

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

**SISTEMA DE INFORMACIÓN DE DETECCIÓN DE PLAGIO EN
DOCUMENTOS DIGITALES USANDO EL MÉTODO DOCUMENT
FINGERPRINTING**

Tesis para optar por el Título de Ingeniero Informático, que presenta el bachiller:

Fernando Emilio Alva Manchego

ASESOR: Ing. Manuel Tupia Anticona

Lima, enero del 2010

Índice de Anexos

Anexo 1 – Product Backlog

Anexo 2 – Sprint Backlog

Anexo 3 – Estándar de Programación

Anexo 1 - Product Backlog

1 Introducción

En el presente documento se especifican los requerimientos del producto a desarrollar, empleando el Product Backlog, documento establecido por la metodología SCRUM. Cabe señalar que, si bien este documento debería originarse a partir de las Historias de Usuario, por la naturaleza del proyecto (no existen usuarios cliente), se decidió realizar directamente el Product Backlog.

2 Especificación de requerimientos

En el siguiente cuadro se presentan los requerimientos del sistema, indicando, para cada uno: Identificador (ID), Nombre, Importancia (ver leyenda para valores posibles), Estimación inicial (ver leyenda para valores posibles), Comprobación (pasos a seguir para poder comprobar que el requerimiento ha sido cumplido) y Notas (indicaciones adicionales sobre el cumplimiento del requerimiento).

ID	Nombre	Imp.	Est. Inicial	Comprobación	Notas
1	Registrar categoría de documento	5	1.5	Ingresar al sistema, seleccionar la opción de “registro de categoría”, indicar el nombre de la categoría y la categoría padre. Salir de la pantalla, ingresar nuevamente y verificar	Las categorías pueden tener varias hijas, en varios niveles, con un máximo de 2.

				que se muestre en el árbol de categorías padre posibles.	
2	Registrar documento en el repositorio	5	3	Ingresar al sistema, seleccionar la opción de “registro de nuevo documento”, indicar los datos del archivo que se requiere registrar. Ir a la sección de “búsqueda de documentos” y verificar que se encuentre el archivo ingresado.	Los datos a considerar son: Nombre del archivo, Autor, Año e ISBN
3	Registrar documento en el repositorio antes de comparar	4	2	Ingresar al sistema, seleccionar la opción “evaluar documento”, seleccionar “registrar nuevo documento para evaluar”. Luego de ingresar los datos, se debe continuar con el proceso de análisis de plagio.	El proceso de registro debe ser exactamente igual al que se sigue en el ítem 2.
4	Detectar similitud total o parcial de un documento	5	4	Ingresar al sistema, seleccionar la opción “evaluar documento”, indicar el documento que se desea comparar con los que se encuentran en el repositorio. Se debe mostrar el resultado de la comparación, indicando si la totalidad o parte del documento en cuestión se encuentra contenida en alguno(s) de los documentos del repositorio.	Se deben indicar los datos de los documentos del repositorio en los que se han encontrado coincidencias.
5	Detectar plagio por copia directa	5	3	Ingresar al sistema, seleccionar la opción “evaluar documento”, indicar el documento que se desea comparar	No se considera detección de parafraseo de oraciones, ni

				con los que se encuentran en el repositorio. Se debe mostrar el resultado de la comparación, permitiendo la detección de textos transcritos completamente, sin modificación alguna, de algún documento del repositorio al documento que es evaluado.	sustitución por sinónimos.
6	Realizar comparación contra un subgrupo del repositorio.	2	2	Ingresar al sistema, seleccionar la opción de “analizar documento”, indicar el documento que se desea evaluar, indicar contra qué grupo de archivos del repositorio se desea realizar la comparación. Se debe mostrar el resultado de la comparación del documento únicamente contra los documentos seleccionados del repositorio.	El documento a evaluar debe estar registrado previamente en el repositorio.
7	Visualizar resultado de comparación (gráfico)	3	4	Ingresar al sistema, seleccionar la opción de “analizar documento”, indicar el documento que se desea comparar con los que se encuentran en el repositorio (o algún subgrupo de este). Se debe mostrar el resultado de la comparación resaltando las oraciones del documento original que han sido copiadas de algún documento del repositorio. Asimismo, se deben indicar los datos del documento al que corresponde dicho texto que ha sido plagiado.	Se puede emplear un color de resaltado diferente para cada documento con el que se presente coincidencias.

8	Visualizar estadísticas de comparación	2	3	Ingresar al sistema, seleccionar la opción de “evaluar documento”, indicar el documento que se desea comparar con los que se encuentran en el repositorio(o subgrupo de este). Se debe mostrar el resultado de la comparación, indicando, mediante un gráfico, el porcentaje del texto evaluado que forma parte de los documentos que se encuentran en el repositorio.	Se puede emplear un gráfico circular, donde cada porcentaje de similitud por documento sea una sección de diferente color.
9	Registrar Preferencias	2	2	Ingresar al sistema, seleccionar la opción “preferencias”, indicar las preferencias correspondientes. Salir de la pantalla, ingresar nuevamente y verificar que los datos ingresados antes se encuentren registrados.	Las preferencias a considerar son las relacionadas con la conexión al repositorio.

Leyenda:

Importancia: 1 – Más bajo, 5 – Más alto.

Estimación: Días-Hombre. Valores válidos: ≥ 0.5

Anexo 2 - Sprint Backlogs

1 Introducción

En el presente documento se describen las tareas a realizar para el cumplimiento de cada requerimiento del *Producto Backlog*, empleando los *Sprint Backlog*, indicados por la metodología SCRUM.

2 Cuadros de descripción de tareas

Se presentan dos cuadros, uno para cada versión del producto (entregables funcionales). Dentro de cada cuadro se especifican los requerimientos que la versión debe cumplir. Para cada requerimiento se indica: Identificador de requerimiento en el *Product Backlog* (ID), Nombre (nombre del requerimiento), Importancia, Estimación Inicial y Tareas (lista de tareas que se deben de realizar para poder cumplir con el requerimiento que se especifica).

1.1 Sprint Backlog Versión Alpha

Se presenta el cuadro correspondiente a los requerimientos de la versión Alpha del producto, la cual incluye los requerimientos mínimos de funcionalidad.

ID	Nombre	Imp.	Est. Inicial	Tareas
1	Registrar categoría de documento	5	1.5	<ul style="list-style-type: none"> • Implementar el formulario de registro de categoría. Debe de ingresarse el nombre de la categoría y la ruta de categorías padres que poseerá (emplear un árbol de categorías). • Agregar los comandos en la UI para el ingreso al formulario de registro de categorías. • Implementar las unidades de prueba correspondientes y ejecutarlas.
2	Registrar documento en el repositorio	5	3	<ul style="list-style-type: none"> • Implementar formulario de ingreso de datos de documento nuevo a registrar. Los datos a registrar deben ser: título (obligatorio), autor (obligatorio), categoría (obligatorio), año (opcional), número estándar (opcional). • Implementar la validación de los datos que se ingresen en el formulario de registro. • Implementar la persistencia de los datos en el repositorio. • Implementar el listado de documentos registrados. • Agregar los comandos en la UI para el ingreso al formulario de registro de nuevo documento y al listado de documentos registrados. • Implementar las unidades de prueba correspondientes y ejecutarlas.

4	Detectar similitud total o parcial de un documento	5	2	<ul style="list-style-type: none"> • Implementar la pantalla de análisis de documento sospechoso. • Agregar los comandos en la UI para el ingreso a la pantalla de evaluación de documento sospechoso. • Implementar las unidades de prueba correspondientes y ejecutarlas.
5	Detectar plagio por copia directa	5	3	<ul style="list-style-type: none"> • Implementar la obtención del texto del documento en forma de una sola cadena. • Implementar el algoritmo <i>Winnowing</i> que reciba como parámetro la cadena que contiene el texto del documento. • Conectar el proceso de registro de documento con el algoritmo <i>Winnowing</i>. • Implementar las unidades de prueba correspondientes y ejecutarlas.

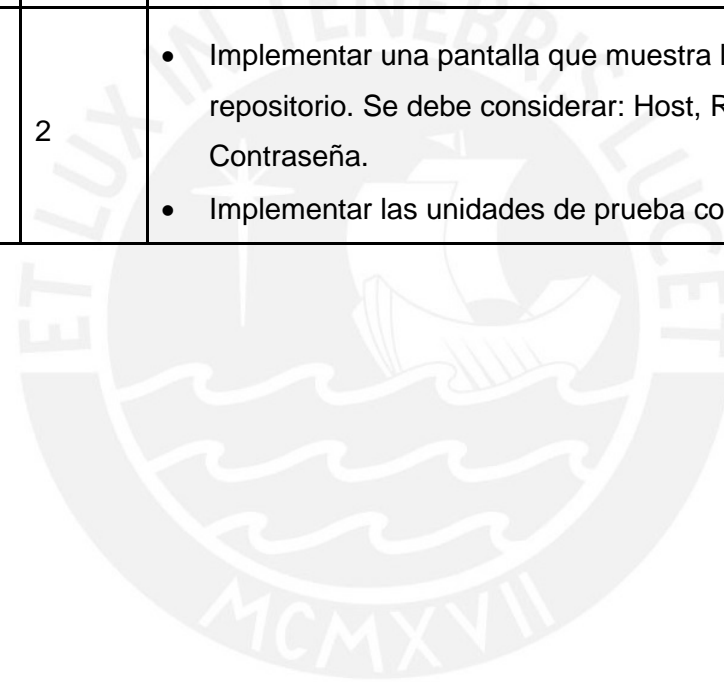
1.2 Sprint Backlog Versión Beta

Se presenta el cuadro correspondiente a los requerimientos de la versión Beta del producto, la cual incluye la totalidad de los requerimientos del producto.

ID	Nombre	Imp.	Est. Inicial	Tareas
3	Registrar documento	3	2	<ul style="list-style-type: none"> • Agregar la opción de registro previa a comparación en el formulario de análisis de

	en el repositorio antes de comparar			<p>documento.</p> <ul style="list-style-type: none"> • Conectar el formulario de registro con la pantalla de evaluación. • Implementar las unidades de prueba correspondientes y ejecutarlas.
6	Realizar comparación contra un subgrupo del repositorio.	2	2	<ul style="list-style-type: none"> • Agregar la opción de seleccionar un grupo de documentos del repositorio (categoría) contra el cual realizar la comparación. • Conectar la pantalla de evaluación de documento sospechoso con la de listado de categorías de documentos en el repositorio y permitir la selección de alguna(s) de ellas. • Modificar la implementación del proceso de comparación para que permita la evaluación de un documento contra un subgrupo del total y no necesariamente contra todo el repositorio. • Implementar las unidades de prueba correspondientes y ejecutarlas.
7	Visualizar resultado de comparación (gráfico)	3	4	<ul style="list-style-type: none"> • Implementar una pantalla donde muestre el resultado de la comparación, resaltando las oraciones del documento original que han sido copiadas de algún documento del repositorio. Emplear un color diferente para cada documento con el que presente coincidencias. • Implementar las unidades de prueba correspondientes y ejecutarlas.
8	Visualizar estadísticas de comparación	2	3	<ul style="list-style-type: none"> • Implementar una pantalla que muestre el resultado de la comparación, indicando mediante un gráfico el porcentaje del texto evaluado que forma parte de los

	(gráfico)			<p>documentos que se encuentran en el repositorio. Se puede emplear un gráfico circular, donde cada porcentaje de similitud por documento sea una sección de diferente color.</p> <ul style="list-style-type: none"> • Implementar las unidades de prueba correspondientes y ejecutarlas.
9	Registrar preferencias	2	2	<ul style="list-style-type: none"> • Implementar una pantalla que muestra las opciones para configurar la conexión al repositorio. Se debe considerar: Host, Repositorio (Base de Datos), Usuario y Contraseña. • Implementar las unidades de prueba correspondientes y ejecutarlas.



Anexo 3 - Estándar de Programación

1 Introducción

En el presente documento se presenta el estándar de programación que se emplea en la implementación del sistema de información que se describe en la memoria del proyecto de tesis. El objetivo del mismo es establecer las normas que se deben seguir al momento de realizar la codificación del sistema, tanto para el estilo a emplear en la misma, como para documentación respectiva.

Para ello, y con el propósito de no “reinventar la rueda”, se emplearán los lineamientos descritos en el documento “The AmbySoft Inc. Coding Standards for Java v17.01d” y que se encuentra disponible en:

<http://www.ambysoft.com/downloads/javaCodingStandards.pdf>

Ese documento es una recopilación de buenas prácticas y estándares de codificación y documentación de código recopilados por Scott W. Ambler, y que servirán como base para la implementación del sistema que se describe en la memoria.

Asimismo, para cada uno de los aspectos particulares de los estándares de programación, se toman en cuenta referencias particulares y se dictan especificaciones independientes, las cuales se describen en las secciones posteriores.

2 Estilo de codificación

Esta sección hace referencia a las consideraciones a tomar en cuenta en el momento de realizar la codificación. Incluye aspectos como nombres de archivos, organización de archivos, sangrías, nomenclatura de paquetes, clases, interfaces, atributos, métodos, variables locales y constantes; inicialización de variables; control de bucles, retorno de valores, entre otros.

Se empleará básicamente lo señalado en el documento “Code Conventions for the Java Programming Language”, publicado por Sun Microsystems y que se encuentra disponible en:

<http://java.sun.com/docs/codeconv/html/CodeConvTOC.doc.html>.

Adicionalmente, se debe de considerar lo señalado al respecto en el documento indicado en la sección 1.

3 Documentación de código fuente

Realizar una buena documentación de código fuente, permite no sólo que el mismo sea más claro y entendible, sino que ayuda a una mejor corrección de errores e implementación de mejoras.

Para poder realizar la documentación del código fuente se emplea la funcionalidad de Javadoc, que, a través del uso de tags en el código, permite la generación automática de la documentación correspondiente a lo implementado.

Es por ello, que se tomarán en cuenta dos documentos publicados por Sun Microsystems, que proveen guías para llevar a cabo esta documentación de forma exitosa, siguiendo las buenas prácticas de los desarrolladores de Java Software.

El primer documento a tomar en cuenta es “Requirements for Writing JAVA API Specifications”, el cual se encuentra disponible en:

<http://java.sun.com/j2se/javadoc/writingapispecs/index.html>.

En él se listan recomendaciones acerca del contenido que debe incluirse al momento de realizar la documentación del código fuente, al mismo tiempo de cómo deben escribirse y en qué posición dentro de los archivos del código fuente.

El segundo documento a tener en consideración es “How to Write Doc Comments for Javadoc.”, que se encuentra disponible en:

<http://java.sun.com/j2se/javadoc/writingdoccomments/index.html>.

En este documento, se indican las convenciones a tomar en cuenta para el uso de los tags a emplear para la generación del javadoc. Indica, entre otras cosas, qué tags utilizar para cada especificación, cuándo no incluir un determinado tag, entre otros.

No debe olvidarse que junto con estos documentos, se deben tomar en cuenta las recomendaciones del documento señalado en la sección 1.