7.5. Lineamientos de diseño

Los siguientes lineamientos enmarcan de manera global las diferentes decisiones y definiciones de la arquitectura del sistema.

- La arquitectura del sistema está basada en los componentes y servicios de la plataforma J2EE.
- La capa de presentación debe ser segura para cross site scripting y técnicas de injection, por tal razón la capa de presentación no deberá asumir ninguna lógica de acceso a las bases de datos sin pasar por un conjunto de capas intermedias (JSTL).
- Todas las validaciones de ingreso de información se realizarán directamente sobre la capa de presentación sin necesidad de entregarlas a una capa posterior.
- Todas las entidades deberán ser mapeadas y persistidas dentro de Entities Bean para representar su estado. Este estado deberá permanecer en memoria para efectos netamente de representación del objeto mientras se ejecuta la solicitud del cliente y su posterior representación dentro del modelo de datos.

- Todos los componentes de reglas de negocio deberán abrir transacciones desde esta capa y no directamente en la capa de acceso a datos con el fin de poder involucrar diferentes entidades dentro de una misma transacción si esta es la finalidad del evento o proceso disparado.
- Con el fin de utilizar el modelo de sesiones de Hibernate, desde los componentes de la capa de negocio se deberá realizar la apertura de la sesión y la transacción que involucra todas las operaciones como controlador del flujo, dicha sesión deberá pasarla como referencia a cada una de las diferentes componentes involucradas en el proceso para que sea compartida la sesión y pueda manejarse la transacción de manera atómica.
- Se realizará la implementación de los patrones encapsulados por la plataforma Hibernate para acceso a datos. Por tal razón se deberá hacer uso de las herramientas de configuración de dicha plataformas a través de las tramas XML. Las sentencias SQL deberán ser incluidas dentro de esta capa de configuración (Solo en el caso que no puedan ser implementadas como procedimientos almacenados).
- Se sugiere el uso de cualquier implementación de lógica DML (Data Manipulation Language) dentro de un procedimiento almacenado de la base de datos y el mapeo de sus parámetros de entrada y retorno dentro de los archivos hbm.xml.
- Los procedimientos almacenados que retornarán la información de las diferentes entidades retornarán parámetros de tipo cursores que entregan un mejor desempeño que un resultset.

7.6. Premisas de la arquitectura

El propósito de esta sección es describir las premisas de diseño y otras consideraciones que se deben tener para la definición de la arquitectura a nivel conceptual, lógico y físico para la implementación de DOCUMENT SEARCH.

1. La aplicación tendrá un escenario de interfaz de usuario y será Web. Esto significa que en las estaciones cliente deberá haber un Browser o navegador de Internet.
2. la interfaz Web del usuario deberá ser compatible con Internet Explorer 6.0 o versiones posteriores.
3. El aplicativo contará con patrones y guías de diseño e implementación para todo el sistema que permitirá estandarizar la forma como se diseñe y se desarrolle el mismo.
4. La arquitectura del software será multinivel, basado en plataforma J2EE con un esquema centralizado.
5. La interfaz de presentación al usuario se desarrollará utilizando la tecnología JSP y Servlets para las interfaces Web.
6. La autenticación para el ingreso al sistema estará controlada por DOCUMENT SEARCH en su base de datos. Esto quiere decir que los usuarios que deseen ingresar al sistema deberán de estar previamente registrados y aptos para utilizar el sistema.
7. La aplicación contará con una base de datos local SQL Server para guardar la información de sus procesos de negocio. El método de autenticación en la conexión a esta base de datos será el manejoado nativamente por SQL Server; para esta autenticación se utilizara un usuario genérico definido para tal fin.
8. El sistema manejará los mensajes de error de validaciones de la presentación del usuario en la capa de presentación.

7.7. Vista conceptual de la arquitectura

El objetivo principal de la vista conceptual es identificar los objetos del proceso en base a la prioridad de los requerimientos funcionales este sin entrar en detalle de

su definición y tipo de información a utilizar. A continuación se ilustra el diseño de la arquitectura conceptual

7.7.1. Document Search

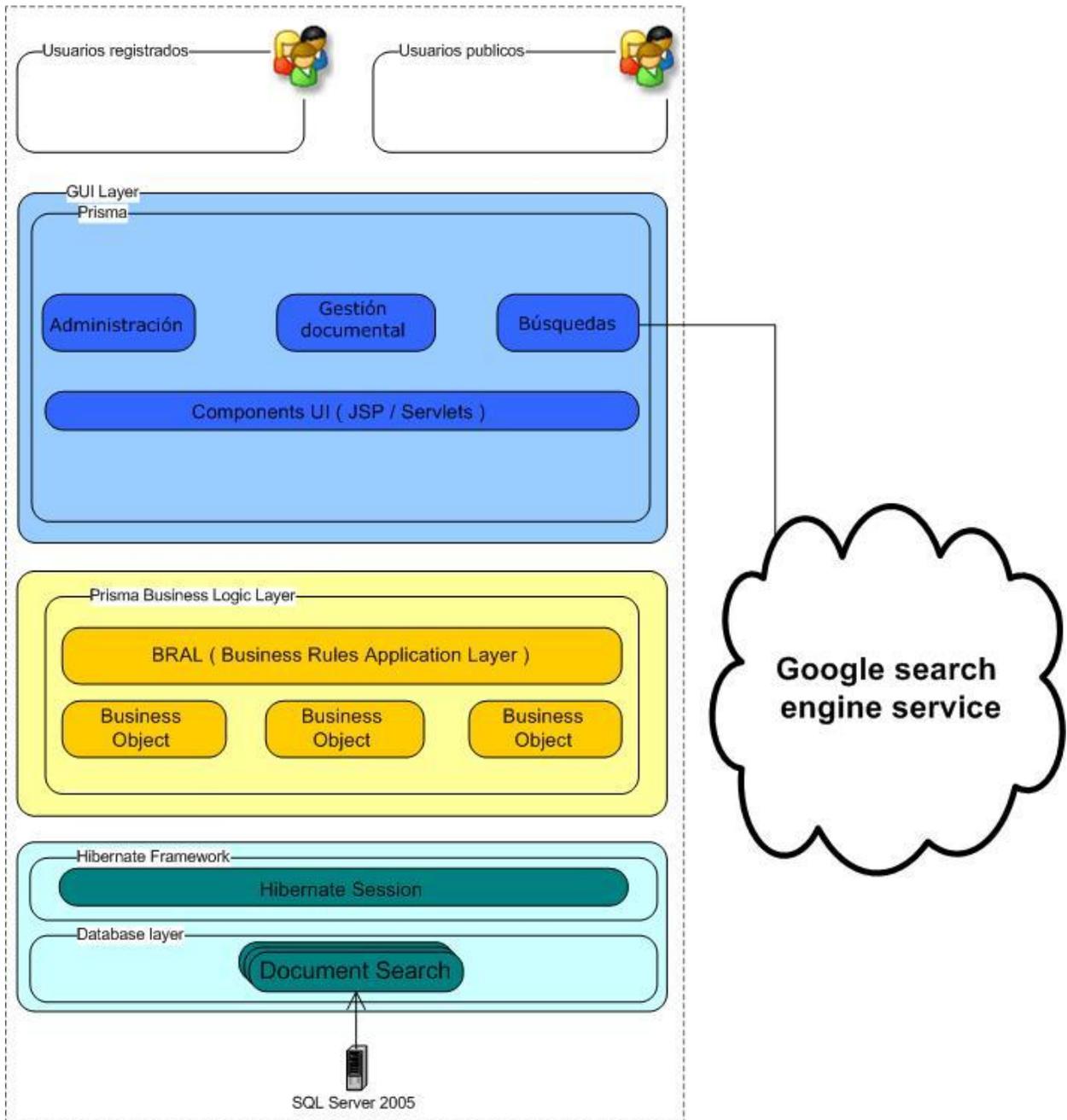


Ilustración 1: Arquitectura de la app

7.7.2 GUI (User Interface) Layer

Este nivel contiene la lógica requerida para exponer los servicios de DOCUMENT SEARCH a los diferentes usuarios involucrados, usando una interfaz gráfica. Está conformado por los siguientes elementos:

7.7.2.1 Usuarios

Los usuarios que utilicen el sistema deben tener acceso al aplicativo a través de sus computadores mediante un browser, dichos usuarios deberán de ser los responsables de la ejecución de las diferentes funcionalidades que ofrece el sistema. Estos pueden ser usuarios no autenticados en la aplicación.

Usuarios Registrados:

Los usuarios de la intranet tendrán disponibles toda la interfase administrativa, creación de documentos y relacion de recursos con articulo existentes en el sistema.

Los requerimientos de los usuarios serán formateados usando HTML, Flash y XML en caso de ser necesario y las validaciones que se requieran realizar en el cliente se harán con JavaScript 1.2.

Usuarios Publicos

Las funciones que serán de exposición abierta a través de la web serán la opción búsqueda y visualización de articulos.

Los requerimientos de los usuarios serán formateados usando HTML, Flash y XML en caso de ser necesario y las validaciones que se requieran realizar en el cliente se harán con JavaScript 1.2.

7.7.3 FrontEnd Document Search

Los usuarios de Document Search podrán interactuar directamente con la aplicación a través del conjunto de opciones de navegación, búsqueda y administración. Para ello se utilizarán interfaces graficas que ayuden al usuario en esta interacción; la implementación de estas interfaces graficas se desarrollaran bajo tecnología JSP y Servlets.

7.7.4 Business Layer

Son los elementos que interactúan directamente con las interfaces del usuario (UI Layer) proporcionándoles la funcionalidad de la aplicación. La capa intermedia contiene los elementos encargados en el sistema de implementar la lógica del negocio así como de implementar el acceso a las fuentes de datos y módulos de integración con sistemas externos.

Este nivel está compuesto por los elementos que se describen a continuación:

7.7.4.1 Objetos y entidades de negocio

Las aplicaciones requieren que los datos sean transmitidos a través de las capas, desde el acceso a datos hasta el usuario final. Ejemplo:

Usuario, Documento, Autor, etc.

7.7.4.2 Flujos del proceso de negocio

Sin importar si un proceso del negocio consta de un “step” o de un flujo del negocio orquestado, el sistema deberá de proveer objetos que implementen la lógica del negocio. En este caso, los componente de negocio deben manejar la

lógica que sea específica de los procesos definidos por Document Search relacionados con el proceso de configuración; los componentes del negocio son los encargados de ejecutar las reglas y tareas asociadas al negocio e interactuar con los objetos del negocio.

Se definen entonces los componentes especializados para el manejo de los procesos intrínsecos dentro de los procesos de Document Search, ejemplo: Usuario, Documento, Autor.

Después de que un dato ingresa al proceso de usuario, se puede dar que este dato ingrese a un proceso de Document Search. Este dato disparará un evento que debe ser fácilmente manipulado y modificado. Por esta razón se implementará un componente que brindará los servicios de atender las solicitudes de acuerdo al evento de negocio, este deberá estar mapeado al componente de reglas para su posterior procesamiento. La implementación de este patrón permitirá aislar los componentes de los eventos generados lo que implicaría mayor facilidad de cambio en los mismos.

7.7.5 Data Layer

Dentro del sistema, esta capa provee los elementos necesarios para comunicarse con el sistema de almacenamiento de datos.

7.7.5.1 Data Access Layer:

Esta capa comprende todos los componentes que serán implementados para realizar la conectividad a la base de datos de Document Search. Al interior de esta capa se deberán construir cada uno de los componentes que representan las

conexiones con la base de datos, aquellos que cohesionadamente implementan todas y cada una de las actividades necesarias para cumplir con las responsabilidades de cada proceso de negocio soportado por las entidades y objetos de la capa superior de negocio. Cada uno de los componentes de acceso a datos que brindan esta funcionalidad a los objetos deberá soportar el uso de transacciones.

Los procesos de acceso a datos se basarán en el framework de Hibernate para la ejecución de sentencias. En la medida de lo posible todas las sentencias SQL que se dispongan para manipulación de datos deberán ser construidas dentro de procedimientos almacenados con las características detalladas en las premisas y lineamientos de este documento.

Para obtener los procedimientos almacenados, se realizará el mapeo de los parámetros de salida y entrada, así como los resultados entregados del mismo como parámetros de tipo cursor, para esto se usaran los tags sql-query de configuración de Hibernate.

Google search engine service

La aplicación usa Google AJAX Search API, la documentación de esta herramienta se encuentra en <http://code.google.com/apis/ajaxsearch/>.

La integración con esta herramienta se encuentra en el archivo /DocumentSearch/WebContent/site/jsp/busqueda.jsp

8. Cronograma

	 Nombre de tarea	Duración	Comienzo	Fin	Pre
1	<input type="checkbox"/> TESIS	110 días	lun 5/7/07	vie 10/5/07	
2	<input type="checkbox"/> Realizar una investigación global sobre las generalidades del tema	15 días	lun 5/7/07	vie 5/25/07	
3	Realizar la investigación global	12 días	lun 5/7/07	mar 5/22/07	
4	Realizar una síntesis sobre los aspectos más relevantes	3 días	mié 5/23/07	vie 5/25/07	3
5	<input type="checkbox"/> Realizar una investigación sobre casos de éxito, en partes específicas del tema que se aborda	10 días	lun 5/28/07	vie 6/8/07	2
6	Realizar la investigación sobre casos de éxito	8 días	lun 5/28/07	mié 6/6/07	
7	Realizar una síntesis sobre dichos casos	2 días	jue 6/7/07	vie 6/8/07	6
8	Realización de la propuesta metodológica	40 días	lun 6/11/07	vie 8/3/07	5
9	<input type="checkbox"/> Implementación de un piloto usando la propuesta metodológica	45 días	lun 8/6/07	vie 10/5/07	8
10	Diseño	10 días	lun 8/6/07	vie 8/17/07	
11	Implementación	30 días	lun 8/20/07	vie 9/28/07	10
12	Pruebas	5 días	lun 10/1/07	vie 10/5/07	11

Ilustración 2: Cronograma - Actividades

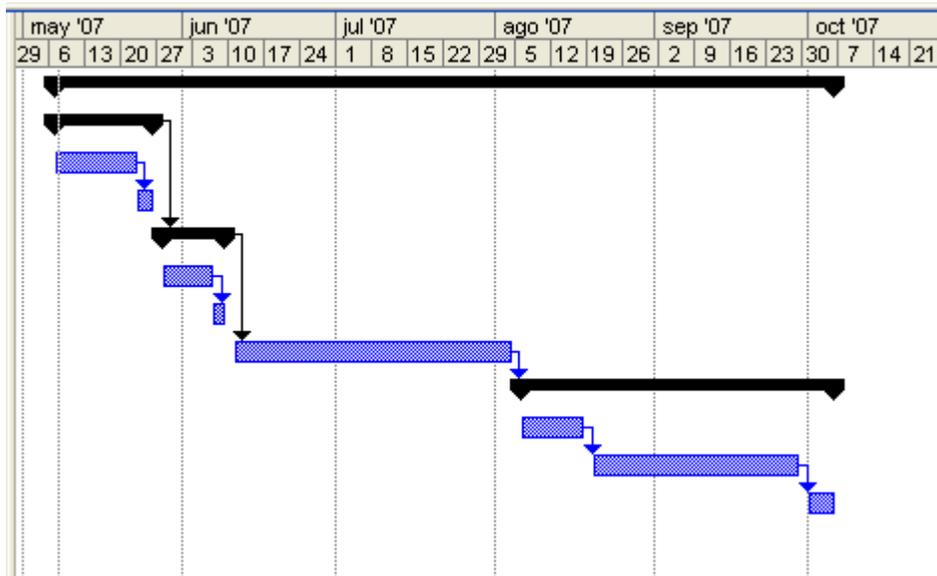


Ilustración 3: Cronograma - Tiempos

9. INTEGRANTES

9.1. *Ejecutores del proyecto*

- Jehudi Castro Sierra
Estudiante de Ingeniería de Sistemas
Código:
E-mail: jcastros@eafit.edu.co
Teléfono:
Celular:

- Juan David Zapata M.
Estudiante de Ingeniería de Sistemas
Código: 200010027010
E-mail: jzapata6@eafit.edu.co
Teléfono: 4160273
Celular: 3006011164

9.2. *Asesor del proyecto*

- Jorge Hernán Abad Londoño
E-mail: jabad@assist.com.co
Teléfono:
Celular:

10. BIBLIOGRAFIA

- <http://exlibris.usal.es/merlo/escritos/calidad.htm>
- <http://www.ts.ucr.ac.cr/~historia/articulos/teo-arch.htm>
- Bearman, David. "Archival strategies", En: American Archivist, Vol. 58. Fall 1995, S.A.A, Chicago, U.S.A.
- Alexander, J.; Tate, M.A. (Alexander; Tate, 1999). Web wisdom: how to evaluate and create information quality on the Internet. Mahwah: Lawrence Erlbaum Associates, 1999.
- Codina, L. (Codina, 2000b). Parámetros e indicadores de calidad para la evaluación de recursos digitales. En Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: retos y soluciones de los profesionales de la información. Bilbao: Universidad del País Vasco, 2000, p. 135-144.
- Cooke, A. Neal-Schuman authoritative guide to evaluating information on the Internet. New York: Neal-Schuman, 1999.