

Algoritmo de Vinculación de Relatos de Denuncias Policiales basado en NER

Mauro Daniel Álvarez^[0009-0005-5288-9651]¹, Leandro Antonelli^[0000-0003-1388-0337]^{2 3}

¹ Facultad de Informática UNLP, La Plata, Bs As, Argentina

² LIFIA – Facultad de Informática, UNLP

³ CAETI – Facultad de Tecnología Informática – Universidad Abierta Interamericana

{mauro.alvarez}@info.unlp.edu.ar
{leandro.antonelli}@lifia.info.unlp.edu.ar

Abstract. Las denuncias policiales, las cuales posteriormente se materializan en casos judiciales, constan de información entre la que se puede destacar el relato narrado por el denunciante. Este relato es una descripción de que ocurrió desde el punto de vista del afectado y constituye el punto de inicio de un caso penal y su futura investigación. Un gran número de estos crímenes son perpetrados por autores desconocidos, es decir, no se puede determinar la autoría del mismo. Debido a esta situación los casos permanecen en estado archivado hasta que nuevos indicios arriben que permitan la continuación de la investigación. A través del uso de procesamiento de lenguaje natural y NER, se propone hacer uso del texto no estructurado provisto por el relato de una denuncia y extraer entidades de interés de manera de poder vincularlas con otros relatos de denuncias existentes con el objetivo de detectar similitudes y generar una representación gráfica de esta situación para proveer asistencia a los investigadores.

Keywords: NLP, NER, relatos denuncias, justicia penal, similitud.

1 Introducción

Los relatos de las denuncias policiales están conformados por texto no estructurado. La información presente en ellos puede ser categorizada, por ejemplo, nombres, lugares, horarios, tipos de delitos entre otros.

Actualmente se cuenta con una amplia base de relatos de denuncias, la misma es denominada corpus. Si bien los relatos constan de texto no estructurado, el contenido de los mismos es de carácter heterogéneo, es decir, estos pueden distar mucho entre sí. Debido a esta situación es preciso reducir la cantidad de relatos a sólo los más similares al relato a analizar.

Este artículo presenta un algoritmo empleando dos ramas de la inteligencia artificial, por un lado, el procesamiento de lenguaje natural y por el otro el reconocimiento de entidades nombradas, con el fin de identificar relatos de denuncias similares a un relato de interés denominado relato pivote.

El algoritmo está constituido por dos fases. La primera fase consiste en la reducción del número de relatos de la base existente a solo aquellos relatos que dispongan de algún grado de similitud con el relato a analizar o pivote.

La segunda fase consiste en la detección y extracción de entidades nombradas de los relatos seleccionados en la fase previa y el relato pivote. A partir de este momento se busca comparar aquellos relatos de denuncias que tengan entidades en común con el relato pivote. Esto permitirá detectar similitudes más profundas y de estilo más semántico en los relatos.

El resto del artículo se organiza de la siguiente manera. La sección dos describe el marco teórico necesario para entender la propuesta. La sección tres describe la propuesta. La sección cuatro detalla la evaluación preliminar realizada. La sección cinco enumera trabajos relacionados. Finalmente, se presentan las conclusiones.

2 Marco teórico

Tanto el procesamiento de lenguaje natural como el reconocimiento de entidades nombradas son ramas pertenecientes a la inteligencia artificial. La primera permite dotar a las máquinas de entendimiento acerca del lenguaje natural, es decir, aquel empleado por los humanos para comunicarse y brindarles la capacidad de poder generar texto en lenguaje natural apto para la comprensión humana.

En el alcance de este trabajo, el procesamiento del lenguaje natural se limita a reducir los datos ruidosos para luego realizar comparaciones sintácticas de forma más eficiente, y de esa manera reducir la cantidad de relatos del corpus.

Para incrementar la eficiencia del procesamiento del lenguaje natural abreviado NLP, es preciso realizar un preprocesamiento de los datos denominado estandarización o normalización [1]. El objetivo de este proceso es reducir el ruido en el texto no estructurado, es decir, remover toda simbología que no aporte significado al texto. Se pueden enumerar ejemplos como los símbolos de puntuación, las mayúsculas, espacios en blanco entre otros.

Este proceso incluye además la extracción de lemas [2]. Los lemas revisten importancia debido a que la utilización de verbos conjugados en el texto de los relatos es frecuente y puede involucrar distintos tiempos verbales. Ejemplos de un mismo verbo en un texto pueden ser extraídos como lemas de manera de ser iguales. El uso de verbos conjugados implica que un verbo conjugado y su lema son dos palabras totalmente independientes sin relación alguna.

Los dos procesos anteriormente enumerados son de vital importancia para incrementar la eficiencia del índice de similitud de Jaccard el cual será detallado a continuación.

2.1 Índice de similitud de Jaccard

El método de Jaccard está definido como la intersección de los conjuntos dividida por la unión de los mismos [3]. A continuación, se puede apreciar la fórmula.

$$A \cap B / A \cup B$$

Este método realiza una comparación puramente sintáctica, por lo que es de vital importancia realizar el proceso de estandarización enunciado anteriormente.

2.2 Reconocimiento de entidades nombradas

La otra rama corresponde a NER, abreviado de Named Entity Recognition [4]. El objetivo es obtener las entidades de interés en el relato. Dichas entidades corresponden al dominio estudiado, es decir, el de la justicia penal.

Al tratarse de un modelo de aprendizaje automático, es necesario proveer un modelo para que el mismo sea entrenado apropiadamente al contexto a utilizar, de otra manera las entidades que extraería carecerían de utilidad para el propósito del algoritmo.

Las entidades extraídas permiten obtener su texto y su categoría. Esto reviste utilidad ya que se puede determinar que entidades figuran con la misma categoría a lo largo de los casos y de esa manera es posible la comparación entre casos mediante la aplicación del índice de similitud de Jaccard.

En resumen, se podrían agrupar aquellos textos no estructurados donde participen las entidades extraídas. Por ejemplo, sería posible agrupar aquellos relatos donde figuren nombres, ubicaciones, tipos de delitos que figuran en el relato pivote, de manera de lograr establecer algún tipo de vinculación entre relatos. Esta idea se ve reforzada por la teoría del patrón delictivo [5]. La generación de estos enlaces entre relato pivote y los relatos relacionados del corpus permitirían detectar patrones de crímenes para individualizar personas y/o grupos de personas que cometen delitos basándose en las entidades extraídas.

3 Contribución

En esta sección se procede a detallar el algoritmo de dos fases. En la siguiente subsección se define un ejemplo de texto no estructurado que corresponde a un relato de denuncia utilizado como principal fuente de entrada del algoritmo. Las demás subsecciones detallan las dos fases del algoritmo y el formato de datos utilizado.

3.1 Ejemplo de relato de denuncia

Un relato de una denuncia consiste en una descripción narrada por parte de la víctima o un tercero afectado directa o indirectamente en el hecho denunciado.

En la figura 1 se muestra un ejemplo de relato de denuncia¹.

¹ El relato de denuncia presentado como ejemplo dispone de datos modificados que no se corresponden con los reales debido a que se tratan de datos sensibles.

```

Caso 120927
Ubicación del hecho: Trelew, Rivadavia Nro 0 - interseccion con calle inmigrantes.
Informe superioridad policial, inicio actuaciones preventionales, en virtud de que en fecha hora lugar mención, denunciante
deja su vehiculo marca volkswagen suran con dominio LLA111,
constata que autor/es ignorado/s previo a dañar cristal d ela puerta trasera
lado izquierdo sustrajo/eron del interior elementos que se detallana a continuación.

Elementos denunciados
Remera talla L color roja.
    
```

Fig. 1. Ejemplo de relato de denuncia

La denuncia está conformada por texto no estructurado en idioma español. Como se puede ver este texto contiene ruido que debe ser eliminado o minimizado ya que atenta contra la precisión de la comparación sintáctica que se lleva a cabo en la fase uno del algoritmo.

Para solventar esta situación, se emplean métodos de estandarización del texto. Se comienza con la conversión del texto en minúsculas, luego se procede con la remoción de las palabras que no aportan significado² y finalmente se detectan los lemas de los verbos conjugados.

En la figura 2 muestra el resultante del procedimiento enunciado anteriormente.

```

['calle', 'superioridad', 'inicio', 'actuación', 'virtud', 'fecha', 'hora', 'lugar', 'mención', 'dejar', 'vehiculo', 'constatar',
'autor', 'ignorar', 's', 'dañar', 'cristal', 'puerta', 'sustraer', 'eron', 'elemento', 'detallana', 'talle', 'color']
    
```

Fig. 2. Palabras o Tokens luego de la estandarización

A partir de este momento, se dispone de conjuntos de palabras o tokens por cada uno de los relatos de la base de relatos, así como también un conjunto que representa al relato pivote.

3.2 Estrategia general

En la figura 3 Se puede apreciar el esquema con la estrategia propuesta.

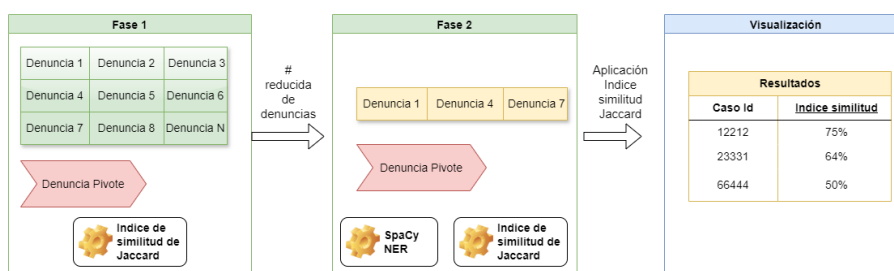


Fig. 3. Estrategia general propuesta

Como se indicó al inicio de la presente sección, el algoritmo consta de dos fases. La primera fase tiene como entrada un relato de denuncia denominado pivote. Esta

² Conocidas como stopwords.

denuncia es la que se intenta resolver. El conjunto de relatos de denuncias existentes constituye el conocimiento que dispone el algoritmo y serán utilizadas para determinar cuáles de las denuncias existentes son similares a la denuncia pivote.

Este proceso arrojará un número reducido de denuncias que serán ingresadas como entrada a la fase dos del algoritmo. En esta etapa se obtienen las entidades nombradas y se vuelve a aplicar el índice de similitud de Jaccard.

Este último refinamiento genera un arreglo con información que puede ser detallada al usuario en forma gráfica o textual. Actualmente se imprime en pantalla una tabla con el índice de similitud con respecto al relato pivote.

En las siguientes subsecciones se detallan las fases del algoritmo.

3.3 Fase uno del algoritmo

La fase uno del algoritmo tiene como precondition la ejecución del proceso anterior correspondiente a la estandarización del texto. Disponiendo de un conjunto de palabras libres de ruido, verbos conjugados, mayúsculas y demás características propias de un relato narrado, se puede proceder a comparar la similitud de los relatos.

Se emplea un índice de selección superior al 25% de similitud entre cada caso y el pivote, aquellos que arrojen un índice de similitud inferior son descartados.

En esta etapa se genera un objeto por cada caso con la estructura que se detalla en la figura 4.

```
Numero caso --> [ 126926 ]
Relato --> Caso 139929 ubicación del hecho: Trelew, Carrasco y urquiza Informe s.s superioridad policial, inicio actuaciones preventivas, fecha, hora, lugar, mención denunciante deja estacionad
o vehículo marca Renault, modelo Sandero, dominio WA-943, constatando que autor/es ignorado/s dañaron cristal trasero derecho ingreso/arón y sustrajo/eron elemento denunciado. Toma intervención p
ersonal de DFI. Elementos denunciados telefono celular, motorola E11, con funda cuero negro, correspondiente al 28012131231.
Jaccard Índex Fase 1 --> 0.7391304347825890
NER --> [ ]
```

Fig. 4. Estructura resultante de la fase uno.

El objeto generado contiene propiedades como número de caso, el texto sin procesar, el resultado del índice de similitud de Jaccard con respecto al relato pivote, el texto procesado luego de la estandarización y una colección vacía que será utilizada en la fase dos del algoritmo.

Esta estructura es repetida en todos los casos que cumplan el criterio de selección especificado, de manera que se dispone de una colección de esos objetos dando fin a la fase uno del algoritmo.

3.4 Fase dos del algoritmo

A partir de este momento se emplea el módulo NER provisto por la librería SpaCy [6].

Utilizando la colección de objetos generada en la fase previa, se extrae de cada uno de los objetos, las entidades nombradas del texto que fue estandarizado.

Al igual que la estandarización es necesaria para la fase uno, la creación de un modelo para el dominio es requerida en la fase dos.

No disponer de un modelo del dominio de interés, en este caso, el dominio judicial penal, haría que las entidades extraídas no sean de utilidad para el propósito del algoritmo. De manera que es necesario entrenar el modelo con el dominio apropiado. Por

ejemplo, se contemplan artículos del Código Penal Argentino [7], jerga del dominio, pertenencia de calles a barrios, individuos que hayan cometido crímenes, etc.

Actualmente la librería SpaCy dispone de un módulo NER dentro de su pipeline³. El mismo puede ser configurado con modelos ya entrenados, pero los mismos tienden a ser genéricos y no ajustarse apropiadamente al problema a resolver.

Por tanto, se propone el entrenamiento de un modelo NER adecuado al entorno judicial penal.

Este dominio incluye conocimiento sobre el Código Penal Argentino, nombres de calles que pertenecen a un barrio, así como nombres o alias según estén disponibles de posibles sospechosos de crímenes. En la tabla 1 se visualiza un ejemplo de cómo podrían ser los datos a entrenar.

Tabla 1. Plantilla de entrenamiento delitos

Palabra	NER	Información adicional
Los individuos Robaron el depósito	Artículo 164	Robo simple.
Los autores ignorados sustrajeron sus pertenencias	Artículo 164	Robo simple.
Hurtar	Artículo 162	Hurto
Dañar	Artículo 183	Vandalismo.

Como se puede apreciar, se trata de vincular determinadas acciones con artículos del código penal. Por ejemplo, el verbo robar o sustraer en sus distintos tiempos verbales podría vincularse con el artículo 164⁴ del código penal que corresponde a robo.

Esta misma idea se puede expandir a las calles que pertenecen a un barrio, o individuos relacionados por sus alias.

Para proceder al entrenamiento del modelo es preciso realizar anotaciones en los relatos. La tarea de anotación es una de las que mayor tiempo consumen debido a su naturaleza manual. Cada relato debe ser analizado en su contexto y anotado apropiadamente.

Una vez creado el corpus de datos, es necesario dividirlo en conjuntos de entrenamiento y validación [8].

Para realizar el entrenamiento del modelo NER, inicialmente se procedió a trabajar definiendo anotaciones manualmente. Dado que se trata de una tarea que lleva mucho tiempo, se reemplazó la definición manual de anotaciones con la herramienta denominada NER Text Annotator [9].

Con el modelo entrenado en el dominio de la justicia penal, se procede a identificar las entidades nombradas de cada uno de los relatos que componen al subconjunto generado por la fase uno del algoritmo.

³ Pipeline o tubería, indica una serie de tareas llevadas a cabo en orden. El módulo NER es solo un componente de dicho pipeline.

⁴ Específicamente Título IV, Delitos Contra la Propiedad, Capítulo II, Robo.

En la figura 5 se puede apreciar un ejemplo de detección de entidades nombradas en un relato cuyos datos reales fueron eliminados, pero que cuenta con la estructura real.

Caso 120927
 Ubicación del hecho: Trelew, Rivadavia Nro 0 Barrio Padre Juan - intersección con calle inmigrantes. Informe superioridad policial, en virtud de que en fecha hora lugar mención,
 denunciante deja su vehículo Artículo 183 marca volkswagen suran Vehículo con dominio LLA111, Patente constata que autor/es ignorado/s Autores ignorados previo a dañar
 cristal d ela puerta trasera lado izquierdo Artículo 183 sustrajo/eron del interior elementos Artículo 164 que se detallana a continuación.
 Elementos denunciados
 Remera talle L color roja.

Fig. 5. Obtención de entidades nombradas de un relato

En este relato se detectó un vehículo, su patente, el barrio donde se produjo el hecho y los artículos involucrados en el delito, también se puede saber que fue cometido por individuos no identificados. En este relato puntual la mayoría de los etiquetados fueron correctos, sin embargo, puede generarse un etiquetado erróneo. Esto ocurre debido a la carencia de ejemplos con el que el modelo fue entrenado, es decir, el modelo visualizó pocos ejemplos de manera que no tiene conocimiento de cómo proceder en el etiquetado.

Para realizar el entrenamiento se utilizó la librería SpaCy. Para entrenar un modelo NER en SpaCy es necesario dividir el total de datos. Una forma clásica es la división 80-20, es decir, 80 por ciento datos de entrenamiento y 20 por ciento restante datos de validación. Esto último es para evitar el sobre entrenamiento u overfitting.

4 Evaluación preliminar

En esta sección se describe la evaluación preliminar del algoritmo. Se emplearon 10 relatos de denuncias y un relato de denuncia a analizar denominada pivote. Luego se realizó el entrenamiento del modelo para el módulo NER. Se utilizó un conjunto de 70 datos que corresponden a temáticas como artículos del código penal, información sobre calles y su pertenencia a barrios. También se incluyó información de patentes o dominios de vehículos, así como marcas y modelos de los mismos. Esta última información puede ser utilizada para vincular casos en los cuales se utilizó un determinado vehículo para cometer un ilícito.

Como se indicó al final de la subsección 3.4, se dividió en conjunto de datos en entrenamiento y en validación, por tanto, se disponen de 56 datos de entrenamiento y 14 de validación. En la primera fase se reportaron los datos que se pueden apreciar en la figura 6.

Caso	Indice de similitud
Caso 120928	0.7391304347826086
Caso 120927	0.3146067415730337
Caso 122633	0.25217391304347825
Caso 122693	0.211864406779661
Caso 122709	0.2
Caso 120924	0.19626168224299065
Caso 122531	0.1889763779527559
Caso 122521	0.17796610169491525
Caso 122355	0.1391304347826087

Fig. 6. Resultado de la comparación de los relatos con respecto al pivote.

Luego se guardaron los 3 primeros casos, ya que los mismos tienen un porcentaje de similitud superior al 25%.

La fase dos consiste en la comparación de las entidades detectadas empleando el método de Jaccard. Esta aplicación permite generar un grado de refinamiento mayor entre los relatos de los casos. Esta comparación solo se efectúa sobre las entidades nombradas reconocidas.

En la figura 7 se puede apreciar el índice de similitud entre el pivote y las denuncias.

Número Caso	Similitud pivote
120928	60.0
120927	66.7
122633	50.0

Fig. 7. Similitud entre casos y pivote en fase dos

5 Trabajos relacionados

Diferentes trabajos y herramientas han sido desarrollados con el objetivo de dotar a investigadores, oficiales de policía y demás involucrados de soportes y ayudas para resolver casos que están actualmente detenidos a la espera de nuevos indicios como lo es la evidencia.

Se pueden citar trabajos como [10] el cual aprovecha el volumen de datos para determinar que crímenes pueden haber sido cometidos por un grupo de individuos.

Otros trabajos [11] [12] [13] [14] están relacionados con la fase dos del algoritmo del presente artículo. En estos artículos se proveen aproximaciones para extracción de entidades en reportes policiales y también se contemplan los idiomas portugués y neerlandés con sus normalizaciones específicas para esos lenguajes. Hafedh et al. [15] y Sedik et al. [16] extraen entidades nombradas de noticias policiales y de distintos recursos disponibles en la red, como blogs, artículos, páginas de noticias, etc.

Otro trabajo interesante que sirvió como punto de entrada para la fase uno del algoritmo del presente artículo fue [17]. En el cual se realiza la aplicación del índice de similitud de Jaccard para determinar qué grado de similitud existe entre escenarios. Como se puede ver, la flexibilidad de esta idea hace posible su utilización en el contexto de justicia penal. En [18] se dispone de extracción de entidades nombradas desde registros médicos. Dichos registros están compuestos por reportes de cirugía, enfermedades, etc. El artículo [19] ofrece información acerca de NER, así como los avances de esta disciplina. Finalmente, en [20] se propone un chatbot que hace uso de NER para mejorar sus respuestas. La extracción de entidades está incluida dentro del entendimiento del procesamiento de lenguaje natural, es decir su NLU ⁵.

6 Conclusiones

El trabajo presentado permite realizar un aprovechamiento del volumen de información disponible en forma de relatos de denuncias policiales. Actualmente esta información solo esta almacenada y es consultada cuando es requerida.

El procesamiento de lenguaje natural brinda herramientas para procesar de forma automática la información existente en las bases de datos, pudiendo abrir nuevos caminos en las investigaciones de delitos al detectar similitudes o patrones. Esta idea esta reforzada por la teoría del patrón delictivo, la cual enuncia que el crimen no es aleatorio, pudiendo detectarse patrones de comportamiento.

Como nota de interés, la puesta en producción de la herramienta, la cual incluye una interfaz gráfica y el algoritmo presentado como núcleo, representa un hito en el organismo donde se desempeña laboralmente el autor principal ya que no se dispone de ninguna tecnología basada en inteligencia artificial actualmente.

Este trabajo presenta una prueba preliminar del algoritmo y no es apta para un entorno de producción, ya que el conocimiento del modelo NER acerca del dominio es escaso.

Como trabajo futuro se propone la mejora del modelo NER, incluyendo 100 nuevos ejemplos de conocimiento variado. La idea es incrementar gradualmente el número de datos de a 100 elementos e ir controlando las métricas arrojadas en cuanto a desempeño.

Otro punto importante es la definición de una interfaz gráfica para cargar el relato de interés a analizar. Actualmente esto es realizado de forma manual y es confuso para un usuario no técnico del dominio del software.

Finalmente se propone el desarrollo de una representación visual de los datos procesados por el algoritmo, de esa manera los investigadores dispondrán de una vista más intuitiva de los casos y sus similitudes.

⁵ NLU se refiere a Natural Language Understanding.

Referencias

1. Aliero, A. & Bashir, S. & Aliyu, H. & Tafida, A. & Kangiwa, B. & Dankolo, N. (2023). Systematic Review on Text Normalization Techniques and its Approach to Non-Standard Words. *International Journal of Computer Applications*. 185. 975-8887.
2. Khyani, D. & Siddhartha B. S. (2021). An Interpretation of Lemmatization and Stemming in Natural Language Processing. *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*. 22. 350-357.
3. von der Brück, T., Pouly, M.: Text similarity estimation based on word embeddings and matrix norms for targeted marketing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies vol 1* pp. 1827-1836. (2019).
4. Salman, N. & Muhammad, G., & Sohaib, & Khalid Alvi, Sohaib & Kiran, Anam & Rehman, Shafique Ur & Murtaza, Ghulam & Campus, Jehlum & Jehlum, Pakistan. (2022). Named Entity Recognition (NER) in NLP Techniques, Tools Accuracy and Performance.
5. Brantingham, P. & Brantingham, P. (2013). Crime pattern theory. *Environmental Criminology and Crime Analysis*. 78-93. 10.4324/9780203118214.
6. Spacy. <https://spacy.io/>. Accedido el 25 de julio de 2024.
7. Código Penal Argentino. <https://servicios.infoleg.gob.ar/infolegInternet/anexos/15000-19999/16546/texact.htm>. Accedido el 22 de julio de 2024.
8. Collet, F. *Deep Learning with Python Second Edition* (2021) Manning publications. ISBN 9781617296864. Pp. 133-138.
9. NER Annotator. <https://tecoholic.github.io/ner-annotator/>. Accedido el 25 de julio de 2024.
10. Cocx, T. & Kosters, W. (2006). A Distance Measure for Determining Similarity Between Criminal Investigations. 511-525. 10.1007/11790853_40.
11. Chau, M. & Xu, J. & Chen, H. (2002). Extracting Meaningful Entities from Police Narrative Reports.
12. Carnaz, G. & Nogueira, V. & Antunes, M. & Ferreira, N. (2019). Named-Entity Recognition for Portuguese Police Reports.
13. Schraagen, M. & Bex, F. & Brinkhuis, M. (2017). Evaluation of Named Entity Recognition in Dutch online criminal complaints. *Computational Linguistics in the Netherlands Journal*. 7.
14. Al-Moslmi, T. & Ocaña G.M. & Opdahl, A. & Veres, C. (2020). Named Entity Extraction for Knowledge Graphs: A Literature Overview. *IEEE Access*. 8. 32862 - 32881. 10.1109/ACCESS.2020.2973928.
15. Hafedh S. & Nazlia O & Khamael, R. (2014). Named Entity Recognition in Crime Using Machine Learning Approach. 280-288. 10.1007/978-3-319-12844-3_24.
16. Sedik, R. & Romadhony, A. (2023). Information Extraction from Indonesian Crime News with Named Entity Recognition. 1-5. 10.1109/KST57286.2023.10086789.
17. Delle Ville J., Torres D., Fernández A., Antonelli L. An Approach to Cluster Scenarios According to their Similarity using Natural Language Processing. Lifa, Fac. De Informática, UNLP, La Plata, Bs As, Argentina.
18. Wen, C. & Chen, T. & Jia, X. & Zhu, J. (2021). Medical Named Entity Recognition from Un-labelled Medical Records based on Pre-trained Language Models and Domain Dictionary. *Data Intelligence*. 3. 1-13. 10.1162/dint_a_00105.
19. N, M. & Suraparaju, V. & Mallaiah, S. (2024). Deep Analysis and Advancements in Named Entity Recognition -NER. 45. 171.
20. Ali, N. (2020). Chatbot: A Conversational Agent employed with Named Entity Recognition Model using Artificial Neural Network. *ArXiv*, abs/2007.04248.