Institutional Repositories as a tool for monitoring and verifying the Sustainable Development Goals. Case study: Argentina

Jose Texier
Universidad Nacional de Chilecito
jtexier@undec.edu.ar
ORCID: 0000-0003-0176-6625; GScholar; ScopusID; PublonsID

Abstract: The general objective of the system is the construction of an automatic classifier of resources of the National System of Digital Repositories in accordance with the 169 goals of the 17 Sustainable Development Goals (SDG) defined by the United Nations Organization. The classifier will consist of a computational development capable of analyzing different metadata (call it title, abstract, keywords and full text) applying text analysis techniques to establish a syntactic and/or semantic relationship with the ODS, which allows the end user to do use of the resources or items of the repository or repositories according to the classification achieved, that is, to date the institutional repositories. This line of research promotes a culture of sharing resources, linking people in similar academic and scientific research and activities, contributing to the generation of new knowledge and reducing the gaps that make it difficult to access information on topics related to the SDGs through the items deposited in Argentine Institutional Repositories. The work plan includes both basic research aspects (improvement and analysis of text processing algorithms), applied research (direct benefit of society on the subject of the SDGs) and field research (collection and availability of digital resources for all the repositories of Argentina for a specific purpose, researcher base research area), as well as knowledge transfer (the project could be integrated as a software component in the different types of existing repository software or as a software system of independently and external to IR).
Keyworks: natural language processing, SDG, RI, datification, Argentina

The development of the system is based on a line of work whose axis is Natural Language Processing (NLP) that uses different approaches applicable locally, regionally and internationally. It involves a line of study that includes basic, applied and field research, as well as knowledge transfer because they contribute to: the improvement and analysis of word processing algorithms in connection with the Sustainable Development Goals (SDG); the application at different educational levels; the collection and availability of resources from the Institutional Repositories of Argentina; the design and implementation of a software as part of the repository and / or as an external system and the articulation with teaching and science. The SDGs are taken as the axis of work because they represent a universal agenda for sustainable development, where the adopting countries must follow a global strategy that combines economic development, social inclusion and environmental sustainability. In the United Nations General Assembly, on September 25, 2015, the 17 SDGs with 169 targets and 232 indicators were approved and declared as universal, indivisible and interdependent (United Nations, 2015).

In recent times, the growth of Argentine Institutional Repositories (IR) has been observed in the international directory of OpenDOAR repositories (Jisc, 2020). The National University of Chilecito (UNdeC) does not have an IR (December 2020), however, that would not be an obstacle for this proposal since it could be implemented in a short period of time and / or its production would be harvested through existing tools. For this reason, a proposal is proposed (with the Argentine IRs and / or the resources of the UNdeC) that is based on the following hypothesis: Could it be that the Argentine Institutional Repositories can contribute added value in the implementation, verification and monitoring of the Sustainable Development Goals of Argentina? In this sense, the general objective is to implement an automatic classifier of resources from the Argentine Institutional Repositories in accordance with the 169 goals of the 17 Sustainable Development Goals defined by the United Nations (UN).

The classifier will function as part of the UNdeC RI or as a repository agnostic technology. It will consist of a novel computational development capable of analyzing different metadata (title,

abstract, keywords and full text) applying text analysis techniques to establish a syntactic and / or semantic relationship with the ODS, which allows the end user to make use of resources or items thanks to the classification achieved and datafication processes. The proposal will be implemented with the Open Science approach (Karisma, 2018; Kramer & Bosman, 2016; UNESCO, 2020; Vicente-Saez & Martinez-Fuentes, 2018) and will be developed under the projective modality (Hurtado, 2008) because it implies the creation and design of a software system. This software will be designed and built with the Agile Unified Process methodology (Ambysoft, 2020).

The specific objectives proposed are:

1. Describe the need to relate the various existing resources in the Argentine IRs according to the SDGs.
2. Analyze the different proposals for identifying resources in the repositories through the different metadata (title, abstract, keywords and full text) of the resources.
3. Characterize the different types of resources and their relationships with the Sustainable Development Goals.
4. Identify the interoperability processes that allow the syntactic and semantic relationship of resources with other repositories / systems in accordance with the Sustainable Development Goals.
5. Anticipate trends in SDG resources in Institutional Repositories in accordance with State policies defined by the Chilean government and the region according to their needs.
6. Reach a greater understanding of the SDGs and their implementation in the local, regional and national context.
7. Design a software architecture that guarantees scalability in the needs and in the adaptation of the relationships between the resources of the repositories and the ODS, within the IR of the UNdeC or external to the repository.
8. Implement a software system that classifies the existing resources of the Institutional Repositories in accordance with the 169 goals corresponding to the 17 Sustainable Development Goals.
9. Generate multidisciplinary networks of researchers, students and scholarship holders related to the line and / or new projects related to the environment (University and region) on the basis of datification and text analysis.
10. Evaluate the functioning of the software developed and its application in teaching to improve and create other innovative proposals that generate new knowledge.

This project lays its foundations on the relationship of Computer Science and Information and Documentation Sciences. The current challenge is not only in the quantity and quality of IR data, but in what should be done with all of them and how they will be managed and integrated with large volumes of structured, semi-structured and unstructured digital information, created by people and / or organisms in all areas of society (Griffiths et al., 2016; Suber, 2012). Therefore, datification is necessary to be able to transform such IR data into information that can be used in any area of knowledge (Lope Salvador et al. 2020; Markus, 2017).

The concept of datification has been developed in academic and scientific publications for more than five years, but its processes (retrieval, transformation and visualization) have been working independently for much longer and it tends to be confused with the concept of Big Data (Lope Salvador et al., 2020; Markus, 2017) that emerges as a technology that allows to process large amounts of data to collect, analyze and use them, (Cf. Rajaraman & Ullman, 2011), especially

because the processing and subsequent analysis of the data and / or information, in many cases, must be carried out in real time to influence timely decision-making.

Now, how the collection, analysis and use of data should be approached is also inherent to datification, therefore, it is understood that Big Data and datification are linked. However, the latter has to do with the action of converting and measuring information, a process or an activity into data thanks to the existing technological infrastructure (Agarwal et al., 2011; Markus, 2017; Mayer-Schönberger & Cukier, 2013 ; Tascón, 2013; Yu & Wang, 2015). In turn, datification is related to digitization. Mayer and Cukier (2013) state that dataification occurred before Information Technologies, a fact that allowed the emergence of digitization. These authors explain that datification refers to the data transformed, through analysis and reorganization, into information that can be used in any area of knowledge or discipline. Instead, digitization (Mayer-Schönberger & Cukier, 2013) is the process by which analog information is converted into digital so that it is legible and processed by a computer. In summary, with the datification, the data can be quantified to be tabulated and analyzed, then, they are stored in different formats to apply Big Data (Marín & Lisbona, 2018), as long as the real time component exists. In other words, data is to analyze before collecting data and this is achieved with education and research and not by a stroke of luck or the simple use of a tool.

This work plan takes as a basis the data and information that, in an orderly and formal way, are found in the Institutional Repositories and other bibliographic sources to transform it (text analysis) into useful information linked to the SDGs. IRs are interoperable computer systems (DSpace, EPrints, Fedora Commons, Greenstone, CKAN, Junar, etc.) that host scientific, academic and administrative resources, described through a set of specific data (metadata) with the objective of collecting, catalog, manage, access, disseminate and preserve such items without legal, technical and access restrictions (Texier, 2013). Thanks to interoperability they communicate with each other and can be harvested or read by other computer systems known as aggregators (the better known Google Scholar). The most relevant interoperability protocols are: OAI-PMH, ResourceSync, SWORD, Signposting, etc., and all follow the international guidelines recommended by OpenAire, La Reference, COAR (2020), among others.

Recovery, transformation, and visualization can be thought of as datification processes (Lope Salvador et al., 2020; Markus, 2017; Mayer-Schönberger & Cukier, 2013). Thus, the recovery process will consist of obtaining reliable data, either from specific media and / or infrastructures or through the application of patterns using existing algorithms and / or software. For its part, the transformation process aims to clean and consolidate incoming data to have accurate, correct, consistent and unambiguous data. And finally, the visualization process is understood as the monitoring of the data produced by the object of datification.

This visualization can be implemented in the software systems of the Institutional Repositories or in external applications that use such repositories, thanks to the different techniques and computational theories such as interoperability. In line with the interoperability of the IRs and relating it to the other datification process called the transformation, it can be stated that this will allow the availability of the various resources from any software system and, therefore, the transformation would be applied in full text and / or in other metadata (title, keywords or abstract) following the quality principles guide, FAIR Data of the European Union (European Commission, 2018).

A computational technique for the transformation would be Natural Language Processing, which is understood as an interrelation between computer science, artificial intelligence and linguistics (Bird et al., 2009; Jurafsky & Martin, 2008; Rivera et al., 2014 ; Sun et al., 2017; Virmani et al., 2017). The various approaches to the PLN are (Acosta et al., 2015; Lossio-Ventura et al., 2014): i) linguistic, ii) statistical, iii) machine learning, and iv) hybrid methods. Some of the most representative topics of the NLP applied on the basis of the aforementioned approaches are (Vallez & Pedraza, 2007): information extraction, automatic generation of summaries, search for answers, retrieval of monolingual and multilingual information, automatic techniques for classifying thematic and non-thematic texts, identification of user profiles, sentiment analysis.

Likewise, there are various techniques for the PLN such as: speech marking (Part-of-speech - POS) which consists of recognizing the named entities and extracting information (Ye et al., 2016), identification of the lexical root (or stemmer) of a set of similar words known as lemmatization (Texier et al., 2017), truncation process (in English it is Chunker) whose purpose is to approximate the words as closely as possible to their lexical root (Stamatatos et al., 2000) , matching with semantic patterns using bootstrapping learning (Gupta & Manning, 2011), machine learning or automatic learning (Alsaffar et al., 2020; Ganegedara, 2018), identification of named entities or Named Entity Recognition - NER (Ritter et al. ., 2011), parse tree (López et al., 2014), identification of rhetorical relationships from Rhetorical Structure Theory (Mann & Thompson, 2009), etc.

The tools that perform NLP have limitations and advantages, in any case, some (in the Java and Python programming languages) that serve as the basis for this and future work (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 cloud by IBM, Google, Microsoft, etc.

After carrying out the process of recovering resources from the different IRs and applying the PLN in the metadata to transform them (according to a specific need) and visualize them within a context, the need for such resources to be related (applying PLN) to the SDGs (Allen et al., 2018; Felipe et al., 2020; Martínez Osés, 2020). In this sense, some publications were found that take the analysis of terms in relation to the SDGs, but none do text processing in institutional repositories or digital libraries or do text analysis from the semantic point of view (Buttigieg et al., 2015 ; Galli et al., 2020; Koch & Krellenberg, 2018; Körfgen et al., 2018, 2019; Meschede, 2020; Molina Gómez et al., 2019). The UN prepared a clarification of terminologies (Buttigieg et al., 2015) that can serve as the basis for the proposed development plan, as can the work of Molina Gómez et al. (2019), which apply text mining on press articles on the internet. The work of Galli et at. (2020) makes an analysis by keywords of the SDGs in the area of medicine and is related to the work of Molina Gómez et al. and the Austrian Universities project on the SDGs known as UniNEtZ (in German Universitäten und Nachhaltige Entwicklungsziele and in English Universities and Sustainable Development Goals) reflected in the works of Koch & Krellenberg, 2018; Körfgen et al., 2018 and Meschede, 2020. This project (UniNEtZ) makes an analysis by ODS keywords of its system against another in English and German. Even so, the most relevant work found that serves as the basis for this plan is that of Molina Gómez et al. (2019), since it will allow the construction of an ODS taxonomy to have a controlled vocabulary in the cataloging of existing resources and to be

deposited with the IR (Texier, 2015).

Although there are some studies that take the SDGs based on the analysis of terms, none, surveyed so far, do word processing in IR or digital libraries or consider this analysis from the semantic point of view (Allen et al., 2018; Buttigieg et al., 2015; Galli et al., 2020; Koch & Krellenberg, 2018; Körfgen et al., 2018, 2019; Felipe et al., 2020; Martínez Osés, 2020; Meschede, 2020; Molina Gómez et al., 2019). In summary, there is a vacancy area that focuses on Natural Language Processing, orienting to problems of society with an impact on the foundations of Computer Science and Information and Documentation Sciences.

The following Table shows the work plan structured in three phases based on the activities and contributions:

Fases	Actividades	Contribuciones
I *In the short term, lay the foundations of the line (exploratory and descriptive)*	**Objective 1:** a. Bibliographic review on text analysis in the SDGs; b. Survey of articles on the use of repositories according to specific topics; c. Implementation of a database to systematize the process. **Objective 2:** a. Review and analysis of the different existing IR software; b. Study of the different approaches and techniques of NLP from the syntactic and semantic; c. Selection of the NLP approach and technique to be applied. **Objective 3:** a. Survey of existing resources compatible with the SDGs; b. Study of the types of texts of the existing IR resources. **Objective 4:** a. Study of existing IN interoperability protocols; b. Structure new proposals and improvements to existing interoperability protocols. **Objective 5:** a. Analysis of the SDGs of the Chilean state for the period 2015-2020; b. Analysis of the current Argentine SDGs; c. Analysis of the SDGs of neighboring countries.	1. Publications on the relationship between ODS and IRs based on datification. 2. Database and Wiki of the survey to be carried out, together with user stories and software architecture. 3. Work methodology. 4. Exercises and new approaches / paradigms in subjects. 5. Comparative analysis of the different SDGs: Argentina and the countries of the southern cone. **Evaluation indicators:** number of academic and scientific publications according to the Open Science paradigm, number of resources and types of texts available and identified in the repositories, applied PLN approach, Wiki, number of: user stories, exercises and new approaches.
II *Medium term, software component*	**Objective 6:** a. Description of the SDGs in the context of the UNdeC; b. Discussion of the utility and the work done **Objective 7:** a. Requirements analysis of software needs; b. Design of the desired software architecture; c. Implementation of the RI of the UNdeC and / or interaction with RIs; d. Collection / harvest of resources (items) with UNdeC affiliation; and. SWOT analysis of the desired system. **Objective 8:** a. System design (component of the repository or agnostic to it); b. System documentation; c. System coding; d. System test cases; and. Installation of the system on test and production servers.	1. Incorporation of new topics on: programming, data structures, databases, machine learning, title work, etc. 2. Validate the software system with the different research units of the UNdeC. 3. Resource classifier (software) and possible UNdeC IR. **Evaluation indicators:** source code lines (GitHub), SWOT, man hours, subjects with new content, IR and / or UNdeC resources.

III *In the long term, achieve transfer*	**Objective 9:** a. Training of the working group; b. Incorporation of new members of the working group (researchers, students and scholarship holders); c. Training of human resources related to the line of research; d. Use of the software in the research, administrative and academic areas; and. Identification of researchers working in similar knowledge areas; F. Integration of the plan / line with other members of the Institute of Engineering and other institutes of UNdeC; g. Consolidation of agreements with institutions and companies. **Objective 10:** a. Dissemination and dissemination of the entire process, the results and the knowledge achieved; b. Design of a development plan for the proposal to be implemented in other contexts; c. Dissemination and creation of new strategies on the use of IR by civil society; d. Application of the knowledge acquired at the basic, intermediate and higher educational level.	1. Consolidation of research groups. 2. End user validation. 3. Formalization of work networks outside the UNdeC. 4. Knowledge transfer plan achieved and, if possible, at all levels of education. 5. Courses, workshops, seminars and / or articles to spread and disseminate. **Evaluation indicators:** framework agreements with institutions, number of human resources trained and trained, number of research groups, students and scholarship holders, number of courses / workshops / seminars carried out, transfer plans, publications in WoS and Scopus.

References

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

Ambysoft Inc. Site Map. (2020). http://www.ambysoft.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 Sypal.

Jisc. (2020). OpenDOAR: Directory of Open Access Repositories. https://v2.sherpa.ac.uk/opendoar/

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

Körfgen, 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

Körfgen, 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., & Martinez-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.