



Facultad de
**Información y
Comunicación**



**UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY**

Nombre del curso: <input type="text" value="Introducción a la programación y análisis de texto con R."/>		Año de elaboración del Programa: <input type="text" value="2023"/>
Nombre abreviado: <input type="text"/> (Será completado por Bedelía)	Código: <input type="text"/> (Será completado por Bedelía)	Estado: <input type="text" value="En Registro"/> Visible en simulador: <input type="checkbox"/>
Carrera: <input checked="" type="checkbox"/> Licenciatura en Comunicación <input checked="" type="checkbox"/> Licenciatura en Archivología <input checked="" type="checkbox"/> Licenciatura en Bibliotecología		Ciclo: <input type="text" value="Intermedio/de Profundización"/>
Curso ofrecido como electiva para otros Servicios Universitarios: <input type="text" value="Seleccionar"/> (Será completado por la Comisión de Carrera)	Obligatorio / Optativo: <input type="text" value="Optativo"/>	Créditos: <input type="text" value="5"/> (Será completado por la Comisión de Carrera)
Tipo Orientación: <input type="text" value="Seleccionar"/> Orientaciones Curriculares <input type="text"/> (Será completado por la Comisión de Carrera)		
Módulo(s): (Será completado por la Comisión de Carrera)	<input checked="" type="checkbox"/> <i>Lenguajes y Medios</i> <input type="checkbox"/> <i>Metodología</i> <input type="checkbox"/> <i>Profesional-Integral</i> <input type="checkbox"/> <i>Sociedad, Cultura y Políticas de Comunicación</i> <input type="checkbox"/> <i>Teoría y Análisis de la Comunicación</i> <input type="checkbox"/> <i>Fundamentos teórico-conceptuales y contexto de la disciplina y la profesión</i> <input type="checkbox"/> <i>Gestión documental y sistemas archivísticos</i> <input type="checkbox"/> <i>Descripción, análisis y recuperación de la información</i> <input type="checkbox"/> <i>Políticas públicas, planificación y gestión de unidades de información</i> <input type="checkbox"/> <i>Investigación en la Ciencia de la Información</i> <input type="checkbox"/> <i>Investigación en la Ciencia de la Información</i> <input type="checkbox"/> <i>Documentación digital y aplicación de las TIC</i> <input type="checkbox"/> <i>Disciplinas complementarias</i> <input type="checkbox"/> <i>Actividades integradoras</i> <input type="checkbox"/> <i>Opcionales / Investigación / Extensión</i> <input type="checkbox"/> <i>Alfabetización Académica</i> <input type="checkbox"/> <i>Fuentes, colecciones y servicios de información</i>	
Departamento, sección o área <input type="text"/>		Semestre/s: <input type="checkbox"/> 1ro: <input type="checkbox"/> 2do: <input type="checkbox"/> 3ro: <input checked="" type="checkbox"/> 4to: <input type="checkbox"/> 5to: <input type="checkbox"/> 6to: <input type="checkbox"/> 7mo: <input type="checkbox"/> 8vo:
Tipo de curso: <input type="text" value="Seleccionar"/> (Será completado por la Comisión de Carrera) Descargar Tipología de Cursos		Turno/s: <input checked="" type="checkbox"/> <i>Matutino</i> : <input type="checkbox"/> <i>Vespertino</i> : <input type="checkbox"/> <i>Nocturno</i> :
Docente/s responsable/s del curso <input type="text"/>		Fernando Filgueira (Fac.Ciencias Sociales) Elina Gómez (Fac.Ciencias Sociales)
A partir de aquí, la información deberá ser aportada por el/los docente/s responsable/s del curso		
Docente/s encargado/s del curso <input type="text"/> (pueden coincidir o no con el docente responsable)		Fernando Filgueira (Fac.Ciencias Sociales) Elina Gómez (Fac.Ciencias Sociales)

Otros integrantes del equipo docente

Forma de evaluación:

- Trabajos Prácticos (Individual)
 Trabajos Prácticos (Grupal)
 Informe (Individual)
 Informe (Grupal)
 Monografía (Individual)
 Monografía (Grupal)
- Producto u Obra (Individual)
 Producto u Obra (Grupal)
 Control de Lectura (Individual)
 Control de Lectura (Grupal)
 Examen Oral (Individual)
 Examen Oral (Grupal)

Régimen de asistencia:

Asistencia obligatoria (75%)

Horas semanales de aula:

2

Horas semanales de trabajo estudiantil (estimadas):

5

Instancias de evaluación:

- Parciales (Individual)
 Parciales (Grupal)
 Examen (Individual)
- Examen (Grupal)
 Trabajo de Pasaje de Curso (Individual)
 Trabajo de Pasaje de Curso (Grupa;)

Descargar Normativa del Plan referida a este asunto

Modalidad de cursado:

Presencial

Créditos incrementales (en caso de corresponder):

No

Explicite en forma más detallada la modalidad de evaluación elegida (sin límite)*:

Aprobación con Trabajo final individual según pauta entregada en clase y que da cuenta del manejo de los diferentes módulos del curso. Trabajos prácticos intermedios a resolver en clase.

Breve resumen de la actividad curricular (máximo 500 caracteres)*: restan:122

El curso busca que los/as estudiantes se familiaricen con el lenguaje de programación en R, dotándolos/as de las herramientas básicas para el manejo del programa estadístico, para luego hacer foco en la exploración de paquetes orientados a la construcción de corpus de datos textuales y análisis a partir de técnicas de minería de texto con sus correspondientes visualizaciones.

Objetivos del curso (sin límite)*:

(generales y específicos)

- 1. Lograr una familiarización el lenguaje de programación R y los conceptos fundamentales que esto implica (vectores, matrices, marcos de datos, listas).*
- 2. Promover que el estudiante pueda obtener datos textuales de diferentes fuentes directamente desde R.*
- 3. Contribuir a que cada estudiante se familiarice con los conceptos de corpus, matriz de términos, así como maneje técnicas vinculadas a la minería de texto.*
- 4. Indagar sobre herramientas para la visualización estática e interactiva de resultados.*
- 5. Introducir al Procesamiento de Lenguaje Natural (PLN) como disciplina que utiliza técnicas de Inteligencia Artificial para el procesamiento de texto o habla.*

Conocimientos previos recomendados (sin límite)*:

Programación en R básica (no obligatorio).

Metodología de trabajo (sin límite)*:

(detallar actividades previstas con los/as estudiantes)

El enfoque del curso es práctico acompañado con breves exposiciones teóricas para cada tema. En el aula se trabaja con estrategia de live-coding y ejercicios prácticos individuales.

Contenidos del curso (sin límite)*:

1. *Introducción a R: Base teórica: ciencias sociales computacionales. R como software libre y gratuito, interfaz gráfica RStudio. Uso de la ayuda y foros; trabajo en proyecto (Rproj); Paquetes y funciones. Operadores relacionales y lógicos; Clases de objetos: vectores, matrices, marcos de datos y listas.*
2. *Fuentes de datos: Contemplar diferentes fuentes de datos textuales. Recuperación de fuentes documentales (PDF e imágenes) con técnicas de OCR; Web scraping de prensa digital, parlamentaria, hermenotecas y otros recursos existentes en el entorno R; Establecer puntos de conexión con APIs para obtención de datos textuales.*
3. *Pre-procesamiento de datos textuales: Procesamiento de strings y abordaje de los requerimientos previos (limpieza y homogenización) para el análisis de textos. Exploración y manipulación básica de datos, contemplando las diferentes fuentes y codificaciones posibles.*
3. *Minería de textos: Noción de corpus, matriz de términos y sus posibilidades analíticas, desde lo más descriptivo a la aplicación de técnicas más complejas: asociación y agrupación de palabras, uso de diccionarios (manuales y automáticos), análisis de sentimiento, modelado.*
4. *Visualización: Exploración de las diferentes posibilidades gráficas de visualización de los resultados del análisis textual (nubes de palabras, frecuencias, grafos). Algunos ejemplos de visualización dinámica e interactiva (con Shiny) para corpus con grandes volúmenes de datos textuales.*
5. *Inteligencia Artificial: Introducción a las posibilidades y alcance de las técnicas de Procesamiento de Lenguaje Natural (PLN) como subdisciplina de la Inteligencia Artificial que busca resolver tareas vinculadas al procesamiento de texto o habla.*

Bibliografía o filmografía básica*:

(máximo 15 entradas)

- Abedin, Jaynal & Kishor Kumar Das 2015. *Data Manipulation with R (2nd Edition)*.
- Caro, J., Díaz-de la Fuente, S., Ahedo, V., Zurro, D., Madella, M., Galán, JM., Izquierdo, L., Santos, JI., y del Olmo, R. (2020). *Terra incógnita: Libro blanco sobre transdisciplinariedad y nuevas formas de investigación en el Sistema Español de Ciencia y Tecnología*.
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, P. P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M. & Helbing, D. (2012). «Manifesto of computational social science». *The European Physical Journal Special Topics* 214(1): 325-46. <http://link.springer.com/10.1140/epjst/e2012-01697-8>.
- Cioffi-Revilla, C. (2021). *The Scope of Computational Social Science. Handbook of Computational Social Science*. Routledge.
- Cioffi-Revilla, C. (2017). *Introduction to Computational Social Science: Principles and Applications. 2nd edition*. Cham, Switzerland: Springer.
- Lazer, D. et al (2020). "Computational Social Sciences: Obstacles and Opportunities". *Science*, Nº 369, 28 ago, pp. 1060-1062. <https://science.sciencemag.org/content/369/6507/1060>. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara T., King G., Macy M., Roy D., Van Alstyne, M. (2009). «SOCIAL SCIENCE: Computational Social Science». *Science* 323(5915): 721-23. <https://www.sciencemag.org/lookup/doi/10.1126/science>.
- Salganik, M (2018) *Bit by Bit: Social Research in the Digital Age*. Princeton Univ. Press.
- Benoit, Kenneth & Michael Laver. 2003. "Estimating Irish party policy positions using computer wordscoring." *Irish Political Studies* 18(1):97–107.
- Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S and Matsuo A 2018. "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*, 3 (30), pp. 774. doi: 10.21105/joss.00774.
- Hu, M. & B. Liu. 2004. Mining and summarizing customer reviews. In *proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, page 168-177,
- Huang, Ronggui. 2018. *RQDA: R-based Qualitative Data Analysis. R package version 0.3-1*. Klemmensen, Robert, Sara BinzerHobolt & Martin Ejnar Hansen. 2007. "Estimating policy positions using political texts: An evaluation of the Wordscores approach." *Electoral Studies* 26(4):746–755.
- Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. 2 ed. Thousand Oaks: Sage.
- Team, R. C. (2000). *Introducción a R. Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos*. CRAN. URL <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>.
- Wickham, H. y Golemund, G. (2017). *R for Data Science*. O'Reilly Media. URL <https://r4ds.had.co.nz/>.

Bibliografía o filmografía recomendada (sin límite)*: