

Los Repositorios Institucionales como una herramienta para el seguimiento y verificación de los Objetivos de Desarrollo Sostenible. Caso de estudio:

Argentina

Jose Texier

Universidad Nacional de Chilecito

jtexier@undec.edu.ar

ORCID: [0000-0003-0176-6625](https://orcid.org/0000-0003-0176-6625); [GScholar](#); [ScopusID](#); [PublonsID](#)

Resumen: El objetivo general del sistema es la construcción de un clasificador automático de recursos del Sistema Nacional de Repositorios Digitales de acuerdo con las 169 metas de los 17 Objetivos de Desarrollo Sostenible (ODS) definidos por la Organización de las Naciones Unidas. El clasificador consistirá en un desarrollo computacional capaz de analizar diferentes metadatos (llámense título, resumen, palabras clave y texto completo) aplicando técnicas del análisis de texto para establecer una relación sintáctica y/o semántica con los ODS, que le permita al usuario final hacer uso de los recursos o ítems del repositorio o repositorios de acuerdo con la clasificación lograda, es decir, datificar los repositorios institucionales. Esta línea de investigación promueve una cultura de compartición de recursos, vinculación de personas en actividades e investigaciones académicas y científicas similares, contribución a la generación de nuevos conocimientos y disminución de las brechas que dificultan el acceso de la información de temas relacionados con los ODS a través de los ítems depositados en Repositorios Institucionales Argentinos. El plan de trabajo contempla tanto aspectos de investigación básica (mejora y análisis de algoritmos de procesamiento de texto), investigación aplicada (beneficio directo de la sociedad sobre el tema de los ODS) e investigación de campo (recolección y disponibilidad de recursos digitales de todos los repositorios de la Argentina para un fin específico, área de investigación base de investigador), así como también transferencia de conocimiento (el proyecto podría integrarse como un componente de software en los diferentes tipos de software de repositorios existentes o como un sistema de software de manera independiente y externo a los RI).

Palabras clave: procesamiento de lenguaje natural, ODS, RI, datificación, Argentina

El desarrollo del sistema se basa en una línea de trabajo cuyo eje es el Procesamiento de Lenguaje Natural (PLN) que emplea diferentes enfoques aplicables en lo local, regional e internacional. Implica una línea de estudio que contempla investigación básica, aplicada y de campo, así como también transferencia de conocimiento por cuanto contribuyen a: la mejora y análisis de algoritmos de procesamiento de textos en vinculación con los Objetivos de Desarrollo Sostenible (ODS); la aplicación en los diferentes niveles educativos; la recolección y la disponibilidad de recursos de los Repositorios Institucionales de Argentina; el diseño e implementación de un software como parte del repositorio y/o como un sistema externo y la articulación con la docencia y lo científico. Se toman los ODS como eje de trabajo porque representan una agenda universal de desarrollo sostenible, donde los países adoptantes deben seguir una estrategia global que combine el desarrollo económico, la inclusión social y la sostenibilidad ambiental. En la Asamblea General de las Naciones Unidas, el 25 de septiembre de 2015 fueron aprobados los 17 ODS con 169 metas y 232 indicadores y fueron declarados como universales, indivisibles e interdependientes (Naciones Unidas, 2015).

En los últimos tiempos se ha observado el crecimiento de los Repositorios Institucionales (RI) argentinos en el directorio internacional de repositorios OpenDOAR (Jisc, 2020). La Universidad Nacional de Chilecito (UNdeC) no cuenta con un RI (diciembre de 2020), no obstante, ese no sería un obstáculo para esta propuesta por cuanto podría implementarse en un lapso corto de tiempo y/o se cosecharía su producción a través de las herramientas existentes. Por ello, se plantea una propuesta (con los RI argentinos y/o los recursos de la UNdeC) que se sustente en la siguiente hipótesis: *¿Será que los Repositorios Institucionales argentinos pueden aportar un valor agregado en la implementación, verificación y seguimiento de los Objetivos de Desarrollo Sostenible de Argentina?* En este sentido, el objetivo general es implementar un clasificador automático de recursos de los Repositorios Institucionales argentinos de acuerdo con las 169 metas de los 17

Objetivos de Desarrollo Sostenible definidos por la Organización de las Naciones Unidas (ONU).

El clasificador funcionará como parte del RI de la UNdeC o como una tecnología agnóstica al repositorio. Consistirá en un desarrollo computacional novedoso capaz de analizar diferentes metadatos (título, resumen, palabras clave y texto completo) aplicando técnicas del análisis de texto para establecer una relación sintáctica y/o semántica con los ODS, que le permita al usuario final hacer uso de los recursos o ítems gracias a la clasificación lograda y a los procesos de datificación. La propuesta se implementará con el enfoque de Ciencia Abierta (Karisma, 2018; Kramer & Bosman, 2016; UNESCO, 2020; Vicente-Saez & Martínez-Fuentes, 2018) y se desarrollará bajo la modalidad proyectiva (Hurtado, 2008) porque implica la creación y diseño de un sistema de software. Este software será diseñado y construido con la metodología del Proceso Unificado Ágil (Ambysoft, 2020).

Los objetivos específicos propuestos son:

1. Describir la necesidad de relacionar los diversos recursos existentes en los RI argentinos según los ODS.
2. Analizar las diferentes propuestas de identificación de recursos en los repositorios a través de los diferentes metadatos (título, resumen, palabras clave y texto completo) de los recursos.
3. Caracterizar los diferentes tipos de recursos y sus relaciones con los Objetivos de Desarrollo Sostenible.
4. Identificar los procesos de interoperabilidad que permiten la relación sintáctica y semántica de los recursos con otros repositorios/sistemas de acuerdo con los Objetivos de Desarrollo Sostenible.
5. Anticipar las tendencias de los recursos ODS en los Repositorios Institucionales de acuerdo con las políticas de Estado definidas por el gobierno chileno y la región según sus necesidades.
6. Alcanzar un mayor entendimiento de los ODS y su implementación en el contexto local, regional y nacional.
7. Diseñar una arquitectura de software que garantice una escalabilidad en las necesidades y en la adaptación de las relaciones entre los recursos de los repositorios y los ODS, dentro del RI de la UNdeC o externo al repositorio.
8. Implementar un sistema de software que clasifique los recursos existentes de los Repositorios Institucionales de acuerdo con las 169 metas correspondientes a los 17 Objetivos de Desarrollo Sostenible.
9. Generar redes multidisciplinarias de investigadores, estudiantes y becarios relacionadas con la línea y/o con nuevos proyectos vinculados con el medio (Universidad y región) sobre la base de la datificación y el análisis de texto.
10. Evaluar el funcionamiento del software desarrollado y su aplicación en la docencia para mejorar y crear otras propuestas innovadoras y generadoras de nuevos conocimientos.

Este proyecto sienta sus bases en la relación de las Ciencias de la Computación y las Ciencias de la Información y Documentación. El desafío actual no está solo en la cantidad y calidad de los datos de los RI, sino en qué se debe hacer con todos ellos y cómo se gestionarán e integrarán con grandes volúmenes de información digital estructurada, semiestructurada y no estructurada, creadas por personas y/o organismos en todos los ámbitos de la sociedad (Griffiths et al., 2016; Suber, 2012). Por tanto, la datificación se hace necesaria para poder transformar tales datos de los

RI, en información susceptible a ser utilizada en cualquier área del conocimiento (Lope Salvador et al. 2020; Markus, 2017).

El concepto de datificación se está desarrollando en publicaciones académicas y científicas desde hace más de cinco años, pero sus procesos (recuperación, transformación y visualización) se trabajan en forma independiente desde hace mucho más tiempo y tiende a confundirse con el concepto de *Big Data* (Lope Salvador et al., 2020; Markus, 2017) que surge como una tecnología que permite procesar grandes cantidades de datos para recolectarlos, analizarlos y usarlos, (Cf. Rajaraman & Ullman, 2011), sobre todo porque el procesamiento y posterior análisis de los datos y/o información, en muchos casos, deben ser realizados en tiempo real para incidir en oportunas tomas de decisiones.

Ahora bien, cómo se debería encarar la recolección, análisis y uso de los datos también es inherente a la datificación, por ello, se entiende que *Big Data* y datificación están vinculados. Sin embargo, esta última tiene que ver con la acción de convertir y medir en datos una información, un proceso o una actividad gracias a la infraestructura tecnológica existente (Agarwal et al., 2011; Markus, 2017; Mayer-Schönberger & Cukier, 2013; Tascón, 2013; Yu & Wang, 2015). A su vez, se relaciona la datificación con la digitalización. Mayer y Cukier (2013) plantean que la datificación ocurrió antes de las Tecnologías de Información, hecho que permitió el surgimiento de la digitalización. Estos autores explican que la datificación hace alusión al dato transformado, por medio del análisis y la reorganización, en información susceptible de ser utilizada en cualquier área del conocimiento o disciplina. En cambio, la digitalización (Mayer-Schönberger & Cukier, 2013) es el proceso mediante el cual la información analógica se convierte en digital para que sea legible y procesada por un computador. En síntesis, con la datificación los datos se pueden cuantificar para ser tabulados y analizados, luego, se almacenan en diferentes formatos para aplicar *Big Data* (Marín & Lisbona, 2018), siempre y cuando exista el componente de tiempo real. En otras palabras, datificar es analizar antes de recolectar datos y esto se logra con educación e investigación y no por un golpe de suerte o por el simple manejo de una herramienta.

Este plan de trabajo toma como base los datos y la información que, de manera ordenada y formal, se encuentran en los Repositorios Institucionales y otras fuentes bibliográficas para transformarla (análisis de texto) en información útil y vinculada con los ODS. Los RI son sistemas informáticos (DSpace, EPrints, Fedora Commons, Greenstone, CKAN, Junar, etc.) interoperables que alojan recursos científicos, académicos y administrativos, descritos por medio de un conjunto de datos específicos (metadatos) con el objetivo de recopilar, catalogar, gestionar, acceder, difundir y preservar tales ítems sin restricciones legales, técnicas y de acceso (Texier, 2013). Gracias a la interoperabilidad se comunican entre sí y pueden ser cosechados o leídos por otros sistemas informáticos conocidos como agregadores (el más conocido Google Scholar). Los protocolos de interoperabilidad más relevantes son: OAI-PMH, ResourceSync, SWORD, Signposting, etc., y todos siguen los lineamientos internacionales que recomiendan OpenAire, La Referencia, COAR (2020), entre otros.

La recuperación, transformación y visualización se pueden pensar como los procesos de la datificación (Lope Salvador et al., 2020; Markus, 2017; Mayer-Schönberger & Cukier, 2013). Así, el proceso de recuperación consistirá en obtener datos confiables, ya sea de medios y/o infraestructuras específicas o a través de la aplicación de patrones usando algoritmos y/o softwares existentes. Por su parte, el proceso de transformación tiene como objetivo realizar la limpieza y

consolidar los datos entrantes para contar con datos precisos, correctos, consistentes y sin ambigüedades. Y, por último, el proceso de visualización se entiende como la monitorización de los datos producidos por el objeto de la datificación.

Esta visualización puede implementarse en los sistemas de software de los Repositorios Institucionales o en aplicaciones externas que usen tales repositorios, gracias a las diferentes técnicas y teorías computacionales como la interoperabilidad. En consonancia con la interoperabilidad de los RI y relacionándolo con el otro proceso de datificación llamado la transformación, se puede afirmar que ésta permitirá la disponibilidad de los diversos recursos desde cualquier sistema de software y, por ende, la transformación se aplicaría en texto completo y/o en otros metadatos (título, palabras clave o resumen) siguiendo la guía de principios de calidad, FAIR Data de la Unión Europea (European Commission, 2018).

Una técnica computacional para la transformación sería el Procesamiento de Lenguaje Natural, que se entiende como una interrelación entre las ciencias de la computación, inteligencia artificial y lingüística (Bird et al., 2009; Jurafsky & Martin, 2008; Rivera et al., 2014; Sun et al., 2017; Virmani et al., 2017). Los diversos enfoques del PLN son (Acosta et al., 2015; Lossio-Ventura et al., 2014): i) lingüístico, ii) estadístico, iii) aprendizaje automático, y iv) métodos híbridos. Algunos de los temas más representativos del PLN aplicados sobre la base de los enfoques antes mencionados son (Vallez & Pedraza, 2007): la extracción de información, la generación automática de resúmenes, la búsqueda de respuestas, la recuperación de información monolingüe y multilingüe, técnicas automáticas de clasificación de textos temáticos y no temáticos, identificación de perfiles de usuarios, análisis de sentimientos.

Asimismo, existen diversas técnicas para el PLN tales como: marcado del discurso (*Part-of-speech - POS*) que consiste en reconocer las entidades nombradas y extracción de información (Ye et al., 2016), identificación de la raíz léxica (o *stemmer*) de un conjunto de palabras similares conocida como lematización (Texier et al., 2017), proceso de truncado (en inglés es *Chunker*) cuya finalidad es aproximar lo más posible las palabras a su raíz léxica (Stamatatos et al., 2000), *matching* con patrones semánticos usando el aprendizaje de *bootstrapping* (Gupta & Manning, 2011), *machine learning* o aprendizaje automático (Alsaffar et al., 2020; Ganegedara, 2018), identificación de entidades nombradas o *Named Entity Recognition - NER* (Ritter et al., 2011), *parse tree* (López et al., 2014), identificación de relaciones retóricas a partir *Rhetorical Structure Theory* (Mann & Thompson, 2009), etc.

Las herramientas que realizan PLN tienen limitaciones y ventajas, en todo caso, se abreviarán algunas (en los lenguajes de programación Java y Python) que sirven de base para este y futuros trabajos (Sun et al., 2017; Texier et al., 2017; Virmani et al., 2017): AnCora-CA, AnCora-ES, TreeBank, RST Spanish, Stanford Tokenizer, Stanford CoreNLP, OpenNLP Tokenizer, NLTK, TreeTagger, Gensim, FudanNLP, Language Technology Platform, NiuParser, MALLET, servicios en la nube por parte de IBM, Google, Microsoft, etc.

Luego de realizar el proceso de recuperación de recursos desde los diferentes RI y aplicar el PLN en los metadatos para transformarlos (según una necesidad específica) y visualizarlos dentro de un contexto, se determinará la necesidad para que tales recursos se relacionen (aplicando PLN) con los ODS (Allen et al., 2018; Felipe et al., 2020; Martínez Osés, 2020). En este sentido, se hallaron algunas publicaciones que toman el análisis de términos en relación con los ODS, pero

ninguna hace procesamiento del texto en repositorios institucionales o bibliotecas digitales ni hacen análisis de textos desde el punto de vista semántico (Buttigieg et al., 2015; Galli et al., 2020; Koch & Krellenberg, 2018; Körfggen et al., 2018, 2019; Meschede, 2020; Molina Gómez et al., 2019). La ONU elaboró una aclaración de las terminologías (Buttigieg et al., 2015) que puede servir de base para el plan de desarrollo propuesto al igual que el trabajo de Molina Gómez et al. (2019), que aplican la minería de texto sobre artículos de prensa en internet. El trabajo de Galli et al. (2020) hace un análisis por palabras clave de los ODS en el área de la medicina y se relaciona con el trabajo de Molina Gómez et al. y el proyecto de las Universidades Austríacas sobre los ODS conocido como UniNEtZ (en alemán *Universitäten und Nachhaltige Entwicklungsziele* y en inglés *Universities and Sustainable Development Goals*) reflejado en los trabajos de Koch & Krellenberg, 2018; Körfggen et al., 2018 y Meschede, 2020. Ese proyecto (UniNEtZ) hace un análisis por palabras claves ODS de su sistema contra otro en inglés y alemán. Aun así, el trabajo más relevante encontrado que sirve de base a este plan es el de Molina Gómez et al. (2019), ya que permitirá la construcción de una taxonomía ODS para tener un vocabulario controlado en la catalogación de los recursos existentes y por depositar al RI (Texier, 2015).

Si bien existen algunos estudios que toman los ODS en función del análisis de términos, ninguno, relevado hasta ahora, hace procesamiento de textos en RI o bibliotecas digitales ni consideran ese análisis bajo el punto de vista semántico (Allen et al., 2018; Buttigieg et al., 2015; Galli et al., 2020; Koch & Krellenberg, 2018; Körfggen et al., 2018, 2019; Felipe et al., 2020; Martínez Osés, 2020; Meschede, 2020; Molina Gómez et al., 2019). En síntesis, hay una área de vacancia que se centra en el Procesamiento de Lenguaje Natural orientando a problemas de la sociedad con un impacto en las bases de las Ciencias de la Computación y las Ciencias de la Información y Documentación.

En la siguiente Tabla se puede observar el plan de trabajo estructurado en tres fases sobre la base de las actividades y contribuciones:

Fases	Actividades	Contribuciones
<p style="text-align: center;">I</p> <p><i>A corto plazo, sentar las bases de la línea (exploratorio o y descriptivo)</i></p>	<p>Objetivo 1: a. Revisión bibliográfica sobre el análisis de texto en los ODS; b. Relevamiento de artículos del uso de repositorios de acuerdo con temas específicos; c. Implementación de una base de datos para sistematizar el proceso.</p> <p>Objetivo 2: a. Revisión y análisis de los diferentes software de RI existentes; b. Estudio de los diferentes enfoques y técnicas de PLN desde lo sintáctico y semántico; c. Selección del enfoque y técnica de PLN que se aplicará.</p> <p>Objetivo 3: a. Relevamiento de recursos existentes y compatibles con los ODS; b. Estudio de los tipos de textos de los recursos de los RI existentes.</p> <p>Objetivo 4: a. Estudio de los protocolos de interoperabilidad de RI existentes; b. Estructurar nuevas propuestas y mejoras de protocolos de interoperabilidad existentes.</p> <p>Objetivo 5: a. Análisis de los ODS del estado Chileno para el período 2015-2020; b. Análisis de los ODS argentinos actuales; c. Análisis de los ODS de los países vecinos.</p>	<p>1. Publicaciones sobre la relación entre ODS y los RI sobre la base de la datificación. 2. Base de datos y Wiki del relevamiento a realizar, junto con historias de usuario y arquitectura de software. 3. Metodología de trabajo. 4. Ejercicios y enfoques/paradigmas nuevos en asignaturas. 5. Análisis comparativo de los diferentes ODS: Argentina y de los países del cono sur.</p> <p>Indicadores de evaluación: número de publicaciones académicas y científicas según el paradigma de Ciencia Abierta, cantidad de recursos y tipos de textos disponibles e identificados en los repositorios, enfoque PLN aplicado, Wiki, cantidad de: historias de usuario, ejercicios y enfoques nuevos.</p>

<p style="text-align: center;">II</p> <p><i>A mediano plazo, componente de software</i></p>	<p>Objetivo 6: a. Descripción de los ODS en el contexto de la UNdeC; b. Discusión de la utilidad y del trabajo realizado</p> <p>Objetivo 7: a. Análisis de requerimientos de las necesidades de software; b. Diseño de la arquitectura de software deseada; c. Implementación del RI de la UNdeC y/o interacción con RIs; d. Recolección/cosecha de recursos (ítems) con filiación UNdeC; e. Análisis FODA del sistema deseado.</p> <p>Objetivo 8: a. Diseño del sistema (componente del repositorio o agnóstico al mismo); b. Documentación del sistema; c. Codificación del sistema; d. Casos de prueba del sistema; e. Instalación del sistema en servidores de pruebas y producción.</p>	<p>1. Incorporación de nuevos tópicos sobre: programación, estructuras de datos, base de datos, aprendizaje de máquina, trabajo de título, etc. 2. Validar el sistema de software con las diferentes dependencias de investigación de la UNdeC. 3. Clasificador de recursos (software) y posible RI de UNdeC.</p> <p>Indicadores de evaluación: líneas de código fuente (GitHub), FODA, horas hombre, asignaturas con contenidos nuevos, RI y/o recursos UNdeC.</p>
<p style="text-align: center;">III</p> <p><i>A largo plazo, lograr transferencia</i></p>	<p>Objetivo 9: a. Capacitación del grupo de trabajo; b. Incorporación de nuevos miembros del grupo de trabajo (investigadores, estudiantes y becarios); c. Formación de recursos humanos relacionados con la línea de investigación; d. Utilización del software en las áreas de investigación, administrativas y académicas; e. Identificación de investigadores que trabajen en áreas de conocimiento similares; f. Integración del plan/línea con otros miembros del instituto de Ingeniería y otros institutos de UNdeC; g. Consolidación de convenios con instituciones y empresas.</p> <p>Objetivo 10: a. Difusión y divulgación de todo el proceso, de los resultados y del conocimiento logrado; b. Diseño de un plan de desarrollo de la propuesta para ser implementado en otros contextos; c. Difusión y creación de nuevas estrategias sobre el uso de los RI por parte de la sociedad civil; d. Aplicación del conocimiento adquirido al nivel educativo básico, medio y superior.</p>	<p>1. Consolidación de grupos de investigación. 2. Validación de los usuarios finales. 3. Formalización de redes de trabajo externas a la UNdeC. 4. Plan de transferencia del conocimiento logrado y de ser posible a todos los niveles de educación. 5. Cursos, talleres, seminarios y/o artículos para difundir y divulgar.</p> <p>Indicadores de evaluación: convenios marcos con instituciones, número de recursos humanos capacitados y formados, cantidad de grupos de investigación, estudiantes y becarios, número de cursos/talleres/seminarios realizados, planes de transferencia, publicaciones en WoS y Scopus.</p>

Referencias bibliográficas

- Acosta, O. L., Aguilar, C. A., & Infante, T. (2015). Reconocimiento de términos en español mediante la aplicación de un enfoque de comparación entre corpus. *Linguamática*, 7(2), 19-34.
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Languages in Social Media*, 30–38.
- Allen, C., Metternicht, G., & Wiedmann, T. (2018). Initial progress in implementing the Sustainable Development Goals (SDGs): A review of evidence from countries. *Sustainability Science*, 13(5), 1453–1467.
- Alsaffar, D., Alfahhad, A., Alqhtani, B., Alamri, L., Alansari, S., Alqahtani, N., & Alboaneen, D. A. (2020). Machine and Deep Learning Algorithms for Twitter Spam Detection. En A. E. Hassanien, K. Shaalan, & M. F. Tolba (Eds.), *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019* (pp. 483-491). Springer International Publishing. https://doi.org/10.1007/978-3-030-31129-2_44
- Ambyssoft Inc. *Site Map*. (2020). <http://www.ambyssoft.com/siteMap.html>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.
- Buttigieg, P. L., McGlade, J., & Coppens, L. (2015). *Clarifying terms in the SDGs: Representing the meaning behind the terminology* [Miscellaneous]. 2nd Meeting of the Inter-Agency Expert Group on Sustainable Development Goal Indicators; United Nations Statistics Division.
- COAR. (2020). *Next Generation Repositories*. Confederation of Open Access Repositories. <https://www.coar-repositories.org/news-updates/what-we-do/next-generation-repositories/>

- European Commission. (2018). *Turning FAIR into reality—Publications Office of the EU*.
<https://doi.org/10.2777/1524>
- Felipe, C., Valbuena, N., & others. (2020). Índice ODS 2019 para América Latina y el Caribe. *Centro de los Objetivos de Desarrollo Sostenible para América Latina y el Caribe*.
<https://doi.org/hdl.handle.net/11520/25484>
- Galli, A., Leuenberger, A., Dietler, D., Fletcher, H. A., Junghanss, T., & Utzinger, J. (2020). Tropical Medicine and International Health and the 2030 Agenda for Sustainable Development. *Tropical Medicine & International Health*, 25(1), e1-e13. <https://doi.org/10.1111/tmi.13368>
- Ganegedara, T. (2018). *Natural Language Processing with TensorFlow: Teach language to machines using Python's deep learning library*. Packt Publishing Ltd.
- Griffiths, M. D., Kuss, D. J., Billieux, J., & Pontes, H. M. (2016). The evolution of Internet addiction: A global perspective. *Addictive behaviors*, 53, 193–195.
- Gupta, S., & Manning, C. D. (2011). Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers. *IJCNLP*.
- Hurtado, J. (2008). *Cómo formular objetivos de investigación* (2da ed.). Fundación Sygal.
- Jisc. (2020). OpenDOAR: Directory of Open Access Repositories. <https://v2.sherpa.ac.uk/opensoar/>
- Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall, 1039.
- Karisma, F. (2018). Declaración de Panamá sobre Ciencia Abierta.
<https://web.karisma.org.co/DeclaracionDePanama/>
- Kramer, B., & Bosman, J. (2016). Innovations in scholarly communication—Global survey on research tool usage. *F1000Research*, 5. <https://doi.org/10.12688/f1000research.8414.1>
- Koch, F., & Krellenberg, K. (2018). How to Contextualize SDG 11? Looking at Indicators for Sustainable Urban Development in Germany. *ISPRS International Journal of Geo-Information*, 7(12), 464.
<https://doi.org/10.3390/ijgi7120464>
- Körffgen, A., Förster, K., Glatz, I., Maier, S., Becsi, B., Meyer, A., Kromp-Kolb, H., & Stötter, J. (2018). It's a Hit! Mapping Austrian Research Contributions to the Sustainable Development Goals. *Sustainability*, 10(9), 3295. <https://doi.org/10.3390/su10093295>
- Körffgen, A., Glatz, I., Maier, S., Scherz, S., Kreiner, H., Passer, A., Allerberger, F., Kromp-Kolb, H., & Stötter, J. (2019). Austrian Universities and the Sustainable Development Goals. *IOP Conference Series: Earth and Environmental Science*, 323, 012156.
<https://doi.org/10.1088/1755-1315/323/1/012156>
- Lope Salvador, V., Vidal Bordes, J., & Mamaqi, X. (2020). La Inteligencia Artificial: Desafíos teóricos, formativos y comunicativos de la datificación. *ICONO*, 18(1). <https://doi.org/ri14.v18i1.1434>
- López, J., Sánchez-Sánchez, C., & Villatoro-Tello, E. (2014). *Laboratorio en línea para el procesamiento automático de documentos*.
- Lossio-Ventura, J. A., Jonquet, C., Roche, M., & Teisseire, M. (2014). Yet Another Ranking Function for Automatic Multiword Term Extraction. *Advances in Natural Language Processing*, 52-64.
https://doi.org/10.1007/978-3-319-10888-9_6
- Mann, W., & Thompson, S. (2009). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3), 243–281.
<https://doi.org/10.1515/text.1.1988.8.3.243>
- Marín, Á. L. G., & Lisbona, C. B. (2018). Datificación en los Archivos Digitales de los Medios de Comunicación: Nuevos Retos. *Investigaciones en datificación de la era digital*.
- Markus, M. L. (2017). Datification, organizational strategy, and is research: What's the score? *The Journal of Strategic Information Systems*, 26(3), 233–241.
- Martínez Osés, P. J. (2020). Hacer realidad la Agenda: Medios de implementación, revisión y seguimiento. *Transformar nuestro mundo, ¿realidad o ficción?*
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Meschede, C. (2020). The Sustainable Development Goals in Scientific Literature: A Bibliometric Overview at the Meta-Level. *Sustainability*, 12(11), 4461. <https://doi.org/10.3390/su12114461>

- Molina Gómez, N. I., Rodríguez, C. A., López, P. A., & Díaz Arévalo, J. L. (2019). Minería de texto y aprendizaje automático para identificar prioridades de desarrollo sostenible. *Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASP)*.
<https://doi.org/10.1109/CASAP.2019.8916682>
- Naciones Unidas. (2015). *Resolución de la Asamblea General de las Naciones Unidas: A/RES/70/1*. ONU.
<https://undocs.org/es/A/RES/70/1>
- Rajaraman, A., & Ullman, J. (2011). *Mining of Massive Datasets | Guide books*. Cambridge University Press. <https://dl.acm.org/doi/book/10.5555/2124405>
- Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011). Named Entity Recognition in Tweets: An Experimental Study. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1524–1534. <http://dl.acm.org/citation.cfm?id=2145432.2145595>
- Rivera, S. J., Minsker, B. S., Work, D. B., & Roth, D. (2014). A text mining framework for advancing sustainability indicators. *Environmental Modelling & Software*, 62, 128-138.
<https://doi.org/10.1016/j.envsoft.2014.08.016>
- Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2000). A Practical Chunker for Unrestricted Text. En D. N. Christodoulakis (Ed.), *Natural Language Processing—NLP 2000* (pp. 139-150). Springer.
https://doi.org/10.1007/3-540-45154-4_13
- Suber, P. (2012). Ensuring open access for publicly funded research. *BMJ : British Medical Journal*, 345.
<https://doi.org/10.1136/bmj.e5184>
- Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 36, 10-25. <https://doi.org/10.1016/j.inffus.2016.10.004>
- Tascón, M. (2013). Introducción: Big Data. Pasado, presente y futuro. *Telos: Cuadernos de comunicación e innovación*, 95, 47-50.
- Texier, J. (2015). *La representación de recursos usando la metodología del Desarrollo Dirigido por Modelos en un repositorio institucional* [Tesis, Facultad de Informática].
<http://hdl.handle.net/10915/45635>
- Texier, J. (2013). *Los repositorios institucionales y las bibliotecas digitales: Una somera revisión bibliográfica y su relación en la educación superior*. 9. <http://eprints.rclis.org/19925/>
- Texier, J., & Zambrano, J. (2019). *La relación entre el Curriculum DL y las Ciencias de la Computación: Una revisión bibliográfica*. Interciencia, 45 número 2.
https://www.interciencia.net/wp-content/uploads/2020/03/01_6205_Com_Texier_v45n2_8.pdf
- Texier, J., Zambrano, J., & Frati, F. E. (2017, octubre). *Framework para el procesamiento lingüístico de artículos científicos*. XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017).
<http://sedici.unlp.edu.ar/handle/10915/63711>
- UNESCO. (2020, febrero 17). Open Science. UNESCO.
<https://en.unesco.org/science-sustainable-future/open-science>
- Vallez, M., & Pedraza, R. (2007). El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines. *Hipertext.Net*. <http://elis.da.ulcc.ac.uk/9973/>
- Vicente-Saez, R., & Martínez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428-436.
<https://doi.org/10.1016/j.jbusres.2017.12.043>
- Virmani, C., Pillai, A., & Juneja, D. (2017). Extracting Information from Social Network using NLP. *International Journal of Computational Intelligence Research*, 13(4), 621–630.
- Ye, D., Xing, Z., Li, J., & Kapre, N. (2016). Software-specific Part-of-speech Tagging: An Experimental Study on Stack Overflow. *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, 1378–1385. <https://doi.org/10.1145/2851613.2851772>
- Yu, Y., & Wang, X. (2015). World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans' tweets. *Computers in Human Behavior*, 48, 392–400.