**Universidad Nacional de Mar del Plata**

**Facultad de Ingeniería**

**Departamento de Informática**

**PROTOCOLO DE PROYECTO FINAL – INGENIERÍA EN INFORMÁTICA**

**NOMBRE DEL PROYECTO**

|  |
| --- |
| NLP APLICADO A ANÁLISIS DE TEXTO |

**SÍNTESIS DEL PROYECTO**

|  |
| --- |
| El proyecto se centrará en el desarrollo de una aplicación para agilizar las investigaciones judiciales en aquellos casos que se requiera del análisis de grandes volúmenes de información en formato de texto.  Para tal fin, se plantea el uso de las herramientas brindadas por la inteligencia computacional, puntualmente aquellas orientadas al procesamiento del lenguaje natural. |

**EQUIPO DE PROYECTO**

Integrantes alumnos

|  |  |  |
| --- | --- | --- |
| **Nombre** | **DNI** | **Carrera** |
| Alias, Gerardo Ismael | 33185304 | Ingeniería en Informática |
| Cassanelli, Rodrigo Nicolás | 38829062 | Ingeniería en Informática |

Director

|  |
| --- |
| Ing. Ana Haydée Di Iorio |

Co-Director

|  |
| --- |
| **I**ng. Santiago Trigo – InFo-Lab |

Referente Funcional

|  |
| --- |
| Ing. Bruno Constanzo - InFo-Lab bconstanzo@ufasta.edu.ar  Ing. Sebastián Lasia - [InFo-Labslasia@ufasta.edu.ar](mailto:InFo-Labslasia@ufasta.edu.ar)  4444 4444 4444 4444  [Rodrigo.ncassaasdasdasdasdasdasdasdasdasdadsasdnelli@gmail.com](mailto:Rodrigo.ncassanelli@gmail.com)  [rodrigo@gmail.com](mailto:rodrigo@gmail.com)  [cassanelli@hotmail.com](mailto:cassanelli@hotmail.com)  [roodriqoo\_03@gmail.com](mailto:roodriqoo_03@gmail.com)  [rodrigo.cassanelli@outlook.com.ar](mailto:rodrigo.cassanelli@outlook.com.ar)  [hola@hotmail.com](mailto:hola@hotmail.com)  [aaaaaa@gmail.com](mailto:aaaaaa@gmail.com)  [asdadasda@gmail.com](mailto:asdadasda@gmail.com)  [asdafjalsjfljaf@gmail.com](mailto:asdafjalsjfljaf@gmail.com)  ddkdkdkd@gmail.com  www.gmail.com |
|  |

**PROBLEMA A RESOLVER**

|  |
| --- |
| La masificación de las redes sociales, aplicaciones de mensajería y correo electrónico, entre otros, hace que hoy en día sea moneda corriente que una investigación criminal implique analizar grandes cantidades de comunicaciones realizadas a través de texto. La actualidad en Argentina en general, y la provincia de Buenos Aires en particular, nos muestra que estas tareas son realizadas por un investigador, generalmente perito informático, en forma manual.  Esta realidad, hace que, cada vez las investigaciones insuman más tiempo, cuando la cantidad de mensajes referidos a la temática a investigar, son un subconjunto del total.  Ya de por sí, hay una dificultad propia en las tareas de extracción de información, a la que se suma, el procesamiento y análisis de esa información. |

**OBJETIVO DEL PROYECTO**

|  |
| --- |
| El objetivo del trabajo es el desarrollo de una aplicación capaz de ofrecer a los usuarios – peritos informáticos y/o investigadores judiciales - un medio para realizar un análisis automatizado del texto extraído de diversos soportes. Dicha aplicación tendrá la finalidad de preseleccionar del total de los mensajes extraídos, un subconjunto de interés a analizar, focalizando la atención del investigador en aquellos fragmentos del texto que tienen mayor probabilidad de contener la información buscada – de acuerdo a la temática de la investigación. |

**PRODUCTO ESPERADO**

|  |
| --- |
| Se espera obtener una aplicación con arquitectura cliente servidor, que permita a los usuarios ingresar texto, ya sea en formato plano o a través de archivos mediante una interfaz web.  El usuario, podrá seleccionar entre determinados tópicos y/o sub-tópicos – temas de interés que tiene la capacidad de detectar la aplicación - a buscar dentro de los textos presentados a la aplicación.  El procesamiento de la información será realizado en un servidor que dispondrá de determinados modelos para reconocimiento de lenguaje natural.  Luego de procesada la información, el servidor devolverá los resultados y será la interfaz de usuario la que proveerá de distintas herramientas de representación para facilitar la interpretación de los resultados obtenidos.  Además, la aplicación proveerá de herramientas administrativas para, entre otras cosas, generar nuevos modelos, entrenar los modelos ya existentes y un sistema de retroalimentación para que los administradores puedan ajustar dichos modelos en función del feedback provisto por los usuarios.  Finalmente, se plantea generar en el marco de este proyecto, algún modelo particular para la investigación temática (de consumo o tráfico de drogas y narcóticos, por ejemplo). |

**DEMANDANTE DEL PROYECTO [[1]](#footnote-1)**

|  |
| --- |
| InFo-Lab – Laboratorio de Investigación y Desarrollo de Tecnología en Informática Forense. Procuración General de la Suprema Corte de Justicia de la Provincia de Buenos Aires, Universidad FASTA y Municipalidad de General Pueyrredón. |

**CARACTERÍSTICAS DE INNOVACIÓN LOCAL DEL PROYECTO [[2]](#footnote-2)**

|  |
| --- |
| El análisis de los textos extraídos de diferentes soportes, principalmente de Smartphones, es una actividad cotidiana de cualquier investigación judicial. Es importante destacar que en nuestro país dicho análisis es realizado de forma manual, es decir, un investigador judicial debe invertir horas en leer estos textos/mensajes. Generalmente, la mayor parte de los textos leídos son irrelevantes para la investigación.  La aplicación propuesta plantea optimizar el uso del tiempo de los investigadores preseleccionando la información relevante vinculada a la investigación judicial.  Si bien a priori parece una tarea tan trivial como una simple búsqueda de texto, es importante destacar que el modo en que se utiliza el lenguaje natural en los diferentes medios informáticos dista mucho de lo sintáctica u ortográficamente correcto. La utilización de redes neuronales permite que haya un comportamiento flexible y adaptativo de forma tal que no sólo preselecciona un texto a partir de reglas preestablecidas sino también a partir de la inferencia resultante del entrenamiento de los modelos.  Contemplar todas las posibilidades de escritura de un solo término en español es un trabajo arduo y tedioso, sin contar las deformaciones al lenguaje o palabras que hacen alusión a otras. Es por ello que la aplicación dispondrá de herramientas para la generación de modelos que faciliten no solo agrupar términos, sino que también permitan deformarlos, conjugar verbos, etc. La idea detrás de todo esto es lograr que generar modelos se reduzca simplemente a detectar términos claves.  La innovación local radica en la utilización de estas tecnologías en las investigaciones penales llevadas a cabo en la provincia de Buenos Aires, ya que actualmente no se utiliza ningún recurso automatizado en las mismas. |

**PLAN DE TRABAJO**

**Fecha de Inicio prevista:** 01/09/2018

**Fecha de Finalización prevista:** 30/03/2019

|  |
| --- |
|  |

**FODA DEL PROYECTO**

|  |
| --- |
| **Fortalezas:**  ● Se cuenta con conocimiento del tema ya que se ha participado en la implementación de un prototipo durante la práctica profesional supervisada desarrollada en el Infolab. Se conocen opciones, herramientas y se tiene un panorama claro de las necesidades del proyecto.  ● Los estudiantes que llevarán a cabo el desarrollo están capacitados en el uso de todas las herramientas, lenguajes y conceptos requeridos para llevar a cabo el proyecto.  **Oportunidades:**  ● Se han tenido reuniones con profesionales en la materia de aplicación del proyecto. En todos los casos se han obtenido respuestas positivas y se ha resaltado el impacto positivo que tendría contar con una herramienta de estas características.  ● La necesidad de este tipo de herramientas ha crecido sustancialmente en los últimos años, con la expansión de diversas aplicaciones siendo una de las más emblemáticas WhatsApp.  **Debilidades:**  ● Los dos estudiantes aún se encuentran cursando materias y, a su vez, uno de ellos trabaja a tiempo completo. Este panorama hace que exista riesgo de que algunas actividades se atrasen.  ● Se cuenta con recursos limitados de hardware a la hora de realizar tareas con elevado requerimiento de poder de cómputo tales como el entrenamiento de las redes neuronales.  **Amenazas:**  ● El proyecto no utiliza una red neuronal propietaria, ni desarrollará un modelo base de lenguaje español propio. Cualquier cambio en las condiciones de uso, discontinuidad en el soporte o mejora de la misma, así mismo como cualquier bug o inconveniente que pueda derivar en la pérdida de acceso a la API puede condicionar en gran medida el desarrollo del proyecto. |

**DERECHOS SOBRE LO PRODUCIDO DEL PROYECTO**

*Los derechos de lo producido en el marco del presente Proyecto son compartidos por la Universidad Nacional de Mar del Plata (50%) y los estudiantes del grupo (50%).*

*Los titulares de tales derechos aceptan ceder la explotación total de los mismos sin costo alguno en favor del InFo-Lab (Procuración General de la Suprema Corte de Justicia de la Provincia de Buenos Aires, Universidad FASTA y Municipalidad de General Pueyrredón). El InFo-Lab se compromete a mencionar la autoría de los alumnos sobre el software producido en el mismo software y en todas las presentaciones y publicaciones que se hagan del mismo.*

**COMPROMISO DE LOS ESTUDIANTES**

*Por la presente, nos comprometemos a desarrollar el Proyecto Final descripto ut supra, comunicando periódicamente las actividades a los docentes supervisores, asistiendo a los seminarios que la Cátedra indique, conduciéndome éticamente con la actitud que corresponde al ámbito universitario.*

*Firmas de los estudiantes: ………………………………………………………*

*………………………………………………………*

*………………………………………………………*

**COMPROMISO DE LOS DIRECTORES**

*Por la presente, me comprometo a dirigir técnicamente la ejecución del presente proyecto final, proveyendo el asesoramiento necesario y en todo de acuerdo con las normas de la Facultad y carrera.*

*Firma Director: ………………………………………………………*

*Firma Co-Director: ………………………………………………………*

*Firma Referente Funcional: ………………………………………………………*

**RECIBIDO DEPARTAMENTO DE INFORMÁTICA**

*Fecha:*

*Firma por Departamento: ………………………………………………………*

**DICTAMEN DE LA CÁTEDRA**

*Fecha:*

*Recomendación:*

* *Autorizar el desarrollo del proyecto tal como fue presentado por el grupo*
* *Autorizar el desarrollo del proyecto con las sugerencias realizadas (pueden no ser atendidas por el grupo)*
* *Autorizar el desarrollo del proyecto siempre que el grupo incorpore las modificaciones solicitadas*
* *No autorizar el desarrollo del proyecto, en base a la fundamentación realizada*

*Sugerencias*

|  |
| --- |
|  |

*Modificaciones solicitadas*

|  |
| --- |
|  |

*Firma Profesores: ………………………………… …………………………………*

*Firma por Departamento: ………………………………………………………*

**Aceptación del Dictamen**

*Por la presente, aceptamos el dictamen de la cátedra interviniente y nos comprometemos a desarrollar el proyecto conforme lo estipulado por la misma.*

*Firmas de los estudiantes: -----------------------------------------------------------------*

*-----------------------------------------------------------------*

*-----------------------------------------------------------------*

1. Se entiende por demandante a la empresa o institución que solicita el desarrollo de una solución tecnológica y que se beneficia de tal desarrollo incorporando lo producido en el proyecto [↑](#footnote-ref-1)
2. Se deben consignar las características que hacen del proyecto o producto del mismo algo innovador, en tanto genera nuevos conocimientos y/o soluciones tecnológicas que incorporen originalidad o novedad cognitiva como un rasgo característico a la luz de las condiciones locales [↑](#footnote-ref-2)