

NOMBRE ACTIVIDAD (Curso / Diploma / Seminario)

**Curso: Desarrollo y Programación de Visualización de Datos Geográficos en Internet**

Docente	:	Cristian L. Hernández Milla
Coordinación Académica	:	EC

Unidad académica / organismo que lo desarrolla	Educación Continua, DEX-FAU
Horas de trabajo presencial	24
Día / hora de realización	Martes y jueves. Desde las 19:00 horas hasta las 22:20 horas. (Primera versión: Octubre 2017)

Dirigido a (público destinatario)

El curso “Desarrollo y Programación de Visualización de Datos Geográficos”, está orientado a todo aquel profesional y técnico que busque expandir sus capacidades de recolección, análisis y visualización de datos en general, con énfasis en la representación de datos geográficos en la web.

En específico el curso se orienta a profesionales que deseen comunicar el resultado de su trabajo a través de Internet en la forma de *visualización de datos (datos espaciales como no-espaciales)*, tales como periodistas trabajando en unidades de periodismo investigativo, profesionales relacionados con la Ciudad y el territorio y la representación de sus fenómenos y procesos, diseñadores que busquen expandir a través de la programación sus habilidades de comunicar información, y/o científicos sociales interesados en humanidades digitales, o la representación de fenómenos sociales mediante nuevos medios.



## Presentación / Descriptor

El constante incremento de la cantidad de datos a nivel global ha permitido tener registro como nunca antes en la historia sobre fenómenos culturales y naturales. Según IBM (<https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>) dichos registros o datos superan hoy en día los 2.5 quintillones de bytes, más aun, el 90% de los datos actuales del mundo, han sido producidos durante los últimos dos años, en lo que se conoce como Big Data (grandes datos).

Acceder a tal magnitud de información y aprovechar sus ventajas implica el desarrollo de habilidades relacionadas con ciencias de la computación, en tanto el rescate y procesamiento de datos y su posterior análisis y visualización. Al conjunto de técnicas y métodos que nos permiten la comunicación e interacción con grandes volúmenes de datos, las llamamos “Ciencia de Datos”, la cual mezcla elementos de programación computacional y estadística.

Este curso pretende guiar a los alumnos(as) a través del proceso completo de ciencia de datos, utilizando bases de datos abiertas para el análisis de diversos fenómenos, y herramientas de programación basadas en el entorno Node.js (Javascript). Mediante el entrenamiento en casos reales de análisis, los participantes podrán aprender habilidades de programación para la recolección de datos, así como el manejo y limpieza de los mismos, realización de análisis estadísticos, y la realización de visualizaciones para comunicar resultados en Internet.

## Objetivos del programa

Objetivo General	:	<i>Aplicar metodologías de análisis de Ciencia de Datos, que implica trabajar todo el ciclo de información, desde la recolección y estructuración, hasta la visualización de datos en Internet, a través del estudio de casos prácticos.</i>
Objetivos Específicos	:	Los objetivos específicos del taller son: a) Conocer y comprender los métodos relacionados con la Ciencia de Datos. b) Desarrollar la habilidad de configurar un entorno de programación para el trabajo con datos diversos, a través de la utilización de Javascript y el ambiente Node.js. c) Desarrollar la habilidad de transformar datos en información susceptible de ser comunicada con audiencias diversas en la forma de visualización de datos en Internet.



## Contenidos / Competencias / Aprendizajes esperados

### Contenidos

Los contenidos del curso se entregarán en cuatro semanas, dos días por semana, para un total de 8 clases en total. Los temas a desarrollar para las ocho clases, son:

#### 1.- Introducción a la Ciencia de Datos.

En esta clase se repasará de forma breve el concepto de ciencia de datos. Se mostrarán aplicaciones avanzadas realizadas por el relator, y se realizará un paso a paso de manera práctica para la instalación de todas las herramientas necesarias que se utilizarán a lo largo del curso (entornos de programación y software - open source).

#### 2.- Introducción a la programación.

En esta clase se enseñará y entrenará a los alumnos en la creación de código. Se enseñarán conceptos como variables, tipos de datos, funciones, control de flujo y objetos. El entorno de programación será Javascript, mediante la utilización de Node.js. Así mismo se entrenará a los alumnos en la utilización de la terminal (*command prompt*), para la interacción con código.

#### 3.- Recolección de datos desde diversas fuentes.

Utilización de aplicaciones construidas en el entorno Node.js para el acceso a datos espaciales y no espaciales desde diversas fuentes públicas, tales como:

- a) Twitter API.
- b) Web Scraping.
- c) Portales abiertos de datos públicos.

En esta clase además se entrenará a los participantes a extraer datos en formatos diversos, tales como csv, json y geojson.

#### 4.- Procesos estadísticos y espaciales.

Entrenamiento para el procesamiento y aplicación de modelos estadísticos y procesos espaciales de datos a través de aplicaciones en el entorno Node.js, una vez que los datos han sido obtenidos desde fuentes diversas. Se pondrá énfasis en el concepto estadístico de regresión y clustering.

#### 5.- Técnicas de análisis de texto y lenguaje natural.

Entrenamiento en la creación y uso de aplicaciones para el análisis de datos en formato de texto escrito, provenientes desde fuentes diversas. Se hará un repaso de la obtención de datos utilizando la técnica de *web scraping* para la obtención de datos desde periódicos en Internet.

Utilizando Node.js se crearán aplicaciones para la creación de un *Corpus* de texto (conjunto de textos), y una aplicación para obtener frecuencia de palabras.



#### 6.- Técnicas y herramientas de visualización de datos.

Se entrenará a los participantes en técnicas de visualización de datos espaciales y no espaciales. Para el caso de datos espaciales, se entrenará a los alumnos(as) en la creación de mapas usando la librería Leaflet, que permite publicar mapas en archivos html en la web.

Para el caso de datos no-espaciales, se utilizará la herramienta en línea desarrollado por el Politécnico de Mlián, `app.rawgraphs` (<http://app.rawgraphs.io/>), la cual permite crear visualizaciones exploratorias, con una salida en formato imagen (png, svg) o modelo de datos (json).

#### 7.- Github para el trabajo colaborativo y la visualización de datos.

Se entrenará a los participantes en la publicación de código, texto (en formato *markdown*) y visualizaciones en la plataforma github (<https://github.com/>). Se realizará un paso a paso, desde la creación de una cuenta (gratuita) en github, hasta la publicación de un producto final de visualización de datos, el cual comprende código, documentación (texto) e imágenes.

#### 8.- Aplicación y creación de visualizaciones de datos

La última clase del curso estará dedicada al trabajo personal del alumno(a) para la creación de una visualización de datos propia, la cual al final de dicha clase será colgada en Internet, utilizando las técnicas entrenadas a lo largo del curso.

#### **Competencias**

El curso está dirigido a personas sin conocimientos de programación o aquellos que se encuentren en el proceso inicial de creación de aplicaciones para la recolección, análisis y visualización de datos. El curso guiará a los participantes paso a paso para su introducción en la ciencia de datos.

En este sentido, las competencias iniciales para la participación del curso son las herramientas necesarias para el aprendizaje de técnicas de programación, dadas por un laptop con una antigüedad no mayor a cinco años. El sistema operativo puede ser Windows, Mac o Linux (en cualquiera de sus distribuciones).

Muchos de los materiales para este curso están en idioma inglés, por lo que se espera al menos un manejo básico de lectura en dicho lenguaje.

#### **Aprendizajes esperados**

Los aprendizajes esperados son:

- Trabajo en el flujo completo de ciencia de datos.
- Programación en Javascript/Node.js
- Creación y uso de herramientas de recolección de datos en Internet.
- Reconocimiento de los formatos de datos espaciales y no-espaciales más utilizados.
- Procesos estadísticos realizados en programación.



- Programación de visualización de datos y su publicación en Internet.
- Trabajo en entornos colaborativos de creación de aplicaciones y software (github).

### Metodología

El método del taller se basa en Workshops, donde cada clase (sesión) corresponde a un taller práctico en donde se combina la transferencia de contenidos teóricos con su aplicación práctica. De esta forma, cada clase tendrá un resultado práctico con el que el/la alumno(a) puede utilizar posteriormente en su trabajo diario.

Cada clase (sesión) está concebida y diseñada en un formato lineal, de tal forma se pueda ir avanzando gradualmente en creciente dificultad, con el fin de llegar al resultado esperado al final del curso, que corresponde a una visualización de datos realizada íntegramente por el/la participante.

### Evaluación

La evaluación final del curso corresponde al cumplimiento/no-cumplimiento de la entrega de una visualización de datos creada por él/ella mismo(a), la cual debe ser colgada en Internet utilizando la plataforma github.

Los componentes de la visualización son:

- Código: los scripts creados tanto para la recolección, procesamiento y visualización de los datos.
- Documentación: texto escrito en formato Markdown que documenta la funcionalidad del código y el objetivo de la visualización de datos.
- Imágenes: de haber producido imágenes complementarias, deben ser dispuestas en el repositorio del alumno(a) en Github.

### Requisitos de Aprobación

Los requisitos para aprobar el curso, son:

- Asistencia mínima de 75% a las actividades del curso.
- Entrega de un producto (trabajo) final (visualización de datos) que incluya código, documentación, imágenes y links al producto final.

### Palabras claves

Ciencia de datos, programación, software, JavaScript, node.js, visualización de datos.

