**Estado del arte de la investigación, análisis de textos**

A partir de la transformación digital, ha sido notorio el cambio de los hábitos de consumo, de aquí, por qué surge la creación de negocios globales a través de internet. Para relacionarse con clientes y satisfacer sus necesidades, cada empresa ha ido generando un cambio adaptativo creando estructuras de trabajo que requieren buscando así perfiles profesionales que se adapten a dichas necesidades, se recomiendan aproximaciones a la demanda del mercado a instituciones académicas, indicando qué conocimientos y capacidades se exigen para perfiles y áreas de trabajo.

Brewer (2013) afirma que muchos jóvenes tienen dificultades para encontrar trabajo por el desajuste entre su formación y las habilidades que exige el mercado laboral. También existen diferentes autores que afirman que la competencia de los estudiantes o graduados son distintas y discrepan entre las que solicitan los empleadores y se necesitan para diversas áreas de su proceso laborar. (Álvarez-Flores, Núñez-Gómez, & Olivares-Santamarina, 2018)

Basado en esto y tomando como punto de partida el gran auge del internet, el cual cobra cada día mayor protagonismo en todos los ámbitos de la vida de las personas, tanto en el plano profesional, educativo, político, social, económico y demás, éste sigue en un imparable desarrollo y cambiando los parámetros tradicionales de la sociedad.

*La comunicación multidireccional y multimedia, la convergencia e instantaneidad, la ubicuidad y la inteligencia colectiva y emocional son algunas de las características que se pueden percibir dentro de éste espacio donde no existe frontera o barrera que impida una efectiva comunicación. (Sánchez Gonzáles & Méndez Muros, 2013)*

El internet como generador de la democratización digital, ha cambiado el hábito de las personas hacia nuevas formas de relación, información y comunicación, tal es así que el número de internautas ha aumentado de forma considerable, un 7,1 por ciento en el último año y se acerca a los 22,2 millones de personas. Éste parte de un estudio realizado en España en el que sacan como conclusión que el resultado podría derivarse del porcentaje de hogares españoles con acceso a Internet, que asciende al 59,1 por ciento frente al 54,0 por ciento del año anterior, según un estudio del Instituto Nacional de Estadística en el 2011 sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los hogares de 2010. (Sánchez Gonzáles & Méndez Muros, 2013)

Los medios sociales, potenciados por la innovación tecnológica, se constituyen en el fenómeno emergente sociocultural de esta época porque han cambiado el hábito de consumo de los internautas, provocando que se origine una sociedad hiperconectada. Cabe resaltar que los usuarios que más tiempo pasan en Internet tienen un rango de edad comprendida entre los 25 y 34 años, seguidos por los de 35 a 44 años, según un Estudio General de Medios (EGM). (Sánchez Gonzáles & Méndez Muros, 2013)

Bajo esta premisa, la sociedad en Red avanza hacia audiencias cada vez más activas inmersas no sólo en el proceso de participación, sino en la producción y distribución de los mensajes. Esto resulta ser artífice del cambio que han protagonizado los usuarios dentro del proceso de comunicación, es decir, dejan de ser receptores pasivos para convertirse en agentes productores del mensaje.

Esto resulta de una tendencia comunicativa que ha despertado el interés de empresas, instituciones y medios de comunicación porque les permite establecer nuevos vínculos con el usuario según los intereses y necesidades de éstos.

*“Las TIC no son sólo una red a la que se suman los individuos, sino que actúan como tecnologías sociales cuyo perfeccionamiento depende tanto de la diversidad de sus funciones (sociales, políticas, cognitivas, económicas, etc.) como de la flexibilidad con que se adapten a nuestra diversidad funcional” (Navarro, 2009:141148)**(Sánchez Gonzáles & Méndez Muros, 2013)*

Ahora, lo que hay que revisar es si estos conocimientos se llevan a cabo en las universidades, tanto españolas (como lo plantean en el artículo) y también en el resto de las universidades del mundo.

La evaluación sobre la demanda laboral de nuevos perfiles profesiones debería ser llevada a cabo a través de técnicas cualitativas que puedan permitir a la institución a adentrarse en una fase de descubrimiento para el acercamiento del perfil profesional más activo en el ámbito laboral por el momento y de esta forma educar a los estudiantes para que se puedan desarrollar en ambientes más conocidos y solicitados.

Desde éste punto se puede tocar el concepto de calidad en el ámbito universitario que conlleva a establecer una analogía entre el entorno productivo y el proceso educativos superior, de tal forma que se pueda responder a la cuestión de quién es el cliente y qué es el producto, existiendo diferentes visiones al respecto. (Arquero-Montaño, Donoso-Anes, Jiménez-Cardoso, & González González, 2009)

Para adentrarse más en esto, se puede considerar como un común acuerdo entre la universidad y la sociedad que establece las bases de la relación entre el estudiante y la universidad.

Así, los clientes iniciales del proceso educativo son la sociedad, lo cual conlleva al siguiente paso que resulta ser el identificar las necesidades de los grupos interesados que la componen, junto con las necesidades que establezca hacia un perfil determinado y que sirva como orientación a la intervención educativa, esto es fundamental tanto para profesores como alumnos para orientar adecuadamente sus acciones.

Bajo estas premisas, es necesario un análisis de los objetivos de formación de la titulación, acorde con las exigencias profesionales, no es únicamente una información útil para la gestión del proceso de cambio, sino que debe prevalecer por encima de otros intereses en el diseño de los grados, no solo para formar estudiantes bajo el concepto de calidad, sino para servir de igual forma a la sociedad, en la que ellos tendrán que desarrollarse como profesionales y deberán cumplir a cabalidad con los criterios dados por las empresas, ya sean entidades públicas o privadas.

En estos criterios se tiene en cuenta si la titulación tiene claramente definido el perfil del egresado, las competencias que lo conforman y si dicha definición está contrastada con estudios socio-económicos desarrollados sobre el tejido empresarial, en los que se determinen las demandas sociales y se cuente con la opinión de los colectivos profesionales implicados. En el mismo sentido, las guías de evaluación de las titulaciones tienen en cuenta la revisión periódica de los programas, en función de los resultados, en términos de inserción laboral y satisfacción de graduados y empleadores mediante planes de seguimiento. (Arquero-Montaño et al., 2009)

Aplicando estos conceptos, y trayéndolos al ámbito americano, especialmente estadounidense, se debe decir que cada audiencia se destaca la valoración del grado en que la Universidad cumple globalmente sus funciones y cada función en particular, es decir, su perfil, así como las diferencias entre distintos estratos de cada audiencia, finalizando con la presentación de las diferencias en los perfiles valorativos de la funcionalidad de la universidad entre las distintas audiencias. (Consenso & Enfermedad, 2009)

La demanda de igualdad en las sociedades contemporáneas desarrolladas ha estimulado una expansión sin precedentes de los servicios educativos. Por ende, las inversiones en educación resultan ser más elevadas a las de otras épocas, de aquí que la oferta curricular sea alta y diversificada, por lo que las compañías tienden a buscar un perfil laboral especifico que se adapte a sus necesidades, pero, si se adentra a los estudiantes que son el principal foco, estos consideran que no se le ha dado suficiente importancia a los aspectos prácticos de la formación y creen que su título carece de calidad; agregándole a esto que su perfil formativo no responde ni a las expectativas sociales ni a las necesidades de las empresas públicas o privadas. (Consenso & Enfermedad, 2009)

Esto lleva a una reestructuración, en la que resulta sugerible que las universidades puedan adaptarse a las necesidades de las empresas con el fin de obtener un proceso formacional adaptativo.

La definición de perfiles y el conocimiento mutuo de las instituciones universitarias favorece el desarrollo de sinergias, el fortalecimiento en el cumplimiento de las misiones de la universidad, la clarificación de las oportunidades de colaboración con la sociedad civil y la consolidación de las redes educativas, económicas y académicas.

Resulta imprescindible la relación universidad, empresa, sociedad está generando actualmente una demanda que obliga a afrontar las necesidades reales de la institución con la mirada puesta en sus misiones y en su dimensionamiento. La extensión universitaria y la comunicación institucional, como formas de proyección y de realización interna y externa de la universidad, son instrumentos incuestionables de la búsqueda de calidad, tanto en el sistema ciencia-tecnología-sociedad, como en el conjunto investigación-universidad-empresa. (Touriñán López, 2005)

En el año 1996 se presentó el informe de la Comisión 12 de las Comunidades Europeas, conocido como el Libro Blanco de la Educación y la Formación. Lo relevante del libro es el planteamiento utilizado, el cual toma el valor educativo de la inversión en capital humano y el valor económico de la educación en un momento de crisis socio-económica. Parte de tres acontecimientos: la sociedad de la información, la mundialización y la civilización científico-técnica en el que se proponen dos objetivos:

1. Revalorizar la cultura general para cualquier tipo de formación
2. Desarrollar la aptitud para el empleo en todos los niveles de formación.

En éste contexto, el de la sociedad del conocimiento y de la información se hace necesario, un rediseño de la vida humana en el que se estructure la educación en conjunto con el trabajo que formen una interacción de la organización de la vida en el siglo XXI. (Touriñán López, 2005)

Éste rediseño no podrá llevarse a cabo sin que se introduzcan las adaptaciones en los sistemas educativos y, en concreto en los universitarios, que deberán afrontar el reto que trae para la institución la extensión de sus áreas de acción en el nuevo entorno para el perfil laborar.

Esto lleva a someter a las universidades a una mayor presión para satisfacer las demandas laborales. Visto desde el punto del trabajador académico, estos procesos se hacen visibles a través de una multitud de nuevas tecnologías sociales a las que deberán adaptarse para estar acorde a las necesidades de las empresas en cuanto a los perfiles educativos de los egresados. (Krejsler, 2006)(Traducción)

Se puede notar la necesidad que genera el tener una visión mucho más global acerca del perfil profesional solicitado por las empresas en cuanto a los egresados de las diferentes profesiones, y la importancia que esto trae consigo tanto para las empresas como para las instituciones y ejercer la calidad llevada en cada uno de sus planes de estudio.

Resulta importante resaltar la variedad de perfiles educativos con los que salen los egresados incluso de la misma área, pero de diferentes instituciones, lo que hace funcional a una sociedad, porque los perfiles educativos de estos necesitan ser adaptativos a las diferentes necesidades de cada empresa.

**Técnicas de Análisis de Textos**

**Que es el análisis de textos:**

Se suele llamar análisis de contenido al conjunto de procedimientos interpretativos de productos comunicativos que proceden de procesos singulares de comunicación previamente registrados, y que, basados en técnicas de medida, a veces cuantitativas (estadísticas basadas en el recuento de unidades), a veces cualitativas (lógicas basadas en la combinación de categorías) tienen por objeto elaborar y procesar datos relevantes sobre las condiciones mismas en que se han producido aquellos textos, o sobre las condiciones que puedan darse para su empleo posterior.[2]

Luego de la Extracción de Datos el siguiente paso a realizar es el análisis de contenido o también conocido como text Mining en el cual se analiza , selecciona , separa la información para la mejor obtención de resultados o también puede ser Explicado cómo.

*“la aplicación de algoritmos y métodos de los campos del aprendizaje de máquina y la estadística sobre los textos con el objetivo de encontrar patrones útiles”*

Lo primero en el text Mining es recopilar los documentos relevantes. Luego la principal tarea es limpiar las muestras y asegurar la calidad, debido a que se pueden encontrar diversos formatos entre los documentos. La transformación del texto a datos numéricos permite que los datos cambien a una codificación clásica de data Mining. Por lo que la presentación de los datos para data Mining y text Mining son solo diferentes en una presentación inicial . Existen muchas variantes de representación de documentos, la mayoría de los enfoques están basados en que las palabras son atributos y los documentos son muestras, formando el conjunto de datos que permitirá desarrollar los distintos métodos de aprendizaje. Las fuentes de conocimiento externas pueden beneficiar enormemente varios de los elementos que componen un sistema de text mining, entregando limitaciones o información adicional sobre los conceptos que se encuentran a través de la aplicación de métodos de aprendizajes, así como servir en la etapa de validación de los resultados[3]

**Tokenización:**

para el manejo de texto es importante separar el stream de caracteres para llevarlo a palabras o, más precisamente, a tokens. El proceso de tokenización es altamente dependiente del lenguaje sobre el cual se está trabajando y las reglas específicas de este, por lo que esta tarea resulta trivial para una persona familiarizada con la estructura del lenguaje, mientras que para un programa computacional puede convertirse en todo un reto [84]. Para identificar las distintas palabras (o tokens) es necesario primero definir los delimitadores de los tokens, los que generalmente corresponden a los signos de puntuación y otros caracteres distintos a las letras del alfabeto . Luego los delimitadores de tokens se separan de las palabras y son reemplazados por un espacio blanco simple. De esta forma cada palabra queda separada por un espacio blanco simple y facilita la tokenización.

**Stopwords**:

Stop words es un método de filtración que consiste en remover palabras comunes que resultan inútiles para la caracterización de un documento. La idea de la aplicación de una lista de stop words es remover palabras que ocurren muy poco o no contienen información útil, como artículos, pronombres, conjunciones, preposiciones, etc. Las palabras que aparecen en la mayoría de los documentos aportan poca información para distinguir los distintos tipos de documentos, así como las palabras que ocurren muy raramente es probable que no tengan una relevancia estadística y pueden ser removidas del diccionario. Al remover las palabras de la lista stop words se puede lograr una reducción significativa del tamaño del diccionario [84], pero no existe una lista fija de stop words que sea universalmente utilizada.

**Stemming:**

El objetivo de stemming es reconocer los conjuntos de palabras que pueden ser tratadas como equivalentes. Muchas veces no hay necesidad para mantener el singular y plural de una misma palabra, así como los verbos pueden ser almacenados es su forma infinitiva. También se puede extender el concepto hacia los sinónimos. Algunos de los efectos de la aplicación de stemming es la reducción del número total de atributos dentro del texto (o reducción del tamaño del diccionario ) y el incremento de la frecuencia de ocurrencia de algunos atributos. Tal como para stop words, no hay un algoritmo de stemming que sea universalmente usado, donde el idioma tiene un papel clave. Por otro lado, la aplicación de stemming debe ser implementado con cautela para no eliminar palabras del diccionario que puedan resultar relevantes, dado que no se considera la semántica de las palabras

**Clasificacion de documentos:**

La clasificación de textos puede ser vista como un proceso de dos pasos . En el primer paso se llama de aprendizaje (o entrenamiento) en el cual se construye el clasificador a partir del análisis de un subconjunto de documentos cuya clasificación asociada es conocida previamente. En el segundo paso, se evalúa la exactitud de predicción del clasificador construido, para lo cual se le aplica el modelo encontrado a un subconjunto de documentos de testeo (independiente del conjunto de entrenamiento) y se le asigna una clase a cada documento, para finalmente comparar esta asignación con la clase original de los documentos de testeo. La clasificación de documentos puede ser single-label o multi-label. En el primer caso cada documento pertenece exactamente a una clase y en el segundo caso cada documento puede estar asociado a más de una clase. El éxito en la clasificación depende en gran parte de si es posible hallar patrones de palabras en distintos subconjuntos de documentos, dichos patrones se forman cuando se producen combinaciones de palabras en una clase y no en el resto.

**Lemmatization (lematización):** técnica que se apoya en la agrupación por morfemas. El algoritmo del sistema de análisis cuenta con un diccionario del idioma e información de la morfología para su asociación (Singh y Gupta, 2017). Esta técnica necesita un conocimiento previo de la lengua para la agrupación de palabras, por lo que podría partir con ventaja con respecto a la anterior, pero según estudios como los de Kettunen, Kunttu y Järvelin (2005), o los de Balakrishnan, Humaidi y Lloyd-Yemoh (2016) los resultados realizados con ambas metodologías y en distintos idiomas han ofrecido diferencias insignificantes.[4]

**Software para analizar los datos**

Los textos pueden analizarse desde dos niveles de complejidad:

1. El nivel de datos, es el nivel más simple desde el punto de vista operacional. El objeto de análisis es el texto puro, entendido como un conjunto de caracteres. Las operaciones automáticas que pueden realizarse son, únicamente, aquellas basadas en la manipulación de los caracteres del texto. Por ejemplo, la localización de todas las palabras que comienzan por un prefijo determinado, las palabras que acompañan a otra dada (colocaciones), o la frecuencia de aparición de un término (C, G, & Cruz, 2018).
2. El nivel de información se relaciona con la interpretación de los textos. A este nivel corresponden las operaciones que necesitan disponer, además de los textos, de la interpretación de los mismos, como por ejemplo la localización de todos los verbos transitivos en una colección de textos o los textos que tratan sobre “las enfermedades del salmón”. Este tipo de consultas más “inteligentes” requieren que el texto tenga marcado de alguna forma este tipo de información. Es importante tener en cuenta que el procesamiento automático de la información es más complicado que el procesamiento de datos y que sólo puede realizarse si previamente se ha preparado (pre-procesamiento) el texto (datos) insertando la interpretación (semántica) de estos datos textuales. Uno de los mecanismos de pre-procesamiento es la inserción en el texto de marcas o etiquetas explícitas con la información asociada a cada elemento textual (proceso de etiquetado)(C et al., 2018).

**Nubes de palabras**

Las nubes de palabras son colecciones de vocablos organizados en forma compacta, en donde el tamaño de la fuente codifica la relevancia de cada palabra en el documento. Para la generación de las nubes de palabras se mapeó el valor de FT-FID de la palabra al tamaño de esta. El tamaño de las palabras va descendiendo de manera logarítmica, con un punto de corte dado por el área disponible y en donde la ubicación de cada una de estas es determinada al azar. Todos los procesamientos y análisis de los datos se realizaron con el lenguaje de programación Python 316, utilizando las bibliotecas NumPy17, Pandas18, Scikit-Learn19 y NLTK20 (Villena & Dunstan, 2019).

**Herramientas informáticas para PLN** :

Existe una amplia variedad de herramientas informáticas para el PLN en diversos idiomas. En este trabajo se seleccionaron algunas Open Source que se pueden utilizar en programas Python . La comunidad científica utiliza en mayor medida el lenguaje interpretado Python por lo que sólo nos centraremos en aquellas bibliotecas que se puedan utilizar en este lenguaje de programación. Cabe mencionar que, si bien no todas son nativas para este lenguaje, como Stanford NLP, existen plugins o APIs que permiten enlazarlas. En el contexto de este trabajo se utilizarán los términos “herramientas”, “librerías”, “bibliotecas” y “paquetes” como sinónimos.[10]

**NLTK (Natural Language Toolkit)** es una plataforma para la creación de programas Python para análisis de texto. Como fue creado con fines educativos en el año 2001, permite la realización de proyectos con distintos propósitos y alcance variado. Provee demostraciones paso a paso de distintos algoritmos. Integra más de 50 corpus y un conjunto de librerías para segmentación, tokenización, etiquetado del habla, análisis sintáctico y semántico, entre otros. Para el español muchas de estas opciones no están disponibles. Su instalación en Python es sencilla debido a que es un paquete que se encuentra en el repositorio del lenguaje. Con una simple instrucción, se instala lo necesario para utilizarlo. El etiquetador POS de NLTK originalmente fue creado para el idioma inglés. Sin embargo, se lo puede extender al español entrenando el etiquetador con algún corpus que incorpora NLTK en esta lengua o usando un etiquetador externo como el de Stanford. Para este trabajo se seleccionó la primera opción con el corpus cess\_esp que tiene 500.000 palabras y 610 archivos. El etiquetado POS que retorna NLTK se corresponde con el formato EAGLES3 . Las etiquetas EAGLES codifican todas las características morfológicas existentes para la mayoría de los idiomas europeos, incluido el español. Estas etiquetas consisten en un conjunto de caracteres de longitud variable donde cada uno corresponde a una característica morfológica. El primer carácter en la etiqueta es siempre la categoría POS que, a su vez, determina la longitud y la interpretación de los elementos siguientes.

**Freeling:**

Freeling4 es una librería de código abierto para el análisis de texto (tokenización, análisis morfológico, detección de entidades nombradas, etiquetado POS, etc.) para una variedad de idiomas. Desde su primera versión, tiene soporte para el español. Fue desarrollado por el Centro de Tecnologías y Aplicaciones del Lenguaje y el Habla (TALP) de la Universidad Politécnica de Catalunya. Se puede utilizar y ampliar los recursos lingüísticos por defecto (diccionarios, lexicones, gramáticas, etc.) adaptándolos a dominios específicos. Su código fue escrito en C++ bajo una arquitectura cliente – servidor. Existen diversas APIs que permiten utilizar Freeling en otros lenguajes, incluido Python. Su instalación no es sencilla, sobre todo en entornos Windows. Sin embargo, subsanado este inconveniente, al importar la API desde el programa Python se puede invocar a todas las funciones implementadas. El analizador morfológico de Freeling realiza el etiquetado POS usando las etiquetas EAGLES.

**Pattern.es**

Pattern.es5 es una biblioteca para Python que contiene herramientas para la conjugación de verbos, la singularización o la pluralización de sustantivos, la división de chunks y permite realizar el etiquetado POS para español. En la versión para el idioma inglés adiciona una interfaz para WordNet. Su instalación y uso son muy sencillos. Para el etiquetado gramatical utiliza Penn TreeBank Tags. Penn TreeBank está basado en el corpus Brown, pionero en etiquetado POS para inglés. A pesar de haberse creado para el inglés, Pattern lo usa también para el español. Cada etiqueta está formada con dos o tres caracteres que indican la función que cumple cada palabra en una oración.

**SPACY**

SpaCy6 es una biblioteca para el procesamiento avanzado de lenguaje natural en Python. Incluye modelos estadísticos pre-entrenados y vectores de palabras, admite tokenización para más de 45 idiomas. Provee etiquetado, análisis y reconocimiento de entidades nombradas y una fácil integración de aprendizaje profundo. Es muy sencilla de utilizar, basta con incorporar desde un programa Python el modelo para el español. Devuelve un etiquetado POS completo, no sólo indica la función de la palabra en la oración, sino otros datos tales como tiempo verbal, persona, número, modo, género, entre otros. La tokenización y el etiquetado gramatical se basan en el corpus OntoNotes5 7 que sigue la sintaxis de Penn TreeBank.

**Stanford NLP**

Stanford NLP es un framework diseñado y desarrollado para Java que provee soluciones para la mayoría de las tareas más comunes en el procesamiento de lenguaje natural que van desde la tokenización hasta el reconocimiento del discurso. No todas las funciones están implementadas para el español. Existen APIs para otros lenguajes de programación, entre ellos, Python8 . NLTK incluye algunas de sus funciones, aunque si se desea hacer un análisis más profundo, siempre se debe correr el servidor de Java en segundo plano y prescindir de NLTK. Permite realizar etiquetado POS con el formato EAGLES y reconoce entidades para el análisis morfológico. Para el análisis sintáctico, identifica grupos lingüísticos o chunks y se puede obtener el árbol de dependencia. Su instalación no es tan sencilla como otras de las herramientas analizadas por el hecho de tener que instalar el servidor Java y definir variables de entorno

**REFERENCIAS**

Álvarez-Flores, E.-P., Núñez-Gómez, P., & Olivares-Santamarina, J. P. (2018). Perfiles profesionales y salidas laborales para graduados en Publicidad y Relaciones públicas: de la especialización a la hibridación. *El Profesional de la Información*, *27*(1), 136. https://doi.org/10.3145/epi.2018.ene.13

Arquero-Montaño, J. L., Donoso-Anes, J. A., Jiménez-Cardoso, S. M., & González González, J. M. (2009). Análisis Exploratorio Del Perfil Demandado Para Administración Y Dirección De Empresas: Implicaciones Para El Área Contable. *Revista de Contabilidad*, *12*(2), 181-213. https://doi.org/10.1016/s1138-4891(09)70006-8

C, A. J. S., G, R. V. B., & Cruz, R. (2018). *A Review of Automatic Requirements Extraction from Natural Language Text lenguaje natural*. (December).

Consenso, P., & Enfermedad, V. D. E. (2009). *a Rchivos V Enezolanos De P Uericultura Y P Ediatría*.

Krejsler, J. (2006). Discursive Battles about the Meaning of University: The Case of Danish University Reform and its Academics. *European Educational Research Journal*, *5*(3-4), 210-220. https://doi.org/10.2304/eerj.2006.5.3.210

Mistral, G. (2019). *ANÁLISIS ESTILOMÉTRICO EN PYTHON*.

Sánchez Gonzáles, H., & Méndez Muros, S. (2013). Correlación entre la demanda laboral y la formación universitaria. *Estudios sobre el Mensaje Periodístico*, *19*, 981-993.

Touriñán López, J. (2005). Universidad, sociedad y empresa: orientaciones estratégicas de extensión universitaria y comunicación institucional. Visión prospectiva desde el marco legal español. *Revista galego-portuguesa de psicoloxía e educación: revista de estudios e investigación en psicología y educación*, *12*(12), 163-186.

Villena, F., & Dunstan, J. (2019). Automatic keyword retrieval from clinical texts: An application of natural language processing to massive data of Chilean suspected diagnosis. *Revista Medica de Chile*, *147*(10), 1229-1238. https://doi.org/10.4067/s0034-98872019001001229