

## Anexo 1: Acta de constitución del Proyecto

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA
Descripción del proyecto	
<p>El proyecto consiste en la implementación de una herramienta que lleve a cabo la simplificación sintáctica de textos de género periodístico en español.</p> <p>Los responsables del proyecto son:</p> <ul style="list-style-type: none"> <li>- BSc. Luis Fernando Muroya – Tesista</li> <li>- MSc. Fernando Alva Manchego – Asesor</li> </ul> <p>El proyecto da inicio el 01 de diciembre del 2014 y tiene como fecha de término el 21 de junio del 2015.</p>	

Definición del producto del proyecto
<p>En el proyecto se implementará una herramienta web que permita simplificar a nivel sintáctico textos periodísticos. Para ello, se implementará las funciones que lleven a cabo dichas operaciones de simplificación, que a saber, son las siguientes:</p> <ul style="list-style-type: none"> <li>- Eliminación</li> <li>- Separación</li> <li>- Cambio</li> <li>- Reordenamiento</li> <li>- Inserción</li> <li>- Mantener (no hacer nada).</li> </ul> <p>Asimismo, se entrenarán clasificadores para tomar la decisión de qué operación de simplificación llevar a cabo. Para ello, se seguirá un marco de trabajo usado para el</p>

aprendizaje supervisado: se recopilará un corpus de textos (el cual se enriquecerá con la ayuda de una herramienta implementada dentro del proyecto) y se extraerá un vector de sus características más importantes.

Cabe mencionar que la implementación de dichos extractores también será parte del proyecto.

El producto será una aplicación web que recibirá como dato de entrada un texto en español y cuya salida será la versión simplificada del mismo, si es que existe.

### Requisitos del proyecto

- El producto deberá estar listo como máximo, 7 días después de finalizado el proyecto.
- Deberá llevarse a cabo reuniones con el asesor, por lo menos, una vez cada quince días.
- El asesor deberá dar su visto bueno a cada entregable del proyecto.

### Objetivos del proyecto

Concepto	Objetivo	Criterio de éxito
Alcance	Implementación de una herramienta web para la simplificación sintáctica de textos y una herramienta de apoyo para la anotación de corpus.	Aceptación del producto por parte del asesor.
Tiempo	Concluir las actividades del proyecto en los tiempos fijados en el cronograma.	Concluir el proyecto, como máximo, el 28 de junio del 2015.
Costo	Concluir el proyecto dentro del presupuesto.	No exceder el presupuesto del proyecto.

Objetivos del producto	
Objetivo	Resultado esperado
Formar un <i>corpus</i> anotado formado por textos de género periodístico en español y sus versiones simplificadas.	Corpus formado por 190 noticias en español anotado con las operaciones de simplificación efectuadas.
Definir el conjunto de características que serán extraídas de los textos periodísticos del <i>corpus</i> .	Funciones de extracción de características implementadas y probadas.
Formar un modelo de clasificadores que tome la decisión de qué operación de simplificación aplicar.	Modelo de clasificadores binarios en serie que definen la operación de simplificación a ser ejecutada.
Implementar el prototipo de una herramienta automática de simplificación de textos en español.	Herramienta que aplica operaciones de simplificación a un texto que recibe como entrada.

Justificación del proyecto
Disminuir tiempos de entrega y costos de simplificación de textos.
Expectativa de éxito debido a resultados positivos en idiomas como portugués e inglés.

Cronograma de hitos del proyecto	
Hito	Fecha programada
Herramienta de apoyo de anotación implementada.	26 de febrero del 2015
Corpus anotado y analizado	24 de marzo del 2015
Extractores implementados y probados	16 de abril del 2015
Clasificadores entrenados con características escogidas.	10 de mayo del 2015
Herramienta web de simplificación de textos implementada	18 de junio del 2015

### Organizaciones que intervienen en el proyecto

Organización	Rol
Grupo Simplext	Proveedor del corpus de estudio.

### Principales amenazas del proyecto

Pérdida de datos del proyecto.
Corpus no adecuado para el trabajo del proyecto.
Curvas de aprendizaje muy grandes.

### Principales oportunidades del proyecto

Diferencia del sistema propuesto con el implementado por Simplext puede llevar a mejores resultados.
--

### Presupuesto inicial

Concepto	Cantidad presupuestada (NS)
Recursos humanos (*)	2900
Otros	200
Reserva de contingencia	200
Reserva de gestión	400

(\*) Monto ficticio. Necesario para el control de tiempos según método de valor ganado (EV).

## Anexo 2: Documento de Gestión de Cambios

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA

Roles de la Gestión de Cambios			
Nombre del rol	Persona asignada	Responsabilidad	Nivel de autoridad
Asesor	FA	Dirimir en decisiones empatadas o tomar decisiones de cambios importantes del proyecto.	Total sobre el proyecto.
Tesista	LM	Decidir qué cambios podrían efectuarse y escalar cambios importantes al asesor.	Autorizar/rechazar solicitudes de cambio.
Profesor	RC	Ver el estado del proyecto y dar sugerencias de mejora.	Solicitar cambios.

Tipos de cambios
<ol style="list-style-type: none"> <li>1. <b>Acción correctiva:</b> este tipo de cambio no pasa por el proceso general del cambio. Es ejecutado inmediatamente después de detectado el error.</li> <li>2. <b>Acción preventiva:</b> este tipo de cambio no pasa completamente por el proceso general del cambio. El asesor puede decidir inmediatamente si llevarla a cabo o no.</li> </ol>

3. **Cambios al plan de proyecto:** este tipo de cambio debe pasar necesariamente por el proceso general del cambio.

<b>Proceso general del Cambio</b>	
Solicitud de cambios: de forma presencial (durante exposiciones o reuniones) o de forma escrita (en correcciones a documentos).	El profesor o el tesista detectan algún cambio posible en el proyecto y lo comunica.
Evaluar impacto: evalúa integralmente el impacto del cambio.	El tesista, en conjunto con el asesor, analiza y evalúan el impacto que tendrá la implementación del cambio sobre el proyecto.
Tomar decisión y re planificar: se toma una decisión en base a la evaluación previa.	El tesista, en conjunto con el asesor, toma una decisión en base a la evaluación (aprobar o no).  Si hay diferencia de opinión con el asesor, se negocia un cambio intermedio.  Si existe una divergencia de opiniones entre el asesor y el profesor, ambos deberán discutir sus opiniones para llegar a una solución.
Implantar el cambio: se ejecuta el cambio.	El tesista lleva a cabo el cambio solicitado y aprobado. Asimismo, registra el cambio en el log del documento o como una nueva versión.
Concluir el proceso de cambio: validar el cambio.	El asesor lleva a cabo la validación de la implantación satisfactoria del cambio.

### Anexo 3: Declaración de Alcance del Proyecto

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA

Descripción del alcance del producto	
Requerimientos	Características
1. Facilidad de uso.	1. Interfaz gráfica de fácil navegación y adaptable al dispositivo.
2. Simplificación debe llegar a nivel establecido como línea base a partir del corpus.	2. Uso de modelo supervisado de aprendizaje de máquina para decidir qué operación aplicar.
3. Convertir un texto (de género periodístico) en otro más sencillo de entender conservando su significado.	3. Uso de una línea base obtenida de aplicar métricas de legibilidad sobre el corpus para una posterior comparación de las métricas del resultado de la herramienta contra esta línea base. 4. Trabaja solo con textos de género periodístico en lenguaje español.

Criterios de aceptación	
Concepto	Criterios de aceptación
1. Técnicos	Interfaz de la herramienta debe ser web. Herramienta debe ser implementada en Java.
2. Calidad	<ul style="list-style-type: none"> <li>- Producto debe ser amigable con el usuario.</li> <li>- Producto no debe tener errores en ningún caso de uso.</li> </ul>

3. Administrativos	Producto debe haberse concluido respetando el cronograma y el presupuesto.
4. Comerciales	No aplican.
5. Sociales	No aplican.

<b>Entregables del proyecto</b>	
<b>Fase</b>	<b>Productos entregables</b>
Obtención y anotación del corpus de noticias en español.	<ol style="list-style-type: none"> <li>1. Corpus original recopilado.</li> <li>2. Herramienta de apoyo para la anotación.</li> <li>3. Corpus anotado.</li> </ol>
Definición e implementación de las funciones extractoras de características.	<ol style="list-style-type: none"> <li>1. Funciones extractoras.</li> <li>2. Pruebas automatizadas de las funciones extractoras.</li> </ol>
Formación del modelo de clasificadores.	<ol style="list-style-type: none"> <li>1. Modelo de clasificadores binarios en serie que toman decisión de efectuar o no operación.</li> </ol>
Implementación de una herramienta cliente-servidor para la simplificación de textos.	<ol style="list-style-type: none"> <li>1. Funciones que ejecutan operaciones de simplificación ("back-end").</li> <li>2. Interfaz web de la aplicación.</li> <li>3. Aplicación en ambiente de producción.</li> </ol>

<b>Exclusiones del proyecto</b>
1. No incluye la recopilación de textos del corpus.
2. No incluye la simplificación manual de los textos del corpus.
3. No se incluye la implementación de los algoritmos empleados en los clasificadores.

**Restricciones del proyecto**

1. Problema del sobreajuste de los datos al corpus de entrenamiento.
2. Disponibilidad de tiempo de los recursos del proyecto.
3. Disponibilidad de los organismos externos (equipo del proyecto Simplext) que colaboran con el proyecto.

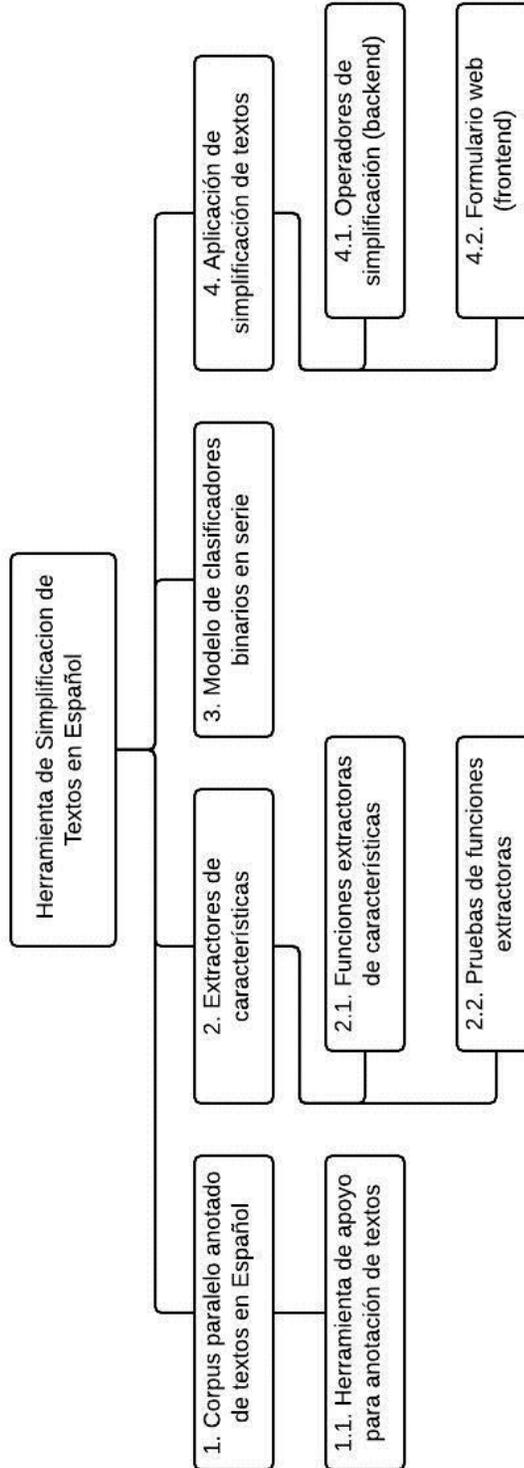
**Supuestos del proyecto**

1. Disponibilidad mínima de 2 horas diarias de los recursos al proyecto.
2. Disponibilidad del presupuesto asignado.
3. Ortografía correcta en las oraciones que componen el corpus de estudio.
4. Disponibilidad continua del asesor y/o profesor.



Anexo 4: EDT y Diccionario del EDT

EDT



## Diccionario del EDT

Componente	Descripción				
<b>Código del paquete de trabajo</b>	1.1				
<b>Descripción del paquete de trabajo</b>	Obtención del corpus paralelo de noticias en español para su posterior anotación.				
<b>Entregable(s)</b>	Corpus paralelo de textos de noticias en español.				
<b>Criterios de aceptación del entregable(s)</b>	<ul style="list-style-type: none"> <li>- Debe estar formado por textos de naturaleza periodística.</li> <li>- Debe haber una distribución uniforme del tipo de noticia.</li> <li>- Las noticias deben estar escritas en español.</li> <li>- Cada noticia debe tener su versión simplificada manualmente.</li> </ul>				
<b>Recursos</b>	Tesista, Asesor.				
<b>Duración estimada</b>	62 días	Fecha inicio:	01/12/2014	Fecha Fin:	31/01/2015
<b>Hito asociado</b>	Ninguno.				
<b>Dependencias</b>	Ninguna.				
<b>Observaciones</b>	Se considera una duración larga debido a las fiestas de Navidad y fin de año.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Ponerse en contacto con el equipo Simplext.</li> <li>- Solicitar al equipo Simplext la autorización para el uso de su corpus de estudio.</li> <li>- Coordinar con el equipo Simplext para el envío del corpus de estudio.</li> </ul>				

Componente	Descripción				
<b>Código del paquete de trabajo</b>	1.2				
<b>Descripción del paquete de trabajo</b>	Análisis, diseño e implementación de una herramienta de apoyo para anotar el corpus de textos con las operaciones efectuadas sobre cada oración del mismo.				
<b>Entregable(s)</b>	Herramienta de apoyo para la anotación de textos				
<b>Criterios de aceptación del entregable(s)</b>	-Cumplir con todos los requerimientos del catálogo de requisitos. -Funcionar para textos de al menos 5 oraciones.				
<b>Recursos</b>	Tesista				
<b>Duración estimada</b>	25 días	Fecha inicio:	01/02/2015	Fecha Fin:	26/02/2015
<b>Hito asociado</b>	Herramienta de apoyo implementada				
<b>Dependencias</b>	1.1				
<b>Observaciones</b>	Las características ofrecidas por el producto deben estar reflejadas y respaldadas en el catálogo der requisitos.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Realizar análisis y diseño de la herramienta.</li> <li>- Implementar la herramienta.</li> <li>- Probar la herramienta.</li> </ul>				

Componente		Descripción			
<b>Código del paquete de trabajo</b>	1				
<b>Descripción del paquete de trabajo</b>	Anotación del corpus de texto.				
<b>Entregable(s)</b>	Corpus de textos de noticias en español anotado con las operaciones de simplificación.				
<b>Criterios de aceptación del entregable(s)</b>	<ul style="list-style-type: none"> <li>- Debe conservarse la información de la oración proveniente del corpus original.</li> <li>- Debe añadirse la información de las operaciones efectuadas sobre cada oración.</li> </ul>				
<b>Recursos</b>	Tesista, Asesor.				
<b>Duración estimada</b>	114 días	Fecha inicio:	01/12/2014	Fecha Fin:	24/03/2015
<b>Hito asociado</b>	Corpus anotado y analizado.				
<b>Dependencias</b>	1.1, 1.2				
<b>Observaciones</b>	Se considera una duración larga debido a las fiestas de Navidad y fin de año.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Obtener el corpus paralelo de noticias en español (1.1).</li> <li>- Analizar propiedades generales del corpus.</li> <li>- Implementar una herramienta para apoyar la anotación del corpus (1.2).</li> <li>- Anotar manualmente el corpus.</li> <li>- Analizar la frecuencia de las operaciones efectuadas sobre el corpus.</li> </ul>				

Componente		Descripción			
<b>Código del paquete de trabajo</b>	2.1				
<b>Descripción del paquete de trabajo</b>	Implementación de las funciones extractoras de características.				
<b>Entregable(s)</b>	Funciones extractoras de características.				
<b>Criterios de aceptación del entregable(s)</b>	- Las funciones deben estar documentadas indicando el propósito de cada una para posterior uso por terceros.				
<b>Recursos</b>	Tesista				
<b>Duración estimada</b>	15 días	Fecha inicio:	25/03/2015	Fecha Fin:	08/04/2015
<b>Hito asociado</b>	Ninguno				
<b>Dependencias</b>	1				
<b>Observaciones</b>	Se creará un extractor por cada característica encontrada. Posteriormente se definirá un subconjunto sobre el cual trabajar.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Definir el conjunto de extractores de características que se necesitará implementar.</li> <li>- Implementar las funciones de extracción de características.</li> </ul>				

Componente	Descripción				
<b>Código del paquete de trabajo</b>	2.2				
<b>Descripción del paquete de trabajo</b>	Pruebas de las funciones extractoras de características.				
<b>Entregable(s)</b>	Pruebas unitarias automatizadas de las funciones extractoras.				
<b>Criterios de aceptación del entregable(s)</b>	- Debe usarse una herramienta de automatización para pruebas unitarias.				
<b>Recursos</b>	Tesista				
<b>Duración estimada</b>	08 días	Fecha inicio:	09/04/2015	Fecha Fin:	16/04/2015
<b>Hito asociado</b>	Extractores implementados y probados.				
<b>Dependencias</b>	2.1				
<b>Observaciones</b>	Ninguna.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Definir los casos de prueba con sus resultados esperados.</li> <li>- Implementar las funciones para llevar a cabo las pruebas unitarias.</li> </ul>				

Componente	Descripción				
Código del paquete de trabajo	2				
Descripción del paquete de trabajo	Implementación y pruebas de las funciones extractoras de características.				
Entregable(s)	Archivos con los vectores de características extraídos del corpus.				
Criterios de aceptación del entregable(s)	- Los archivos deberán tener la estructura y extensión requerida por Weka para ser su fuente de datos (ARFF).				
Recursos	Tesista, Asesor				
Duración estimada	15 días	Fecha inicio:	25/03/2015	Fecha Fin:	19/04/2015
Hito asociado	Ninguno.				
Dependencias	2.2				
Observaciones	Ninguna.				
Tareas	<ul style="list-style-type: none"> <li>- Implementar los extractores de características (2.1).</li> <li>- Probar las funciones implementadas (2.2).</li> <li>- Aplicar las funciones sobre el corpus y dar estructura de archivo ARFF.</li> </ul>				

Componente	Descripción				
<b>Código del paquete de trabajo</b>	3				
<b>Descripción del paquete de trabajo</b>	Formación del modelo de clasificadores.				
<b>Entregable(s)</b>	Modelo de clasificadores definido.				
<b>Criterios de aceptación del entregable(s)</b>	<ul style="list-style-type: none"> <li>- Deberá escogerse el modelo de clasificadores con mejor desempeño según métrica de la medida F con validación cruzada.</li> <li>- Deberá escogerse el mejor subconjunto de características necesarias para ese modelo con método "Leave one out".</li> <li>- Deberá existir un análisis estadístico que justifique la elección.</li> </ul>				
<b>Recursos</b>	Tesista, Asesor				
<b>Duración estimada</b>	21 días	Fecha inicio:	20/04/2015	Fecha Fin:	10/05/2015
<b>Hito asociado</b>	Clasificadores entrenados con características escogidas.				
<b>Dependencias</b>	2				
<b>Observaciones</b>	La experimentación del modelo de clasificadores puede llevarse a cabo de manera paralela con la de los subconjuntos de características.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Experimentar con modelos de clasificadores.</li> <li>- Experimentar con subconjuntos de características.</li> <li>- Escoger el mejor modelo de clasificadores con su respectivo subconjunto de características.</li> </ul>				

Componente	Descripción				
<b>Código del paquete de trabajo</b>	4.1				
<b>Descripción del paquete de trabajo</b>	Implementación de las funciones de simplificación de textos.				
<b>Entregable(s)</b>	Funciones de simplificación de textos.				
<b>Criterios de aceptación del entregable(s)</b>	<ul style="list-style-type: none"> <li>- Las funciones deberán llevar a cabo las operaciones de simplificación escogidas en (1).</li> <li>- Las oraciones deben conservar su significado (a excepción del caso de eliminación de oración completa).</li> <li>- Las oraciones deben ser más sencillas según la métrica de Flesch-Szigrizt (aproximarse más a la línea base formada por la aplicación de esta métrica al corpus).</li> </ul>				
<b>Recursos</b>	Tesisista.				
<b>Duración estimada</b>	28 días	Fecha inicio:	11/05/2015	Fecha Fin:	07/06/2015
<b>Hito asociado</b>	Ninguno.				
<b>Dependencias</b>	3				
<b>Observaciones</b>	Ninguna.				
<b>Tareas</b>	<ul style="list-style-type: none"> <li>- Implementar las operaciones de simplificación.</li> <li>- Probar las operaciones de simplificación.</li> </ul>				

Componente	Descripción				
Código del paquete de trabajo	4.2				
Descripción del paquete de trabajo	Implementación del "front-end" de aplicación cliente-servidor para la simplificación de textos.				
Entregable(s)	Prototipo funcional de aplicación cliente-servidor para la simplificación de textos.				
Criterios de aceptación del entregable(s)	- El formulario de la aplicación debe ser simple y sencillo de usar.				
Recursos	Tesista.				
Duración estimada	7 días	Fecha inicio:	08/06/2015	Fecha Fin:	14/06/2015
Hito asociado	Herramienta de simplificación implementada y probada.				
Dependencias	4.1				
Observaciones	Los requisitos de la aplicación se encuentran en el catálogo de requerimiento. Los flujos de información, en las especificaciones de casos de uso.				
Tareas	<ul style="list-style-type: none"> <li>- Diseñar la interfaz gráfica.</li> <li>- Implementar las funcionalidades de la aplicación.</li> <li>- Desplegar en ambiente de pruebas.</li> </ul>				

Componente	Descripción				
Código del paquete de trabajo	4				
Descripción del paquete de trabajo	Implementación aplicación cliente-servidor para la simplificación de textos.				
Entregable(s)	Aplicación cliente-servidor para la simplificación de textos.				
Criterios de aceptación del entregable(s)	<ul style="list-style-type: none"> <li>- El formulario de la aplicación debe ser simple y sencillo de usar.</li> <li>- Comunicación con el "back-end" por medio de servicios web.</li> </ul>				
Recursos	Tesista, Asesor.				
Duración estimada	42 días	Fecha inicio:	11/05/2015	Fecha Fin:	21/06/2015
Hito asociado	Ninguno.				
Dependencias	4.2				
Observaciones	Ninguna.				
Tareas	<ul style="list-style-type: none"> <li>- Implementar las operaciones de simplificación (4.1).</li> <li>- Implementar servicios web.</li> <li>- Implementar la interfaz gráfica de la herramienta (4.2).</li> <li>- Desplegar en ambiente de producción.</li> </ul>				

## Anexo 5: Cronograma del Proyecto

EDT	Tarea	Duración	Comienzo	Fin	Pred.	Recurso
	<b>Implementación de Herramienta de Simplificación</b>	<b>203 días</b>	<b>01/12/2014</b>	<b>21/06/2015</b>		
<b>1</b>	<b>Anotación del corpus de textos</b>	<b>114 días</b>	<b>01/12/2014</b>	<b>24/03/2015</b>		
1.1	Obtener un corpus paralelo de textos en Español	62 días	01/12/2014	31/01/2015		Tesista,Asesor
<b>1.2</b>	<b>Herramienta de apoyo para la anotación del corpus</b>	<b>25 días</b>	<b>01/02/2015</b>	<b>25/02/2015</b>		
1.2.1	Realizar analisis y diseño de la herramienta	4 días	01/02/2015	08/02/2015	2	Tesista
1.2.2	Implementar la herramienta	10 días	07/02/2015	18/02/2015	4	Tesista
1.2.3	Probar la herramienta	5 días	19/02/2015	25/02/2015	5	Tesista
1.2.4	Herramienta implementada	1 día	23/02/2015	23/02/2015	6	Asesor, Tesista
1.3	Anotar manualmente el corpus	24 días	26/02/2015	22/03/2015	7	Tesista

1.4	Analizar las frecuencias de las operaciones de simplificación	1 día	23/03/2015	24/03/2015	8	Tesista
1.5	Corpus anotado y analizado	0 días	24/03/2015	24/03/2015	9	Tesista,Asesor
<b>2</b>	<b>Extractores de características</b>	<b>26 días</b>	<b>25/03/2015</b>	<b>19/04/2015</b>		
2.1	Definir los extractores que se van a implementar	3 días	25/03/2015	27/03/2015	9	Tesista,Asesor
2.2	Implementar los extractores	12 días	28/03/2015	08/04/2015	12	Tesista
2.3	Aplicar las pruebas unitarias automatizadas a los extractores	8 días	09/04/2015	16/04/2015	13	Tesista
2.4	Extractores implementados y probados	0 días	16/04/2015	16/04/2015	14	Tesista,Asesor
2.5	Aplicar los extractores a los textos del corpus	3 días	17/04/2015	19/04/2015	14	Tesista
<b>3</b>	<b>Modelo de clasificadores</b>	<b>21 días</b>	<b>20/04/2015</b>	<b>10/05/2015</b>		

3.1	Experimentar con modelos de clasificadores	21 días	20/04/2015	10/05/2015	16	Tesista[50%]
3.2	Experimentar con subconjuntos de características	21 días	20/04/2015	10/05/2015	18CC	Tesista[50%]
3.3	Clasificadores entrenados con características escogidas.	0 días	10/05/2015	10/05/2015	19	Tesista,Asesor
<b>4</b>	<b>Herramienta para simplificación</b>	<b>42 días</b>	<b>11/05/2015</b>	<b>21/06/2015</b>		
4.1	Implementar operaciones de simplificación	28 días	11/05/2015	07/06/2015	19	Tesista[90%]
4.2	Implementar interfaz web	7 días	08/06/2015	14/06/2015	22	Tesista[90%]
4.3	Implementar servicio web	4 días	15/06/2015	18/06/2015	23	Tesista[90%]
4.4	Herramienta implementada y probada	0 días	18/06/2015	18/06/2015	24	Tesista,Asesor
4.5	Desplegar herramienta en servidor web	42 días	11/05/2015	21/06/2015	22CC	Tesista[10%],Asesor

Anexo 6: Costeo del proyecto

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA

Entregable	Tarea	Recursos	Trabajo	Costo
Corpus original recopilado.	Obtener un corpus paralelo de textos en Español	Tesista, Asesor	992 horas	S/. 1,240.00
Herramienta de apoyo para la anotación.	Realizar análisis y diseño de la herramienta	Tesista	32 horas	S/. 32.00
	Implementar la herramienta	Tesista	80 horas	S/. 80.00
	Probar la herramienta	Tesista	40 horas	S/. 40.00
	Herramienta implementada	Asesor, Tesista	16 horas	S/. 20.00
Corpus anotado	Anotar manualmente el corpus	Tesista	192 horas	S/. 192.00
	Analizar las frecuencias de las operaciones de simplificación	Tesista	8 horas	S/. 8.00
Funciones extractoras	Definir los extractores que se van a implementar	Tesista, Asesor	48 horas	S/. 60.00
	Implementar los extractores	Tesista	96 horas	S/. 96.00
Pruebas automatizadas de extractores	Aplicar las pruebas unitarias automatizadas a los extractores	Tesista	64 horas	S/. 64.00
	Aplicar los extractores a los textos del corpus	Tesista	24 horas	S/. 24.00
	Experimentar con modelos de clasificadores	Tesista [50%]	84 horas	S/. 84.00
Modelo de clasificadores binarios en serie.	Experimentar con subconjuntos de características	Tesista [50%]	84 horas	S/. 84.00
	Implementar operaciones de simplificación	Tesista [90%]	201.6 horas	S/. 201.60
Back-end	Implementar servicio web	Tesista [90%]	28.8 horas	S/. 28.80
Interfaz web	Implementar interfaz web	Tesista [90%]	50.4 horas	S/. 50.40
Aplicación web	Desplegar herramienta en servidor web	Tesista [10%], Asesor	369.6 horas	S/. 537.60

## Anexo 7: Línea Base de Calidad del Proyecto

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA

Línea Base de Calidad					
Factor de calidad	Métrica	Objetivo de calidad	Frecuencia de medición	Frecuencia de reporte	Relevancia
Desempeño del proyecto	SPI (Schedule Performance Index)	SPI $\geq$ 0.95	Semanal	Quincenal, al asesor.	1
Desempeño del proyecto	CPI (Cost Performance Index)	SPI $\geq$ 0.95	Semanal	Quincenal, al asesor.	2

## Anexo 8: Guía de Anotación de Corpus

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA

Caso encontrado	Elim	Red	Sep	Cab	Ord	Ins	Mant
Oración eliminada	X						
<b>Oración no eliminada</b>							
Oración Separada			X				
Eliminación de [1..N-1] partes	X		X				
Inserción de una definición			X			X	
Reordenamiento de una parte			X		X		
Cambio de sujeto de una parte			X	X			
Cambio de voz de una parte			X	X			
Parafraseo de una o más partes		X	X				
No se hace nada más			X				
<b>Oración No Separada</b>							
Reordenamiento de una parte					X		
Cambio de sujeto de una parte				X			
Cambio de voz de una parte				X			
Inserción de una definición						X	
Hacer explícito un verbo reflexivo						X	
<b>Parafraseo</b>							
Dif. Longitud $\geq 10$		X					
Dif. Longitud $< 10$							X
No se hace nada							X

## Anexo 9: Pruebas de las funciones analíticas

Nombre del proyecto	Siglas del proyecto
Implementación de una herramienta de automatización para la simplificación de textos en Español	SIMPLIFICA

**Objetivo:** documentar las pruebas unitarias de funcionamiento de las funciones de extracción de características.

### Pruebas de funcionamiento

cuentaSustantivosPropios		
Oración	Esperado	Obtenido
Juan corrió a su casa.	1	1
Ana y Teresa quemaron su comida en el horno.	2	2
La verde cocina está nueva y limpia.	0	0
La calle de Juan era muy grande.	1	1
La calle de Ana y Sofia era muy grande.	2	2
La calle de Pablito fue limpiada.	1	1

cuentaPronombresPersonales		
Oración	Esperado	Obtenido
Yo corrí a mi casa.	1	1
Ana y Teresa quemaron su comida en el horno.	0	0
Ustedes salieron al parque conmigo.	2	2
Usted cree que a ella le gusta él.	4	4
Lo traje consigo.	2	2
Vosotros se quieren.	2	2

<b>cuentaPronombresDemostrativos</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Tal es mentira.	1	1
Aquellos son míos.	2	2
Ésa es de Pepe.	1	1
Eso me asusta.	2	2
Esto es feo.	1	1
Aquel borrador es verde.	0	0

<b>cuentaPronombresPosesivos</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
El nuestro es así.	1	1
Es el suyo.	1	1
Eso es mío.	2	2
Nuestro padre es el mejor.	0	0

<b>cuentaPronombresIndefinidos</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Algunos se dejan llevar.	2	2
Hablan bastantes personas en una habitación y ninguna en la otra.	2	2
Pocos saben que me gusta el limón.	2	2
Cualquiera sabe que la calle es grande.	1	1
Venden varios jugos.	0	0
La calle es gris.	0	0

**cuentaPronombresInterrogativos**

Oración	Esperado	Obtenido
¿Dónde está Juan?	1	1
¿Cuándo ocurrió el evento?	1	1
¿Quién y por qué hizo eso?	2	2
¿Soy?	0	0
¿Juan, adónde crees que vas?	1	1
Toda noticia debe indicar el porqué de las mismas.	0	0

**cuentaPronombresRelativosYExclamativos**

Oración	Esperado	Obtenido
Voy donde cortan el pelo.	1	1
La señora cuyos gatos se extraviaron.	2	2
Come como perro.	0	0
El cuando vaya.	1	1
Son ellos quienes te abandonaron y cuyas familias rompieron.	4	4
¡Qué!	1	1

**cuentaAdverbios**

Oración	Esperado	Obtenido
Juan corrió a su casa rápidamente.	1	1
Ana quemó su comida en el horno y tristemente la botó.	1	1
La verde cocina está siempre limpia.	1	1
La calle por donde corría Juan ahora es muy grande.	2	2
Mi mamá corre.	0	0
Lo supo antes que yo.	1	1
Mañana será un gran día.	1	1
Quizá venga solamente yo alrededor tuyo.	3	3
El cuadro fue vendido y Juana lloró.	1	1

<b>cuentaAdjetivos</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan corrió a su verde casa.	2	2
Ana quemó su rica comida en el horno y la botó.	2	2
La verde cocina está nueva y limpia.	3	3
La gris calle por donde corría Juan era muy grande.	2	2
Mi padre es un mecánico.	1	1
Juana es más alta que María.	1	1
Pedro es menos talentoso que Alex.	1	1
La casa es tan pequeña como un departamento.	1	1
La pizza me gusta.	0	0
Soy el primer puesto.	1	1
Son los primeros en llegar.	1	1
Somos la segunda compañía.	1	1
La decimocuarta potencia.	1	1
El undécimo y duodécimo puesto.	2	2

<b>cuentaPreposiciones</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan corrió a su casa rápidamente.	1	1
Ana quemó su comida en el horno y tristemente la botó.	1	1
Estuve con ella durante el sismo.	2	2
Jugamos versus tu equipo.	1	1
Jugamos contra su equipo.	1	1
Se paró ante él.	1	1
Lo compramos mediante Internet.	1	1

**cuentaDeterminantes**

Oración	Esperado	Obtenido
Aquel borrador me pertenece.	1	1
Aquellas cartucheras son mías y esa cartuchera es tuya.	2	2
Tus padres son médicos.	0	0
No toques mis cosas.	0	0
¿Cuántos platos has comido?	1	1
¿Qué sabor de pizza te gusta?	1	1
Ninguna chica quiere jugar contigo.	1	1
Toda torta se la das.	1	1
Hay varias cosas.	1	1
Otros gatos son de ella.	1	1

**cuentaIndicativos**

Oración	Esperado	Obtenido
Aquel borrador me pertenece.	1	1
La señora me vendió una limonada.	1	1
Me compré un videojuego.	1	1
Viajaré a París mañana.	1	1
Me gusta jugar pelota y me gusta montar patineta.	2	2
Soy de Lima	1	1

**cuentaSubjuntivo**

Oración	Esperado	Obtenido
Ojalá me venda su videojuego.	1	1
Ojalá te mejorases más rápido.	1	1
Me compré un videojuego.	0	0

<b>cuentaImperativos</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
¡Ciérrate la puerta!	1	1
!Traed el jugo de naranja!	1	1
Me compré un videojuego.	0	0
Trae tú el jugo y ciérrate la ventana.	2	2

<b>cuentaInfinitivos</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Me gusta jugar videojuegos y montar patineta.	2	2
!Traed el jugo de naranja!	0	0
Me compré un videojuego.	0	0
Mirar televisión.	1	1

<b>cuentaParticipio</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
La gata fue cuidada por su dueño.	1	1
El auto fue comprado por aquella señora.	1	1
Me compré un videojuego esta mañana.	0	0
Yo he adquirido un vaso de agua y he vendido una bottela.	2	2
Ella había sido estafada.	2	2

<b>cuentaGerundio</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Estaba jugando pelota.	1	1
He estado hablando por teléfono.	1	1
Me compré un videojuego.	0	0
Estaba mananejando cuando llamste.	1	1
Ella habia sido estafada.	0	0

<b>cuentaConjuncionesCoordinativas</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Estaba jugando pelota y sonó mi celular.	1	1
He estado hablando por teléfono pero he estado atento.	1	1
Me compré un videojuego, mas no gasté mucho.	1	1
No te hablaba a ti, sino a ella.	1	1
Fue una hermosa visita.	0	0

<b>cuentaConjuncionesSubordinativas</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
lamaste cuando estaba en el baño.	1	1
Hizo travesuras mientras dormías.	1	1
Crei qye es así porque sí.	2	2
No te hablaba a ti, sino a ella.	0	0
Del texto se deduce, entonces, que dos más dos es cuatro.	1	1

<b>cuentaFrasesVerbales</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan corrió a su casa.	1	1
Alan fue al cine y Ana jugó con sus muñecas.	2	2
Se quitó la chompa porque tenía calor.	2	2
La casa fue vendida al mejor postor.	1	1
El cuadro fue vendido y Juana lloró.	2	2

**cuentaFrasesNominales**

Oración	Esperado	Obtenido
Juan corrió a su casa.	2	2
Alan fue al cine y Ana jugó con sus muñecas.	4	4
Se quitó la chompa porque tenía calor.	2	2
La casa fue vendida al mejor postor.	2	2
El cuadro fue vendido y Juana lloró.	2	2

**calculaTamañoPromedioFrasesVerbales (caracteres)**

Oración	Esperado	Obtenido
Juan corrió a su casa.	6	6
Alan fue al cine y Ana jugó con sus muñecas.	4	4
Se quitó la chompa porque tenía calor.	5	5
La casa fue vendida al mejor postor.	10	10
El cuadro fue vendido y Juana lloró.	8	8

**calculaTamañoPromedioFrasesNominales (caracteres)**

Oración	Esperado	Obtenido
Juan corrió a su casa.	4	4
Alan fue al cine y Ana jugó con sus muñecas.	5	5
Se quitó la chompa porque tenía calor.	6	6
La casa fue vendida al mejor postor.	5	5
El cuadro fue vendido y Juana lloró.	6	6

**cuentaFrasesPreposicionales**

Oración	Esperado	Obtenido
Juan corrió a su casa.	1	1
Alan fue al cine y Ana jugó con sus muñecas.	2	2
Se quitó la chompa porque tenía calor.	0	0
La casa fue vendida al mejor postor.	1	1
El cuadro fue vendido y Juana lloró.	0	0

**calculaTamañoPromedioFrasesPreposicionales (caracteres)**

Oración	Esperado	Obtenido
Juan corrió a su casa.	7	7
Alan fue al cine y Ana jugó con sus muñecas.	10	10
Se quitó la chompa porque tenía calor.	0	0
La casa fue vendida al mejor postor.	14	14
El cuadro fue vendido y Juana lloró.	0	0

**vozPasiva**

Oración	Esperado	Obtenido
Juan corrió a su casa.	0	0
Alan fue al cine y Ana jugó con sus muñecas.	0	0
Se quitó la chompa porque tenía calor.	0	0
La casa fue vendida al mejor postor.	1	1
El cuadro fue vendido y Juana lloró.	1	1

**cuentaFrasesAdverbiales**

Oración	Esperado	Obtenido
Juan corrió a su casa.	0	0
Alan fue al cine y Ana jugó con sus muñecas muy emocionada.	1	1
Se quitó la chompa porque tenía mucho calor.	0	0
La casa fue vendida al mejor postor.	0	0
El cuadro fue vendido y Juana lloró.	0	0

**cuentaClausulasRelativas**

Oración	Esperado	Obtenido
La revista que te presté está sobre la mesa.	1	1
Las candidatas por quienes voté ghanaron las elecciones.	1	1
La chica que me gusta se llama Stephanie.	1	1
El hombre cuyo nombre no recuerdo.	1	1
Ela mesa de billar es verde.	0	0

**cuentaClausulasCoordinadas**

Oración	Esperado	Obtenido
Juan fue al cine y Ana regresó a su casa.	2	2
El gato subió al techo y Pedro lo estaba esperando.	2	2
Juan y Pedro jugaron a las escondidas.	0	0
Ana y María fueron al cine.	0	0
Juan y María jugaron cartas y Pedro y Ana vieron televisión.	2	2

**cuentaClausulasSubordinadas**

Oración	Esperado	Obtenido
Mi sueño es que tenga buena nota.	1	1
Iré cuando me digas.	1	1
Juan y Ana jugaron cartas-	0	0
Cuándo no es importante.	0	0
La televisión informa de que no hay huelga	1	1

**esOracionCompuesta**

Oración	Esperado	Obtenido
Juan fue al cine y Ana regresó a su casa.	1	1
El viejo barco de Luis zapó por la mañana.	0	0
Fue al médico cuando me sentí mal.	1	1
Ana jugó con muñacas y Luisa fue al médico cuando estaba mal.	1	1

**calcuProfundidad**

Oración	Esperado	Obtenido
Juan fue al cine y Ana regresó a su casa.	5	5
El viejo barco de Luis zapó por la mañana.	4	4

**sujetoPrimero**

Oración	Esperado	Obtenido
Juan fue al cine y Ana regresó a su casa.	1	1
El viejo barco de Luis zapó por la mañana.	1	1
Juega videojuegos.	1	1
Ana jugó con muñacas y Luisa fue al médico cuando estaba mal.	1	1
Anoche llovió mucho.	0	0
Juguete.	0	0
Se murió el perro.	0	0

<b>cuentaCaracteres</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan corrió a su casa.	18	18
Ana quemó su comida en el horno.	26	26
La verde cocina está nueva y limpia.	30	30
La calle de Juan era muy grande.	26	26

<b>cuentaPalabras</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan corrió a su casa.	5	5
Ana quemó su comida en el horno.	7	7
La verde cocina está nueva y limpia.	7	7
¿Hola?	1	1
La calle de Juan era muy grande.	6	6

<b>calculaTamañoPromedioPalabrasEnOracion</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan corrió a su casa.	4	4
Ana quemó su comida en el horno.	4	4
¿Hola?	5	5

<b>cuentaSignosDePuntuacion</b>		
<b>Oración</b>	<b>Esperado</b>	<b>Obtenido</b>
Juan, ve a tu casa.	2	2
¿Qué haces, Pedro?.	4	4
¿Hola?.	3	3

cuentaExpresionesNumericas		
Oración	Esperado	Obtenido
Juan tomó 3 gaseosas.	1	1
Tengo 220 soles en el banco.	1	1
Controla el 80% de las acciones.	1	1
Compra 1 litro de leche.	1	1
Juan corre todo el día.	0	0

**Herramienta usada:** JUnit 4.0 en Netbeans 8.0.

**Conclusión:** las funciones de extracción actúan de acuerdo a lo esperado.



## Anexo 10: ejemplos de simplificación de oraciones

### Oración compuesta coordinada copulativa

**Oración original:** Juan vive al final de la calle y va al cine todos los días.

**Oración simplificada:** Juan vive al final de la calle. Juan va al cine todos los días.

### Oración compuesta coordinada adversativa exclusiva

**Oración original:** No he comprado un coche sino una bicicleta.

**Oración simplificada:** No he comprado un coche. He comprado una bicicleta.

### Oración compuesta coordinada disyuntiva

**Oración original:** Tienes que salir o no llegarás a tiempo.

**Oración simplificada:** Tienes que salir. No llegarás a tiempo.

### Oración compuesta coordinada distributiva

**Oración original:** Ya baila mucho, ya bebe mucho.

**Oración simplificada:** Baila mucho o bebe mucho.

### Oración compuesta coordinada consecutiva

**Oración original:** Juan salió a la calle porque sintió un temblor.

**Oración simplificada:** Juan salió a la calle. Juan sintió un temblor.

### Oración compuesta subordinada adjetiva explicativa

**Oración original:** Los jugadores, que estaban cansados, no jugaron.

**Oración simplificada:** Los jugadores estaban cansados. Los jugadores no jugaron.

### Construcciones de participio

**Oración original:** Se les regaló un libro editado por el ayuntamiento.

**Oración simplificada:** Se les regaló un libro. El libro es editado por el ayuntamiento.

### Construcciones con elipsis en objeto directo

**Oración original:** Se le acusa de robar un caramelo, hecho de chocolate.

**Oración simplificada:** Se les acusa de robar un caramelo. El caramelo está hecho de chocolate.

### Construcciones con elipsis en objeto indirecto

**Oración original:** Se le acusa de maltratar a John Doe, un soldado.

**Oración simplificada:** Se les acusa de maltratar a John Doe. John Doe es un soldado.

### Cambio de estructura de reporte

**Oración original:** “Vamos a jugar”, dijo un niño a su amigo.

**Oración simplificada:** Un niño dijo: “Vamos a jugar”.