

DESARROLLO DE UN SISTEMA DE IDENTIFICACIÓN DE PROCESOS  
JUDICIALES EN IMÁGENES DE CARTELERAS DE NOTIFICACIONES POR  
ESTADO

ANDRÉS TORRES CANO

ESTEBAN SANTA GARCÍA

Anteproyecto

UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
FACULTAD DE INGENIERÍAS  
INGENIERÍA EN SISTEMAS Y COMPUTACIÓN

2013

# CONTENIDO

	pág.
<b>1. Título</b>	<b>1</b>
<b>2. Definición del problema</b>	<b>2</b>
<b>3. Justificación</b>	<b>2</b>
<b>4. Objetivos Generales y específicos</b>	<b>3</b>
4.1. Objetivo General	4
4.2. Objetivos Específicos	4
<b>5. Marco Referencial</b>	<b>5</b>
5.1. Marco de Antecedentes	5
5.2. Marco Teórico	5
5.2.1. Coincidencia aproximada de cadenas	5
5.2.2. Edit Distance	6
5.3. Marco Conceptual	6
5.3.1. Proceso judicial	6
5.3.2. Notificación por estado	6
5.3.3. Auto	6
5.3.4. Trámite	6
5.3.5. Coincidencia Aproximada de cadenas	7
5.3.6. Reconocimiento de texto	7
5.3.7. OCR	7
5.3.8. Probabilidad	7
<b>6. Método o estructura de la unidad de análisis, criterios de validez y confiabilidad</b>	<b>8</b>
<b>7. Diseño metodológico</b>	<b>8</b>
7.1. Hipótesis	9
7.2. Tipo de investigación	9
7.3. Población	9
7.4. Unidad de análisis	9
7.5. Muestra	9
7.6. Variables	10
7.7. Diseño de instrumentos	10
<b>8. Esquema Temático</b>	<b>12</b>

9. Personas que participan en el proyecto	13
10. Recursos disponibles	14
11. Cronograma	15
12. Bibliografía	16
13. Posibilidades de publicación	17

# **1. Título**

”Desarrollo de un sistema de identificación de procesos judiciales en imágenes de carteleras de notificaciones por estado.”

## **2. Definición del problema**

Es necesario disminuir el tiempo que toma encontrar una lista de procesos judiciales en diferentes documentos digitales de notificaciones por estado de distintos juzgados.

Para varias empresas que prestan servicios para abogados, el proceso de análisis de las fotografías de las carteleras de notificaciones por estado publicadas en los juzgados de Pereira, se convierte en un problema dado que no existe una herramienta informática con capacidad de detectar en su totalidad la información contenida en dichas imágenes, forzando a revisar cientos de procesos de forma manual.

### 3. Justificación

Según la sociología jurídica, rama encargada del estudio de la relación entre la justicia y la sociedad, los procesos judiciales se generan por interacciones sociales, ya que estas relaciones están vinculadas a temas como la violencia política, la democracia, el narcotráfico, los derechos humanos y efectos del desarrollo de la economía como lo es la migración del campo a la ciudad (Revista Prolegómenos - Derecho y Valores pp 119-121 2011). Es Colombia un país considerado una economía en desarrollo, donde se presentan avances significativos en la economía con crecimiento del Producto Interno Bruto de 4.0 % para el 2010, 6.6 % para el 2011 y 4.0 % para el 2012 (fuente: [www.dane.gov.co/files/investigaciones/.../pib/cpPIBIVtrim12.pdf](http://www.dane.gov.co/files/investigaciones/.../pib/cpPIBIVtrim12.pdf)). *Con un índice de desarrollo humano*

En un servicio integral de gestión de procesos judiciales es importante contar con información oportuna y a tiempo. Actualmente los juzgados publican la información en carteleras llamadas notificación por estado, las personas a cargo del servicio de gestión de procesos judiciales se debe desplazar diariamente a consultar estas carteleras, tomando fotografías que posteriormente se procesan manualmente para integrar estos datos a la plataforma. Sumando el factor de crecimiento del volumen de procesos en la rama judicial, con un indicador interno empresarial de crecimiento de clientes que va ligado al número de procesos registrado en el servicio, la tarea de revisión y comparación de la información generada en los juzgados con la registrada en el servicio se convierte en algo colosal.

## **4. Objetivos Generales y específicos**

### **4.1. Objetivo General**

Desarrollar una herramienta que indique las probabilidades que un proceso judicial se encuentre en un conjunto de documentos digitales, buscando por radicado o nombre del demandante.

### **4.2. Objetivos Específicos**

- Realizar un estudio de diferentes métodos de búsqueda de texto.
- Realizar un módulo de preprocesamiento de documentos para prepararlos para la búsqueda.
- Diseñar e implementar una aplicación que permita encontrar el radicado y nombre del demandante en un conjunto de documentos.
- Validar el nivel de efectividad la aplicación implementada.
- Analizar los resultados obtenidos.

## 5. Marco Referencial

### 5.1. Marco de Antecedentes

Los juzgados en Colombia están obligados a publicar la información de los estados judiciales a primera hora hábil del día siguiente al que fueron generados, para cumplir con esta exigencia la gran mayoría de juzgados instala carteleras en las cuales pegan una o varias hojas donde se encuentra impresa la lista de procesos que se notifican por estado ese día. Estas listas son de carácter público y cualquier persona que esté interesada en ellas puede consultarlas libremente.

Para los abogados es importante contar con la información de sus procesos de manera oportuna y para lograrlo cada uno utiliza el método que considere más conveniente, algunos visitan los juzgados diariamente, otros contratan personas que se encargan de revisar los listados y recientemente ayudados por las tecnologías de la información han surgido servicios donde se satisfacen estas necesidades. Las personas encargadas de la prestación de este servicio basado en las tecnologías de la información, realizan recorridos diarios tomando fotografías o escaneando las listas de estados en cada juzgado, para posteriormente procesar la información y enviarla a sus clientes.

En la tesis de maestría de Fernando Casanovas Martín, llamada "Approximate string matching algorithms in art media archives" se intenta crear un sistema para centralizar y facilitar el acceso de los usuarios a un número significativo de archivos multimedia en Europa. Aquí abordan el problema de buscar en los archivos los nombres de los autores, donde el sistema desarrollado debe encontrar las coincidencias tolerando errores de escritura y abreviaciones.

El anterior documento es importante porque en el desarrollo de la investigación se evalúan distintos algoritmos de coincidencia aproximada de cadenas, dando un punto de partida importante para el desarrollo de la investigación que se piensa llevar a cabo.

A pesar de que existen diversas herramientas que permiten la identificación de texto contenido en imágenes, el proceso de revisión de estados es una actividad que se realiza manualmente, por lo tanto, es necesario desarrollar una herramienta que ayude a agilizar este proceso.

### 5.2. Marco Teórico

#### 5.2.1. Coincidencia aproximada de cadenas

La coincidencia aproximada de cadenas o también conocido como búsqueda difusa de cadenas, contempla encontrar patrones de texto dentro de un texto más grande, permitiendo alguna cantidad de errores en la coincidencia. Esto significa que para que



una cadena coincida con otra, no es necesario que tengan la misma longitud ni que coincidan exactamente todos los caracteres.

Para poder buscar estas cadenas difusas, "lo primero que se necesita es crear un modelo de error, el cual define qué tan diferentes son dos cadenas. Esta idea de qué tan diferentes son dos cadenas es llamado distancia entre cadenas." (Casanovas Martín, Fernando. Approximate string matching algorithms in art media archives p 7)

### **5.2.2. Edit Distance**

Edit distance es una distancia que mide la diferencia entre dos secuencias o cadenas de texto sobre un alfabeto. Esta distancia mide el mínimo número de operaciones necesarias para convertir una cadena en la otra. Una operación se considera la inserción, la sustitución o la remoción de un solo caracter.

## **5.3. Marco Conceptual**

### **5.3.1. Proceso judicial**

Es el mecanismo encargado de proteger el derecho al debido proceso establecido por la constitución, donde se garantiza un juez o un tribunal para que se encargue de juzgar las actuaciones judiciales y administrativas bajo el amparo de la ley. Quien sea sindicado tiene derecho a la defensa, a ser representado por un abogado, a presentar pruebas y controvertir las decisiones que se tomen en su contra. (constitución política de Colombia artículo 29, Código general del proceso)

### **5.3.2. Notificación por estado**

Notificación de la decisión de un juez que según lo establecido en la ley no deba hacerse personalmente, esta se hace por medio de anotación en estados los cuales deben ser publicados a la primera hora laboral del día siguiente a la fecha en que es generado.

### **5.3.3. Auto**

Son todas las decisiones de un juez, cualquiera fuere la instancia en que se pronuncien, que no sean decisiones sobre las pretensiones de la demanda o las excepciones que no tengan el carácter de previas.

### **5.3.4. Trámite**

Si el recurso se formula por escrito, este se mantendrá en la secretaría por dos días en traslado a la parte contraria, sin necesidad de que el juez, lo ordene; surtido el traslado se decidirá el recurso. El secretario dará cumplimiento al artículo. La reposición interpuesta en audiencia y diligencia se decidirá allí mismo, una vez oída la parte contraria si

estuviere presente. Para este fin cada parte podrá hacer uso de la palabra hasta por quince minutos. (código del procedimiento civil artículo 349)

**Procesos con trámite:** Proceso que se encuentra abierto y en determinado periodo se ha desarrollado acorde lo que se define como trámite en el código del procedimiento civil.

**Procesos sin trámite:** Proceso que se encuentra abierto pero en determinado periodo no se ha desarrollado acorde a lo que se define como trámite en el código del procedimiento civil.

### **5.3.5. Coincidencia Aproximada de cadenas**

”La coincidencia aproximada de cadenas es una coincidencia de cadenas que permite errores. El objetivo es realizar una coincidencia de cadenas en un texto donde una o ambas fuentes han sufrido algún tipo de corrupción.”(Casanovas Martín, Fernando. Approximate string matching algorithms in art media archives p73 )

### **5.3.6. Reconocimiento de texto**

”El reconocimiento de texto es ejecutado para convertir una imagen que contiene texto impreso o manuscrito en un formato que puede ser entendido por un computador (Ej. ASCII o unicode)” (Introduction to Document Analysis and Recognition)

### **5.3.7. OCR**

”Los sistemas de reconocimiento de texto frecuentemente dividen las palabras en caracteres y luego asignan una clase a cada objeto aislado. Cuando se trata de texto impreso, esta aproximación es generalmente conocido como Reconocimiento ptico de Caracteres (OCR por sus siglas en inglés)”(Introduction to Document Analysis and Recognition)

### **5.3.8. Probabilidad**

La probabilidad permite expresar numéricamente la posibilidad que un evento ocurra. Se afirma que algo va o no va a ocurrir, pero no se está seguro de ello. La probabilidad es el grado de confianza en la validez de dicha afirmación.(Probabilidad y estadística para ingeniería y ciencias)

## **6. Método o estructura de la unidad de análisis, criterios de validez y confiabilidad**

Para garantizar la confiabilidad de la información arrojada por el sistema, se tomarán como entrada documentos que han sido revisados de manera manual y se sabe con exactitud cuántos y cuáles procesos deben coincidir. Los resultados de la salida del aplicativo serán comparados con éstos, y así se podrá validar el grado confianza de las coincidencias halladas. De la misma manera, se verificará que al ejecutar la aplicación varias veces con un mismo conjunto de documentos, debe generar salidas idénticas en cada una de las ejecuciones.

## **7. Diseño metodológico**

### **7.1. Hipótesis**

¿Es posible desarrollar una herramienta informática que automatice la tarea de encontrar una lista de procesos judiciales en diferentes documentos de notificaciones por estado de distintos juzgados?.

### **7.2. Tipo de investigación**

Esta investigación tomará un enfoque cuantitativo.

### **7.3. Población**

Los documentos impresos de notificaciones por estado de los juzgados civiles municipales de menor cuantía de la ciudad de Pereira.

### **7.4. Unidad de análisis**

Fotografías de los documentos mencionados en la población.

### **7.5. Muestra**

La muestra está compuesta por las fotografías de estados y traslados de los juzgados civiles municipales de Pereira de los días:

- 19 de septiembre de 2013.
- 20 de septiembre de 2013.
- 24 de septiembre de 2013.
- 25 de septiembre de 2013.
- 26 de septiembre de 2013.
- 27 de septiembre de 2013.
- 07 de octubre de 2013.
- 08 de octubre de 2013.
- 09 de octubre de 2013.
- 11 de octubre de 2013.

- 15 de octubre de 2013.
- 21 de octubre de 2013.
- 22 de octubre de 2013.
- 23 de octubre de 2013.
- 24 de octubre de 2013.
- 25 de octubre de 2013.
- 28 de octubre de 2013.
- 29 de octubre de 2013.
- 30 de octubre de 2013.
- 31 de octubre de 2013.
- 01 de noviembre de 2013.

## **7.6. Variables**

Las variables serán las siguientes:

- Porcentaje de coincidencias reconocidas correctamente en total.
- Porcentaje de coincidencias reconocidas por cada juzgado.
- Porcentaje de seguridad con el que se reconoció cada proceso.

## **7.7. Diseño de instrumentos**

Para determinar el porcentaje de coincidencias reconocidas por juzgado se usará el formato especificado en la tabla 1, donde se pueden registrar fácilmente la cantidad de procesos que se notifican por estado diariamente, los procesos que debe reconocer y las coincidencias reconocidas, para que de esta manera se pueda registrar un porcentaje de reconocimiento diario y total correspondiente a todos los días medidos para ese juzgado.

Para determinar el total de coincidencias reconocidas se usará el formato descrito en la tabla 1, donde se pueden identificar fácilmente el total de coincidencias resultantes de las medición en cada juzgado y el total de coincidencias reconocidas para que de esta forma se pueda identificar el porcentaje total.

Variable	Definición Operacional	Categorías	Indicadores	Items
Grado afinidad con el perfil solicitado	Grado afinidad con el perfil buscado, de acuerdo a la cantidad de check-ins realizados en lugares marcados como estratégicos.	Sitios Clase A	Recurrencia a lugares con perfiles idénticos o parecidos al ofertado	Cantidad de check-ins por día
				Cantidad de check-ins por semana
				Cantidad de check-ins por mes
				Cantidad de check-ins por año
				Día de la semana que más realiza check-ins
				Hora del día que más realiza check-ins
				Cantidad de comentarios realizados en el lugar
				Cantidad de Amigos encontrados en el lugar
				Frecuencia con que encuentra amigos en el lugar
		Sitios Clase B	Recurrencia a lugares con perfiles complementarios al solicitado	Cantidad de check-ins por día
				Cantidad de check-ins por semana
				Cantidad de check-ins por mes
				Cantidad de check-ins por año
				Día de la semana que más realiza check-ins
				Hora del día que más realiza check-ins
				Cantidad de comentarios realizados en el lugar
				Cantidad de Amigos encontrados en el lugar
		Sitios Clase C	Recurrencia a lugares aleatorios a lugares con perfiles idénticos o parecidos al ofertado	Cantidad de check-ins por día
				Cantidad de check-ins por semana
				Cantidad de check-ins por mes
				Cantidad de check-ins por año
				Día de la semana que más realiza check-ins
				Hora del día que más realiza check-ins
				Cantidad de comentarios realizados en el lugar
				Cantidad de Amigos encontrados en el lugar
		Tipo de usuario	Determina la frecuencia y el tipo de uso que se le da a la red social forsquare	Frecuencia con que encuentra amigos en el lugar
				Cantidad de medallas
				Cantidad de alcaidías
				Tipo de sitio donde más realiza check-ins
				Tipo de sitio del que posee la alcaidía
				Cantidad de check-ins realizados en días promocionales
				Cantidad de comentarios realizados en el lugar
				Frecuencia con que realiza comentarios

**Figura 1.** Tabla de diseño de instrumentos

*aquí va una tabla*

Para medir el porcentaje de reconocimiento acertado por proceso, se usará el formato descrito en la tabla 1, el cual permite registrar claramente el proceso donde se ha reconocido una coincidencia, el día y el porcentaje de reconocimiento que se identificó para este caso.

*aquí va otra tabla*

## **8. Esquema Temático**

El posible esquema temático del documento final propuesto es el siguiente:

**Introducción**

**Problema**

**Justificación**

**Objetivos**

**Marco Referencial**

**Metodología**

**Resultados y discusión**

**Conclusiones**

**Recomendaciones**

## 9. Personas que participan en el proyecto

**Esteban Santa:** Estudiante Ingeniería de Sistemas y Computación, encargado del módulo de preprocesamiento.

**Andrés Torres:** Estudiante Ingeniería de Sistemas y Computación, encargado del módulo de reconocimiento.

**Sebastián Gómez:** Docente Ingeniería de Sistemas y Computación, asesor del proyecto de grado.



## 10. Recursos disponibles

Se cuentan con los siguientes recursos para llevar a cabo el proyecto:

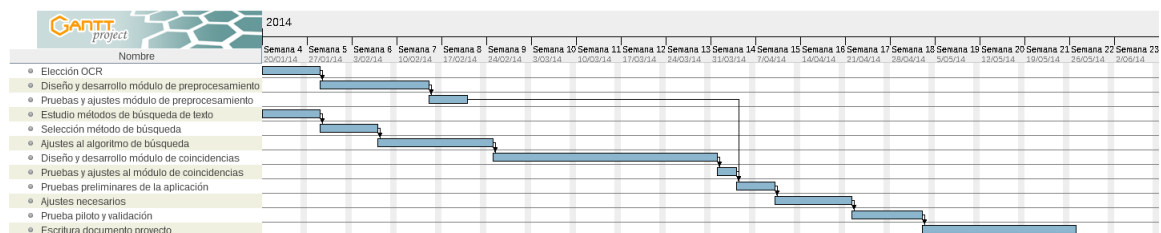
- 3 computadores portátiles.
- 1 computador de escritorio.
- Diferentes distribuciones de sistemas operativos GNU/Linux con licencia GNU GPL Versión 3.
- 2 licencias de Microsoft Windows 7 Professional.

## 11. Cronograma

A continuación se listan las actividades con su estimación de tiempo:

- Elección OCR: 8 días.
- Diseño y desarrollo módulo de preprocesamiento: 15 días.
- Pruebas y ajustes módulo de preprocesamiento: 5 días.
- Estudio métodos de búsqueda de texto: 25 días.
- Selección método de búsqueda: 8 días.
- Ajustes al algoritmo de búsqueda: 15 días.
- Diseño y desarrollo módulo de coincidencias: 30 días.
- Pruebas y ajustes al módulo de coincidencias: 3 días.
- Pruebas preliminares de la aplicación: 5 días.
- Ajustes necesarios: 10 días.
- Prueba piloto y validación: 10 días.
- Escritura documento de proyecto: 20 días.

La figura 2 contiene un diagrama Gantt del cronograma, empezando en enero 20 de 2014:



**Figura 2.** Diagrama Gantt

## 12. Bibliografía

- [1] BAKSHI, Kapil. Considerations for Big Data: Architecture and Approach. En: Aerospace Conference IEEE. Memorias. Big Sky, MT. 2012.
- [2] CUZZOCREA, Alfredo. Analytics over Large-Scale Multidimensional Data: The Big Data Revolution!. En: Conference on Information and Knowledge Management. Memorias. ACM New York, NY, USA 2011. p. 101-104.
- [3] Proceedings of the VLDB Endowment Volume 5 [Online]. Istanbul: LABRINIDIS, Alexandros. Challenges and opportunities with big data August 2012. Disponible en internet: <<http://dl.acm.org/citation.cfm?id=2367572>>
- [4] CHAUDHURI, Surajit, What Next? A Half-Dozen Data Management Research Goals for Big Data and the Cloud. En: International Conference on Management of Data. Memorias. ACM New York, NY, USA 2012. p. 1-4.
- [5] GOPALKRISHNAN, Vivekanand, Big Data, Big Business: Bridging the Gap. En: Knowledge Discovery and Data Mining. Memorias. ACM New York, NY, USA 2012. p. 7-11.
- [6] BORKAR Vinayak, Inside Big Data Management: Ogres, Onions, or Parfaits?. En: Extending Database Technology. Memorias. ACM New York, NY, USA 2012. p. 3-14.
- [7] ZIKOPOULOS, Paul et al. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill. 2012.
- [8] USENIX Association Berkeley [Online]. CA, USA: Dean Jeffrey y Ghemawat Sanjay. MapReduce: simplified data processing on large clusters. 2004 Disponible en internet <<http://dl.acm.org/citation.cfm?id=1251264>>
- [9] MANYIKA, James et al. Big Data: The next frontier for innovation, competition and productivity. Mckinsey Global Institute. May 2011.
- [10] SAS Institute. Big Data: Harnessing a game-changing asset. Reporte de Economist Intelligence Unit. Septiembre 2011.
- [11] YEUNG, Ken. Now counting 6.2 million daily, Foursquare has crossed the 3 billion check-in mark. The Next Web [Online]. 21 noviembre 2012 [Citado 27 noviembre 2012]. Disponible en internet <<http://thenextweb.com/location/2012/11/21/foursquare-has-its-3-billionth-check-in-seeing-growth-of-x/>>

[12] Foursquare. About foursquare[Online]. Disponible en internet <[https:// foursquare.com/about/](https://foursquare.com/about/)>

[13] Harvard Bussines Review. Hbr Insigth Center[Online]. Harvard Bussines Review Octubre - Noviembre 2012 [citado 27 Noviembre 2012]. Disponible en internet <<http://hbr.org/special-collections/insight/big-data>>

[14] Foursquare. Foursquare connects people to businesses[Online]. Disponible en internet <[http:// foursquare.com/business](http://foursquare.com/business)>

[15]Foursquare. Advertise with Foursquare[Online]. Disponible en internet <[http:// business.foursquare.com/advertise/](http://business.foursquare.com/advertise/)>

[16] Hoot Suite. Gestion de redes sociales[Online]. Disponible en internet <[http:// hootsuite.com/features/social-networks](http://hootsuite.com/features/social-networks)>

[17]SCHIFFMAN , Leon G. KANUK , Leslie L. El comportamiento del consumidor. Pearson Educación (Agosto 2005).

[18]KOTLER, Philip. Fundamentos de Marketing (6 edición). Pearson Educación de México. 2003

[19]King, Kevin F., Geolocation and Federalism on the Internet: Cutting Internet Gamblings Gordian Knot (July 14, 2009). Columbia Science and Technology Law Review, Vol. XI, 2010.

## 13. Posibilidades de publicación

Los resultados de la investigación completa serán publicados en un informe final escrito. Algunos resultados parciales que lo ameriten serán publicados en la revista *Scientia et Technica* de la Universidad Tecnológica de Pereira. Entre estos resultados parciales estarían el informe de la *API* de FourSquare, el informe sobre modelos de reconocimiento y el informe sobre paradigma *MapReduce*.

Todos estos escritos serán publicados bajo la licencia Creative Commons Reconocimiento Unported 3.0.