UNIVERSIDAD DE COSTA RICA
Escuela de Filología, Lingüística y Literatura
Departamento de Lingüística

FL-1051 Procesamiento automatizado de textos en Filología y Lingüística

- Créditos: 3
- Horas: 3 horas
- Requisitos: -
- Correquisitos: -
- Ciclo: Segundo.
- Clasificación: Propio
- Modalidad: Virtual

1. Descripción general

En este curso se aborda la manipulación de grandes cantidades de texto en Filología y Lingüística como una tarea automatizable. Además, considerando las características propias del lenguaje humano, se resuelven problemas específicos del procesamiento de texto. También se establecen las bases para identificar las necesidades y soluciones en proyectos complejos que requieran tareas de programación en dichas disciplinas.

2. Objetivo general

Brindar las herramientas de programación necesarias para la resolución de tareas específicas de procesamiento automatizado de textos en Filología y Lingüística.

3. Objetivos específicos

Al final del curso, el estudiante será capaz de:

- Efectuar manipulaciones de textos por medio de un lenguaje de programación.
- Identificar tareas programables en Lingüística y Filología.
- Resolver problemas de programación básicos en Lingüística y Filología.
- Adaptar soluciones de programación de otras áreas temáticas a la Filología y la Lingüística.

4. Contenidos

1. Corpus lingüísticos
2. Consultas de corpus lingüísticos (*Expresiones regulares*)
3. Lenguajes de marcado (HTML y XML) y Hojas de estilo en cascada (CSS)
4. Introducción al lenguaje de programación Python

5. Aplicaciones en filología y lingüística
6. Perspectivas de desarrollo y recursos auxiliares

5. Metodología

- Clases magistrales.
- Prácticas de cómputo de los estudiantes.
- Discusión de los resultados de las investigaciones de los estudiantes.
- Uso de la plataforma Mediación Virtual para actividades, tareas y evaluaciones. Es posible inscribirse por medio de la siguiente dirección: https://mediacionvirtual.ucr.ac.cr.

6. Cronograma

Tema	Duración en semanas
Introducción en el procesamiento automatizado de textos	1
Corpus lingüísticos	3
Lenguajes de marcado (HTML y XML) y Hojas de estilo en cascada (CSS)	4
Python	4
Problemas de programación en Lingüística y Filología	4
Total	**16**

7. Evaluación

Descripción	Porcentajes (%)
Tareas	20%
Primer trabajo	35%
Trabajo final	45%
Total	100%

8. Bibliografía

- Acera García, M.A. (2012). *XML. Edición 2012.* Anaya Multimedia.
- Apache Software Foundation (2020). Apache HTTP (Versión 2.4.43) [Software computacional]. Estados Unidos. Disponible en https://httpd.apache.org
- Anthony, L. (2019). AntConc (Versión 3.5.8) [Software computacional]. Tokio, Japón: Waseda University. Disponible en https://www.laurenceanthony.net/software
- Baquero Velásquez, J. M. (2010). Lingüística computacional aplicada. Editorial Universidad Nacional de Colombia. https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/ereader/sibdi/127585?page=1

- Bernal Chávez, J. A. y Bernal Chávez, J. A. (2016). Introducción a la lingüística computacional. Ediciones de la U. https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/70290
- Boulanger, T. (2015). *XML Práctico: bases esenciales, conceptos y casos prácticos.* Barcelona: ENI Ediciones.
- Cuevas Álvarez, A. (2016). Python 3: curso práctico. RA-MA Editorial. https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/106404
- Gauchat, J. (2012). *El gran libro de HTML5, CSS3 y JavaScript.* Barcelona: Marcombo.
- Jiménez Millán, J. A. (2014). *Compiladores y procesadores de lenguajes*. Servicio de Publicaciones de la Universidad de Cádiz. https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/33847
- López, F., & Romero, V. (2014). *Mastering python regular expressions*. Recuperado de https://ebookcentral-proquest-com.ezproxy.sibdi.ucr.ac.cr
- Martelli, A. (2008). *Python: Guía de referencia.* Anaya Multimedia / O'Reilly. Madrid.
- Murphy, P y Persson, N. (2009). *HTML y CSS.* Anaya Multimedia.
- Pemberton, S. et al. (2000). *XHTML 1.0TM. El lenguaje de Etiquetado Hipertextual Extensible-REC-xml-20000126*. [Página web: http://home.worldonline.es/jlgranad/xhtml/xhtml1.htm].
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada No 46, pgs. 93–119.*
- *Parodi, G. (2010). Lingüística de corpus: de la teoría a la empiria. Editorial Iberoamericana / Vervuert. https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/36857*
- Pons, O. et al. (2008). *Introducción a los sistemas de bases de datos.* Madrid: Paraninfo.
- O'Keeffe, M., McCarthy (eds.) (2010). *The Routledge Handbook of Corpus Linguistics*. London: Routledge.
- Raya González, L. y Toharia Rabasco, J. P. (2015). Lenguajes de marcas y sistemas de gestión de la información. RA-MA Editorial. https://elibro-net.ezproxy.sibdi.ucr.ac.cr/es/lc/sibdi/titulos/62483
- Gerardo Sierra Martínez (2017). Introducción a los Corpus Lingüísticos. Universidad Nacional Autónoma de México-Instituto de Ingeniería, México.
- Silberschnatz, A. et al. (2006). *Fundamentos de bases de datos.* McGraw-Hill.
- The World Wide Web Consortium. *The World Wide Web Consortium.* [Página web: http://www.w3.org/].
- The World Wide Web Consortium. *W3Schools.* [Página web: https://www.w3schools.com/].