



**UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA
COMISIÓN ACADÉMICA DE POSGRADO**

DESCRIPCIÓN DEL PROGRAMA DE POSGRADO

Nombre del Programa: *Maestría en Ciencia de Datos y Aprendizaje Automático*

Montevideo – 22 de febrero de 2018

1. IDENTIFICACIÓN

DE LA CARRERA

Nombre del Programa: Maestría en Ciencia de Datos y Aprendizaje Automático

Programa: Maestría

ÁREA ACADÉMICA

Facultad de Ingeniería

Institutos vinculados al Área:

- Instituto de Ingeniería Eléctrica
- Instituto de Computación
- Instituto de Matemática y Estadística

Contacto institucional del Programa

Instituto de Ingeniería Eléctrica

Nombre: Federico Lecumberry

Teléfono: 27110974 x1122

E-mail: lecumberry@fing.edu.uy

Instituto de Matemática y Estadística

Nombre: Paola Bermolen

Teléfono: 27110621 x112

E-mail: paola@fing.edu.uy

Instituto de Computación

Nombre: Lorena Etcheverry

Teléfono: 27142714 int 12148

E-mail: lorenae@fing.edu.uy

2. UBICACIÓN FÍSICA DEL PROGRAMA:

Lugar y dirección completa de la sede del programa:

Facultad de Ingeniería - Julio Herrera y Reissig 565 - 11300 Montevideo - Uruguay

Nombre y teléfono de un contacto en la Institución Sede:

Nombre: Federico Lecumberry
Teléfono: 27110974 x1122

Personal, instalaciones y materiales disponibles para la realización del programa:

El dictado de los cursos y la coordinación se realiza con docentes de, al menos, los institutos de Matemática y Estadística (IMERL), Ingeniería Eléctrica (IIE) e Ingeniería en Computación (INCO). La SCAPA de la carrera también estará integrada por docente representando estos institutos y sus disciplinas. Participan del dictado de las unidades curriculares relacionadas con el programa de maestría al menos 27 docentes de Facultad de Ingeniería; 22 de ellos poseen título de Doctorado.

La Facultad de Ingeniería cuenta con salas de clases para cursos expositivos de posgrado así como salas de informática donde se pueden realizar actividades prácticas o *hands-on*. Asimismo los institutos participantes cuentan con instalaciones similares que pueden complementar el dictado de los cursos y talleres.

Participan de los cursos, como coordinadores y como docentes, 17 docentes nacionales con título de Doctor, 8 docentes con título de Magister y 3 docentes con titulación de ingeniero y licenciaturas. Los 4 docentes del exterior que participan como coordinadores de los cursos tienen titulación de Doctor. Desde el punto de vista edilicio se dispondrá de 4 salones de clase para cursos expositivos de posgrado, 2 salones equipados con PC para el dictado de cursos que incluya la aplicación de herramientas informáticas.

En la Biblioteca Central de Facultad de Ingeniería, así como en las bibliotecas de los distintos institutos se dispone de más de 45000 documentos (libros, revistas, artículos, reportes técnicos, tesis, etc.) vinculados con temáticas asociadas directa e indirectamente a la maestría. Se tiene además acceso a través del Portal Timbó (www.timbo.org.uy), a las bases de datos: Science Direct, Springer, Scopus, IEEE (entre otras). El mismo permite el acceso a 1884 publicaciones académicas y 2500 documentos aproximadamente (libros, conferencias, revistas, informes, etc.) vinculados a las disciplinas de la ciencia de datos y el aprendizaje automático, con acceso a texto completo.

OBJETIVO DEL PROGRAMA

FINALIDAD

La Ciencia de Datos es la disciplina que busca extraer conocimiento de forma sistemática y computacionalmente eficiente a partir de los datos de un dominio. Para esto, utiliza principalmente métodos y técnicas de la matemática y la estadística, la computación y la visualización de datos. Entre las disciplinas relacionadas con la Ciencia de Datos y áreas de aplicación se destacan el aprendizaje automático, el modelado estadístico, el procesamiento de señales, el análisis de datos de alta dimensionalidad, el procesamiento de datos multimedia, el almacenamiento y la recuperación de información en bases de datos distribuidas, su aplicación a la visión artificial y el procesamiento de lenguaje natural, entre otras.

En la Universidad de la República y en Facultad de Ingeniería en particular existe una larga trayectoria y amplios antecedentes de diferentes e importantes líneas de investigación vinculadas a la Ciencia de Datos y el Aprendizaje Automático. El Instituto de Matemática y Estadística (IMERL, FIng) tiene en su plantel docentes con dedicación total que trabajan en aspectos fundamentales de la ciencia de datos. Por un lado, en estadística -en particular para datos de altas dimensiones- y análisis multivariado, prestando especial atención a los problemas relacionados con el aprendizaje automático y con el análisis de datos en general. Por otro lado, en optimización y otros métodos de análisis numérico, que se encuentran presentes en cualquier algoritmo de aprendizaje automático. El Instituto de Ingeniería Eléctrica (IIE, FIng) a través del Departamento de Procesamiento de Señales tiene una amplia trayectoria de más de dos décadas trabajando en análisis de datos, en particular en el área de procesamiento y análisis de señales, tanto con enfoques basados en modelos como basados en datos. Las aplicaciones tradicionales han sido la transmisión y comunicación de datos, la biología, la biomedicina, la biometría y el agro, entre otras, tanto a partir de señales unidimensionales como de imágenes (microscopía, cámaras, satelitales, radar, etc) o datos multidimensionales. Más recientemente, con el aumento vertiginoso de los volúmenes de datos y de la potencia de Cálculo, el aprendizaje automático y los enfoques data driven ha venido cobrando un protagonismo cada vez mayor entre las metodologías para resolver este tipo de problemas. La investigación y producción científica del Instituto de Ingeniería Eléctrica ha ido acompañando esta tendencia. En el Instituto de Computación (INCO, FIng), diferentes grupos de investigación, han estudiado algoritmos de aprendizaje supervisado y no supervisado como parte de los métodos principales para sus dominios, vinculados a diferentes aspectos de la inteligencia artificial como, por ejemplo, robótica, procesamiento de lenguaje natural, o investigación de operaciones. La larga tradición de investigación en Sistema de Información ha generado asimismo conocimiento en el análisis y organización de grandes volúmenes de datos. La experiencia generada en estos temas ha posibilitado que exista ya un perfil "Inteligencia Artificial" en la Licenciatura en Computación, que incluye materias vinculadas a temas de Ciencia de Datos.

OBJETIVOS ESPECÍFICOS:

Los objetivos de la Maestría en Ciencia de Datos y Aprendizaje Automático son promover la profundización y actualización de los conocimientos en las diferentes disciplinas relacionadas con la Ciencia de Datos y el Aprendizaje Automático. Se busca formar recursos humanos altamente capacitados para la investigación y el desarrollo científico-tecnológico, así como para el ejercicio innovador en la profesión y la resolución de problemas específicos de la industria nacional, promoviendo el desarrollo de la misma. Asimismo se busca

formar profesionales con una actitud crítica ante las nuevas posibilidades y tecnologías, y con la capacidad de aplicarlas a la resolución de las necesidades de la sociedad uruguaya en el área.

PERFIL DEL EGRESADO

El egresado adquiere formación en diferentes disciplinas de la Ciencia de Datos y el Aprendizaje Automático. Esto lo capacita en la recolección, depurado, modelado, representación y análisis de datos para resolver problemas como los que se presentan en la industria o la academia. De igual modo adquiere capacidades para interpretar, visualizar y comunicar los resultados obtenidos.

Asimismo es capaz de entender los fundamentos matemáticos y algorítmicos detrás de los métodos de la ciencia de datos y del aprendizaje automático, y comprender la literatura especializada de la disciplina. Por lo tanto, es capaz de proponer y desarrollar nuevos métodos y algoritmos adaptados al tipo de datos y al problema que tenga que resolver.

3. ORGANIZACIÓN Y NORMAS DE FUNCIONAMIENTO

Duración prevista del programa: **2 años**

Número de plazas previstas (incluyendo becas si es aplicable): **30**

Número mínimo de alumnos para realizar el programa: 6

Requisitos para obtener el título

Número de créditos mínimos de Actividad Programada: 65. Estos créditos están divididos en 42 créditos mínimos en Áreas de Formación Básica (Ciencia de Datos: 10 créditos; Estadística: 8 créditos; Aprendizaje Automático: 8 créditos; Gestión de datos e información: 8 créditos; Optimización: 8 créditos) y 15 créditos mínimos en Áreas de Formación Específica.

Horas presenciales mínimas de Actividad Programada: 500.

Tesis: 45 créditos

Ingreso

Perfil de ingreso: El perfil de ingresos al programa se identifica como egresados de las carreras de Licenciatura o Ingeniería en Computación, Ingeniería Eléctrica, la Licenciatura en Matemáticas o la Licenciatura en Estadística.

Requisito de ingreso: Pueden ingresar a la Maestría en Ciencia de Datos y Aprendizaje Automático quienes posean antecedentes académicos de acuerdo a lo expresado en el Artículo 19 del Reglamento General de las Actividades de Posgrado y Educación Permanente de la Facultad de Ingeniería (RGP-FING).

Los aspirantes deben tener conocimientos sólidos en computación, Cálculo, álgebra, probabilidad y

estadística. Aquellos aspirantes que a juicio de la SCAPA-CDAА así lo requieran, previo a su admisión al programa, deberán completar actividades de nivelación para asegurar el completo aprovechamiento de las actividades de posgrado.

Criterios de selección de los candidatos: La admisión tendrá en cuenta los antecedentes profesionales y académicos pudiendo realizar una entrevista a los aspirantes para complementar la información presentada. La CAP resolverá la admisión de los candidatos en base al informe de la SCAPA- CDAА.

4. CUERPO DOCENTE Y SUS ACTIVIDADES

La siguiente es una lista no exhaustiva de docentes de Facultad de Ingeniería que trabajarán en la Maestría. Incluye solamente docentes con Grado 3 o superior, o con título de Doctorado. Como parte de la formación de los estudiantes, y en acuerdo con la SCAPA del Programa, podrán incluirse nuevos docentes, ya sea en el marco de nuevos cursos que se ofrezcan, o como parte de otras actividades programadas.

	Nombre/titulación/instituto	Horas aula anuales dedicadas al programa	N° previsto de candidatos a orientar	Horas anuales de otras actividades vinculadas al programa
1	Bermolen, Paola / Dra / IMERL	60	2	40
2	Bourel, Mathias/ Dr. / IMERL	60	2	40
3	Cancela, Pablo / Dr. Ing. / IIE	40	1	20
4	Castro, Alberto / Dr. Ing./ INCO	60	1	20
5	Delbracio, Mauricio / Dr. Ing. / IIE	80	2	40
6	Etcheverry, Lorena / Dra. Ing./ INCO	60	2	40
7	Etcheverry, Mathias / Mag. Ing./ INCO	40	1	20
8	Fariello, María Inés / Dra. /IMERL	60	2	60
9	Fernández, Alicia / Prof. Ing. / IIE	60	2	40
10	Fiori, Marcelo /Dr. Ing. / IMERL	60	2	40
11	Garat, Diego / MSc. Ing./ INCO	60	1	20
12	Gómez, Álvaro / MSc. Ing. / IIE	60	1	20

13	Lecumberry, Federico / Dr. Ing. / IIE	60	2	60
14	León, José/ Dr. / IMERL	60	1	20
15	Lezama, José / Dr. Ing. / IIE	60	2	40
16	Marotta, Adriana / Dra. Ing./ INCO	60	1	20
17	Martin, Álvaro / Dr. Ing./ INCO	60	1	20
18	Moncecchi, Guillermo / Dr. Ing. / INCO	60	2	40
19	Musé, Pablo / Dr. / IIE	60	2	40
20	Nesmachnow, Sergio / Dr. Ing. / INCO	60	1	20
21	Prada, Juan José / MSc. Ing. / INCO	40	1	20
22	Piria, Alfredo / MSc. Ing. / IMERL	60	0	0
23	Ramírez, Ignacio / Dr. Ing. / IIE	60	2	20
24	Randall, Gregory / Dr. Ing. / IIE	40	1	20
25	Rocamora, Martín / Dr. Ing. / IIE	60	2	20
26	Rodríguez Bocca, Pablo / Dr. Ing. / INCO	60	1	20
27	Rosá, Aiala / Dra. Ing./ INCO	60	1	20
28	Simon, María / Prof. Ing. / IIE	20	1	20
29	Wonsever, Dina / Dra. Ing./ INCO	40	1	20

5. CURRÍCULA

La siguiente es una lista no exhaustiva de cursos ofrecidos regularmente por Facultad de Ingeniería que pueden ser considerados para formar parte de la actividad programada de los estudiantes. Existen otros cursos que pueden formar parte de esta lista pero no fueron agregados y que se pueden ajustar dada la opcionalidad de cada perfil.

<p>Asignatura N° 01 Reconocimiento de patrones</p> <p>Responsable: Profesora Titular Alicia Fernández Instituto: IIE N° de créditos: 12 Horas presenciales: 70</p>
<p>Objetivos: En el curso se verán los principales conceptos teóricos y algoritmos clásicos utilizados para resolver problemas de reconocimiento y clasificación de patrones a partir de datos sensoriales, y se analizará su aplicación a problemas prácticos concretos. Al finalizar el curso, el estudiante estará en condiciones de re-interpretar problemas de clasificación y reconocimiento con un enfoque basado en la teoría de reconocimiento de patrones, y estudiar su solución en base a técnicas estándar. Los conceptos teóricos presentados en el curso podrán ser fuente de inspiración para el desarrollo de nuevas técnicas.</p>
<p>Conocimientos previos exigidos: Los estudiantes deberán tener conocimientos de Cálculo diferencial e integral, probabilidad y estadística.</p>
<p>Metodología de enseñanza:</p> <ul style="list-style-type: none"> • Horas clase (teórico): 44 • Horas clase (práctico): 8 • Horas clase (laboratorio): • Horas consulta: 12 • Horas evaluación: 6 ◦ Subtotal horas presenciales: 70 • Horas estudio: 20 • Horas resolución ejercicios/prácticos: 60 • Horas proyecto final/monografía: 30 ◦ Total de horas de dedicación del estudiante: 180
<p>Forma de evaluación: Entrega de una serie de ejercicios obligatorios (en papel y en máquina), entrega de respuestas a preguntas teóricas y realización de un proyecto final con defensa oral.</p>
<p>Temario: Tema 1: Introducción, modelo de un sistema de reconocimiento de patrones. Tema 2: Teoría de la decisión Bayesiana, estimación paramétrica. Tema 3: Técnicas de clasificación no paramétricas. Tema 4: Selección y extracción de características. Tema 5: Aprendizaje no supervisado y agrupamiento. Tema 6: Funciones discriminantes lineales, Support Vector Machines. Tema 7: Redes neuronales multicapas. Tema 8: Combinación de clasificadores</p>

Bibliografía:

- Pattern Classification– Duda, Hart and Stork, John Wiley & Sons (ISBN-10-0471056693)-2001.
- Pattern Recognition and Machine Learning C. M Bishop, Springer (ISBN-13-9780387310732)-2006
- Pattern Classification- Jürgen Schürmann, John Wiley & Sons (ISBN-10-0471135348)-1996
- Pattern Recognition: A Statistical Approach – P. Devijver and J. Kittler, Prentice Hall (ISBN-10-136542360)-1982
- Combining Pattern Classifiers: Methods and Algorithms – L. I. Kuncheva, WileyInterscience, (ISBN-10: 0471210781)-2004
- Statistical Pattern Recognition, A. Webb, John Wiley & Sons (ISBN-13-9780470682289)
- Data Mining, I. H. Witten, E. Frank, Elsevier inc. (ISBN: 0-12-088407-0 1 -2005)
- Learning with kernels - B. Scholkopf, A.J. Smola, Mit Pr., (ISBN-13-9780262194754)-2002
- The Elements of Statistical Learning, Hastie, Tibshirani and Friedman Springer-Verlag, (ISBN-13-9780387848587) (2008)

Asignatura Nº 02 Aprendizaje Profundo para visión Artificial

Responsable: Profesor Adjunto Dr. Mauricio Delbracio

Instituto: IIE

Nº de créditos: 8

Horas presenciales: 48

Objetivos: En los últimos años, el aprendizaje profundo (deep learning) se ha convertido en una herramienta fundamental en el aprendizaje de máquinas para una amplia variedad de dominios y aplicaciones. Uno de sus mayores éxitos ha sido su aplicación a la visión artificial, donde el desempeño en problemas como el reconocimiento de objetos y acciones ha permitido importantes mejoras en la última década.

El aprendizaje profundo es una técnica que emplea redes neuronales para aprender representaciones a partir de una serie de datos observados, que puedan ser de utilidad para resolver problemas de alto nivel como ser predicción, restauración o clasificación de señales. Dichas representaciones se denominan profundas por estar construidas a partir de una jerarquía compuesta de sucesivas capas que representan las observaciones con un nivel creciente de abstracción. El aprendizaje profundo pretende descubrir las propiedades intrínsecas de grandes volúmenes de datos construyendo representaciones distribuidas, tanto en contextos supervisados como no supervisados. Este curso brinda una Introducción al aprendizaje profundo con un enfoque práctico en la visión artificial. El objetivo principal es presentar a los estudiantes los principales aspectos de modelado, algorítmicos y de optimización de forma de que ellos mismos puedan implementar (diseñar, entrenar y validar) sus propios modelos. Se presentarán modelos simples basados en redes convolucionales que permiten explicar matemáticamente algunas de las propiedades claves para su éxito. Todos los conceptos serán ilustrados con aplicaciones específicas en los campos de visión artificial y procesamiento de imágenes.

Conocimientos previos exigidos: Conocimientos básicos de Cálculo diferencial, Álgebra Lineal,

Probabilidad y Estadística y Programación (conocimientos sólidos en al menos un lenguaje de programación).

Metodología de enseñanza:

- Horas clase (teórico): 27
- Horas clase (práctico): 9
- Horas clase (laboratorio): 0
- Horas consulta: 8
- Horas evaluación: 4
- Subtotal horas presenciales: 48
- Horas estudio: 20
- Horas resolución ejercicios/prácticos: 36
- Horas proyecto final/monografía: 16
- Total de horas de dedicación del estudiante: 120

Forma de evaluación: La evaluación consiste en la entrega de una serie de ejercicios obligatorios en máquina, la entrega de respuestas a preguntas teóricas y la realización de un mini-proyecto final con defensa oral. El mini-proyecto final consistirá en el estudio de un trabajo publicado recientemente o bien en evaluar y profundizar en alguna de las técnicas discutidas durante el curso.

Temario:

1. Introducción al aprendizaje profundo, motivación, reseña histórica
2. Aprendizaje supervisado. Métodos lineales de clasificación, regla del k-vecino más cercano
3. Representaciones de alto nivel, características en imágenes
4. Formulación del aprendizaje como un problema de optimización. Algoritmo de "backpropagation".
5. Redes neuronales totalmente conectadas y su entrenamiento.
6. Redes neuronales de convolución, arquitectura, capas, operadores, y su entrenamiento.
7. Análisis de redes neuronales, visualización de representaciones. Propiedades: invarianza, covarianza, redundancia e invertibilidad.
8. Redes neuronales recurrentes (RNN), redes de corta-larga memoria (LSTM). Entrenamiento y ejemplos.
9. Redes adversas generativas (GAN). Entrenamiento y ejemplos.
10. Aprendizaje profundo en la práctica (bibliotecas). Aplicaciones (clasificación y subtítulo de imágenes, predicción de video, superresolución, eliminación de ruido y borrosidad).

Bibliografía:

- Goodfellow, I., Bengio, Y., Courville, A. "Deep learning". MIT Press, 2016.
- Hastie, T, Tibshirani, R. Friedman, J. "The Elements of Statistical Learning". NY Springer, 2001
- LeCun, Y., Bengio, Y., Hinton, G. "Deep learning". Nature 521.7553 (2015): 436-444.
- Mallat, S., "Understanding deep convolutional networks" Phil. Trans. R. Soc. A 374.2065 (2016): 20150203.
- Krizhevsky, A., Sutskever, I., Hinton, G. "Imagenet classification with deep convolutional neural networks". NIPS, 2012.
- Zeiler, M. D., and Fergus, R. "Visualizing and understanding convolutional networks". ECCV, 2014.
- Li, Fei-Fei, L., Karpathy, A., Johnson, J. "CS231n: Convolutional neural networks for visual recognition". Stanford University, notas de curso, 2015.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. "Generative adversarial nets". NIPS, 2014
- Anh, N., Yosinski, J., Clune, J. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." CVPR, 2015
- Johnson, J., Alahi, A., Fei-Fei, L. "Perceptual losses for real-time style transfer and super-resolution". ECCV 2016.

Asignatura N° 03 Tratamiento de imágenes por computadora

Responsable: Profesor Titular Dr. Gregory Randall

Instituto: IIE

N° de créditos: 10

Horas presenciales: 62

Objetivos: Se introducirán los conceptos principales del tratamiento de imágenes por computadora. Se abarcarán los distintos aspectos de un área muy extensa de manera de dejar claros los conceptos generales subyacentes y abrir la puerta a un estudio más detallado por parte del estudiante. Al finalizar el curso el estudiante comprenderá los fundamentos del tratamiento de imágenes por computadora, tendrá experiencia en programación de algoritmos de tratamiento de imágenes y podrá encarar proyectos de aplicación en esta área.

Conocimientos previos exigidos: Muestreo. Programación básica. Matemáticas de la Ingeniería.

Metodología de enseñanza:

- Horas clase (teórico): 40
- Horas clase (práctico): -
- Horas clase (laboratorio): 16
- Horas consulta: (Las consultas se realizan junto con los laboratorios)
- Horas evaluación: 6
 - Subtotal horas presenciales: 62
- Horas estudio: 20
- Horas resolución ejercicios/prácticos: 30
- Horas proyecto final/monografía: 40
 - Total de horas de dedicación del estudiante: 152

Forma de evaluación: Trabajo individual de programación a realizar por los estudiantes. Proyecto final grupal (típicamente de 2 o 3 estudiantes).

Temario:

1. Introducción: Procesamiento de imágenes: problemas y aplicaciones. Relaciones con disciplinas vecinas. Pasos fundamentales en el procesamiento de imágenes: Esquema general de un sistema de visión por computador.
2. Percepción de imágenes. El sistema visual humano. Representación de Marr
3. Representación digital de una imagen. Arreglos de datos multidimensionales. Imágenes vectoriales. Discretización espacio temporal. Cuantificación.
4. Modelo de imagen. Relación del modelo con el modelo de la visión humana. Modelo de color. Modelo de ruido. Concepto de Apertura. Relación con la multiresolución. MTF. Ejemplos prácticos.
5. Adquisición de imágenes. Sistema de adquisición, iluminación, óptica, etc. Arquitecturas. Formatos de archivos. Temas diversos asociados.
6. Transformaciones geométricas. Transformadas homogéneas. Interpolación. Histograma. Operaciones sobre el histograma.
7. Transformadas. Transformadas de imágenes: Transformada de Fourier. Transformada de Karhunen-Loeve. Propiedades principales: Ejemplos de otras transformadas 2D y sus aplicaciones principales.
8. Restauración. Iluminación no uniforme: modelo multiplicativo, estimación. Distorsiones geométricas: modelo, estimación, calibración. Modelo de la degradación Métodos de restauración.
9. Mejoramiento. Planteamiento del problema. Filtros lineales: promediado. Filtros no lineales: mediana. Filtros en el espacio de frecuencias. Difusión anisotrópica.
10. Segmentación. Detección de discontinuidades vs regularidades. Detección de bordes: Aproximación local sin información a priori: Sobel, Canny. Discontinuidades en un espacio de características: Texturas. Aproximación Global con información a priori: Transformada de Hough. Detección de regiones. Segmentación y umbrales. Segmentación MAP. Split and Merge. Segmentación utilizando información contextual.
11. Morfología binaria. Erosión y Dilatación. Cerradura y Apertura. Esqueleto Idempotencia.
12. Estructuras de representación. Etiquetado. Chain coding. Grafos Búsqueda de puntos con máxima curvatura. Interpolación de curvas: puntos y splines.
13. Descriptores. Momentos Factor de forma Medidas geométricas. Medidas estadísticas.

Bibliografía:

- Digital Image Processing. Rafael C. Gonzales y Richard E. Woods. Addison-Wesley Pub Co, 1992. ISBN: 201508036
- Computer and Robot vision. Robert M. Haralick y Linda G. Shapiro. Addison Wesley Publishing Co., 1992. ISBN 0-201-10877-1
- La visión. David Marr. Alianza Editorial, Madrid, 1982. ISBN: 84-206-6512-6
- Organization in vision: Essays on Gestalt Perception, Gaetano Kanizsa, 1979. ISBN-10: 0275903737 ISBN-13: 978-0275903732
- Fundamentals of Digital Image Processing. Anil K. Jain. Prentice Hall Inc., New Jersey, 1989. ISBN 0-13-336165-9
- Computer vision, A Modern Approach. David A. Forsyth & Jean Ponce, 2002 , ISBN-10: 0130851981 ISBN-13:978-0130851987
- Multiple View Geometry in Computer vision, Richard Hartley & Andrew Zisserman, 2004, ISBN-10:

0521540518 ISBN-13: 978- 0521540513

- Computer vision: Algorithms and Applications, Richard Szeliski, 2010. ISBN-10: 1848829345 ISBN-13: 978-1848829343

Asignatura N° 04 Procesamiento digital de señales de audio

Responsable: Profesor Adjunto Dr. Martín Rocamora

Instituto: IIE

N° de créditos: 8

Horas presenciales: 55

Objetivos: Se introducirán los conceptos y aplicaciones principales del procesamiento digital de señales de audio. Al finalizar el curso el estudiante comprenderá los fundamentos del procesamiento digital de señales de audio, tendrá experiencia en diversas técnicas clásicas y en la programación de algoritmos para resolver problemas típicos, lo que le permitirá abordar proyectos de aplicación en esta área.

Conocimientos previos exigidos: Muestreo y procesamiento digital. Programación básica. Matemáticas de la ingeniería.

Metodología de enseñanza:

- Horas clase (teórico): 36
- Horas clase (práctico): 14
- Horas clase (laboratorio): -
- Horas consulta: -
- Horas evaluación: 5
- Subtotal horas presenciales: 55
- Horas estudio: 10
- Horas resolución ejercicios/prácticos: 25
- Horas proyecto final/monografía: 30
- Total de horas de dedicación del estudiante: 120

Forma de evaluación: Trabajo individual de resolución de ejercicios entregables y proyecto final.

Temario:

1. Introducción al procesamiento de audio: objetivos, aplicaciones, conceptos básicos de señales de audio (e.g. inventariado, densidad espectral de potencia, autocorrelación, etc)
2. Señales de voz y audio: aparato fonador, modelo de producción de voz, modelos para señales de audio en general
3. Percepción auditiva: fisiología del sistema auditivo, psicoacústica, sonoridad, bandas críticas, enmascaramiento, percepción de altura, modelos del sistema auditivo, análisis de panorama sonoro computacional (CASA)
4. Análisis de tiempo corto de señales de audio: Transformada de Fourier de tiempo corto,

- Espectrograma, variantes multiresolución: Constant-Q Transform, Multiresolution FFT
5. Análisis Homomórfico: cepstrum y cepstrum complejo, estimación de envolvente espectral, detección de pitch, deconvolución, coeficientes cepstrales de frecuencia mel (MFCC)
 6. Análisis por predicción Lineal: modelo todo polo, Cálculo de LPC, orden, inestabilidad, estimación de formantes y otras aplicaciones
 7. Codificación de voz y audio: cuantización, codificación con y sin pérdidas, codificación de audio, codificación de voz, CELP
 8. Procesamiento tiempo-frecuencia: overlap-add, convolución rápida, phase vocoder, aplicaciones y transformaciones
 9. Análisis por modelado espectral: modelado espectral, estimación de pitch, seguimiento de parciales, aplicaciones
 10. Filtros digitales con aplicaciones en audio: resonadores, notch, peine, pasa-todo, interpoladores y aplicaciones en efectos de audio y síntesis de sonido
 11. Síntesis de sonido: principales técnicas (aditiva, AM, FM, granular, modelado físico)
 12. Extracción de información musical (MIR): problemas clave (separación de fuentes, transcripción automática de música, reconocimiento de instrumentos, etc)

Bibliografía:

- “Digital Processing of Speech Signals”, Rabiner, L.R. & Schafer, R.W., Prentice Hall, ISBN 978-0132136037, 1978
- “Discrete-Time Speech Signal Processing: Principles and Practice”, Quatieri, T.F., Prentice Hall PTR, 0-13-242942-X, 2002
- “DAFX: Digital Audio Effects”, Zölzer, U., Wiley, ISBN 978-0-471-49078-4, 2002
- “Digital Audio Signal Processing”, Zölzer, U., Wiley, ISBN 978-0-470-99785-7, 2008
- “Percepción auditiva”, Basso G., Universidad Nacional de Quilmes, ISBN 987-558-082-1, 2006
- “Computational Auditory Scene Analysis: Principles, Algorithms, and Applications”, Wang D. & Brown G. (editors), IEEE Press – Wiley, ISBN 978-0-471-74109-1, 2006
- “Digital Signal Processing Primer: With Applications to Digital Audio and Computer Music”, Steiglitz K., Prentice Hall, 978-0805316841, 1996
- “Audio Signal Processing and Coding”, Spanias, A., Painter, T. & Atti, V., Wiley-Interscience, 978-0471791478, 2007
- “Introduction to Digital Filters with Audio Applications”, Smith, J. O., W3K Publishing, 978-0974560717, 2007
- “Physical Audio Signal Processing: for Virtual Musical Instruments and Digital Audio Effects”, Smith, J. O., W3K Publishing, 978-0974560724, 2010
- “Spectral Audio Signal Processing”, Smith, J. O., <http://ccrma.stanford.edu/~jos/sasp/>, online book, 2011
- “Signal Processing Methods for Music Transcription”, Klapuri, A. & Davy, M (editors), Springer, 978-387-30667-4, 2006

Asignatura N° 05 Estimación y predicción en series temporales

Responsable: Profesor Titular Dr. Pablo Musé

Instituto: IIE

N° de créditos: 10

Horas presenciales: 65

Objetivos: Las técnicas que se presentan son parte fundamental del cuerpo teórico de la disciplina procesamiento digital de señales. Se refieren a métodos diseñados para procesar señales que, en el mejor de los casos, pueden caracterizarse en forma estadística, o que se encuentran contaminadas por distintos tipo de ruidos. Este curso debe tomarse en este sentido, y verse como un segundo (o tercer) curso en la materia.

El objetivo es presentar las ideas principales y sus herramientas asociadas, de forma que el/la alumno/a pueda aplicarlas a problemas concretos y a su vez tenga acceso a la vasta literatura del área.

Conocimientos previos exigidos: Transformadas discretas, filtros digitales, probabilidad, Introducción a procesos estocásticos, álgebra lineal. Programación en algún lenguaje de programación científico (Matlab, Octave, R, Python).

Metodología de enseñanza:

- Horas clase (teórico): 40
- Horas clase (práctico): 16
- Horas clase (laboratorio): 0
- Horas consulta: 8
- Horas evaluación: 1
- Subtotal horas presenciales: 65
- Horas estudio: comprendidas dentro de la resolución de ejercicios
- Horas resolución ejercicios/prácticos: 65
- Horas proyecto final/monografía: 35
- Total de horas de dedicación del estudiante: 150

Forma de evaluación: El curso se aprueba con la siguiente evaluación:

1. Entrega de 4 hojas de ejercicios.
2. Realización de una monografía en forma individual. El tema de esta será elegido conjuntamente entre el docente y el estudiante. El estudiante deberá entregar un informe escrito sobre la monografía una semana antes de la presentación.
3. Presentación de la monografía (45 minutos + 15 de preguntas aproximadamente). Se evaluará la calidad de la presentación.

Temario:

1. Introducción (1 clase)
2. Primera Parte: Caracterización de estimadores
 - Estimadores insesgados de varianza mínima (MVU)
 - Cota inferior de Cramer-Rao para la varianza de un estimador insesgado MVU para el

- caso de modelos lineales
 - Estimadores lineales insesgados y de varianza mínima (BLUE) Estimación de parámetros por máxima verosimilitud (MLE)
 - Enfoque Bayesiano: estimación MAP
3. Segunda Parte: Métodos de estimación aplicados
- Caracterización de Procesos Estacionarios Procesos Autorregresivos (AR)
 - Filtros de Wiener
 - Filtros Adaptivos
 - Algoritmo de Máxima Pendiente, Algoritmo LMS
 - Filtro de Kalman, Filtro de Kalman Extendido, Filtro de Kalman sin Perfume (UKF) Algoritmo de mínimos cuadrados recursivo (RLS)

Bibliografía:

- "Fundamentals of Statistical Signal Processing", Volume I: Estimation Theory, Steven M. Kay, Prentice Hall; 1st edition, ISBN 0-13-345711-6, 1993.
- "Adaptive Filter Theory", 3rd Edition, Simon Haykin, Prentice-Hall, New Jersey, ISBN 0-13-004052-5, 1995.
- "Statistical Digital Signal Processing and Modeling", Monson H. Hayes, Wiley, New York, ISBN 0-471 59431-8, 1996.
- "Optimal Filtering", Brian D. O. Anderson and John. B. Moore, Dover Publications, New York, ISBN 0-486-43938-0, 2005.
- "Introduction to Random Signals and Applied Kalman Filtering", R. G. Brown and P. Hwang, John Wiley & Sons, New York, ISBN 0-471-12839-2, 1996.
- "Matrix Computations", 3rd edition, G. H. Golub and C. F. Van Loan, Johns Hopkins Univ Press, ISBN 0-8018-3739-2, 1996

Asignatura N° 06 Procesamiento Semántico del Lenguaje Natural

Responsable: Profesora Titular Dra. Dina Wonsever

Instituto: INCO

N° de créditos: 8

Horas presenciales: 34

Objetivos: Adquisición de conocimientos y experimentación práctica en distintos tipos de representaciones semánticas para el lenguaje natural y en métodos y procesos para su obtención a partir de enunciados lingüísticos. Conocimiento de recursos existentes, de metodologías de construcción de nuevos recursos y de propuestas de transferencia entre distintas lenguas.

Conocimientos previos exigidos: Lógica de predicados Probabilidad y estadística Aprendizaje automático Programación Procesamiento de Lenguaje Natural Los conocimientos requeridos en los primeros cuatro ítems se adquieren en la carrera de Ingeniería en Computación. Los conocimientos previos requeridos en Procesamiento de Lenguaje Natural se adquieren en alguno de estos dos cursos : Introducción al Procesamiento de Lenguaje Natural o Gramáticas Formales para el Lenguaje Natural. Esto significa que al menos uno de estos dos cursos es previo.

Metodología de enseñanza:

- Horas clase (teórico): 30
- Horas clase (práctico): 0
- Horas clase (laboratorio): 0
- Horas consulta: 0
- Horas evaluación: 4
 - Subtotal horas presenciales: 34
- Horas estudio: 30
- Horas resolución ejercicios/prácticos: 0
- Horas proyecto final/monografía: 60
 - Total de horas de dedicación del estudiante: 124

Forma de evaluación: El curso se evalúa en base al proyecto / artículo científico final (contribuye con 30% al puntaje total), la presentación de un artículo en clase (contribuye con 30% al puntaje total) y una prueba individual escrita (contribuye con 40% al puntaje total). Cada una de estas partes tiene un mínimo de suficiencia de un 60%; es imprescindible llegar a estos mínimos para la aprobación del curso.

Temario:

1. Revisión de nociones básicas de Procesamiento de Lenguaje Natural
2. Representaciones semánticas en PLN: semántica lógica/composicional, semántica léxica, etiquetas semánticas en dependencias, semántica específica orientada a acciones un dominio.
3. Anotación semántica, esquema AMR, esquemas particulares para algunos fenómenos: temporalidad, relaciones espaciales, afectividad, otros. Esquemas transversales y esquemas orientados a dominio.
4. Recursos para el análisis semántico:
 - Repositorios léxicos, relaciones entre unidades léxicas: diccionarios, tesauros, bases de datos léxicas (WordNet). Repositorios para el español y el inglés
 - Corpora: texto puro, texto anotado, representaciones vectoriales.
5. Construcción de representaciones semánticas para expresiones lingüísticas (parsing semántico):
 - Aprendizaje supervisado en CCG
 - Aprendizaje semi-supervisado
 - Algoritmos basados en grafos

Bibliografía:

- Consulta (disponible en la biblioteca del InCo) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, (2nd edition), D. Jurafsky y J. H. Martin. Prentice Hall, Upper Saddle River, N.J. ISBN-10: 0131873210. 2008.

Artículos:

- Liang, P., Jordan, M. I., and Klein, D. (2011). Learning dependency-based compositional semantics. In *Proceedings of ACL-11* pages 590-599. (Disponible web, http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00127) Liang, P. and Potts, C. (2015).
- Bringing machine learning and compositional semantics together. *Annual Reviews of Linguistics*, 1(1):355-376. (disponible web en versión casi final, <http://web.stanford.edu/~cgpotts/manuscripts/liang-potts-semantics.pdf>)
- Graph-based Algorithms for Semantic Parsing Jeffrey Flanigan, Samuel Thomson, David Bamman, Jesse Dodge, Manaal Faruqi, Brendan O'Connor, Nathan Schneider, Swabha Swayamdipta, Chris Dyer and Noah A. Smith ACL 2014 (disponible web) Wong, Y. W. and Mooney, R. J. (2006).
- Learning for semantic parsing with statistical machine translation. In *Proceedings of HLT-NAACL-2006*, pages 439-446. (disponible web : <http://dl.acm.org/citation.cfm?id=1220891>)

Asignatura Nº 07 Teoría y Algoritmia de Optimización

Responsable: Dr. Marcelo Fiori, IMERL

Instituto: IMERL

Nº de créditos: 10

Horas presenciales: 60

Objetivos: Este curso incluye el estudio de problemas de optimización continua, con especial énfasis en optimización no lineal, y algoritmos modernos. En particular, se estudiarán algoritmos generales para problemas de optimización, así como métodos de optimización de funciones no diferenciables, o algoritmos utilizados para grandes volúmenes de datos.

Conocimientos previos exigidos: Conocimientos básicos de Cálculo, Álgebra lineal, y Métodos Numéricos. Recomendados: Conocimientos básicos de Optimización, Investigación Operativa, y algún lenguaje de programación interpretado (Matlab, Octave, R, Python, etc.).

Metodología de enseñanza:

- Horas clase (teórico): 40
- Horas clase (práctico): 20
- Horas clase (laboratorio): 0
- Horas consulta: 0
- Horas evaluación: 0
- Subtotal horas presenciales: 60
- Horas estudio: 25
- Horas resolución ejercicios/prácticos: 25
- Horas proyecto final/monografía: 40
- Total de horas de dedicación del estudiante: 150

Forma de evaluación: Entrega de un conjunto de ejercicios teóricos y prácticos.
Aprobación de un examen o proyecto final (sobre un tema a definir entre el estudiante y los docentes).

Temario:

- 1) Introducción
Condiciones de optimalidad. Nociones básicas de convexidad.
Ejemplos de problemas clásicos.
- 2) Métodos de gradiente
Análisis y tasa de convergencia. Elección de paso de gradiente.
Projected Gradient Descent, Frank-Wolfe.
- 3) Multiplicadores de Lagrange
- 4) Dualidad
Subgradientes. Resultados de dualidad. Algoritmo Primal – Dual.
- 5) Métodos sobre funciones no diferenciables y otros métodos
Métodos proximales. ADMM (alternating direction method of multipliers) y variantes.
Avances en algoritmos de optimización.
- 6) Stochastic Gradient Descent

Bibliografía:

- Convex Optimization. S. Boyd, L. Vanderberghe, Cambridge Univ. Press, 2004.
- Nonlinear programming. D. Bertsekas, Athena Scientific, 2016.

Asignatura N° 08 Gramáticas formales para el lenguaje natural

Responsable: Profesora Adjunta Dra. Aiala Rosá

Instituto: INCO

N° de créditos: 13

Horas presenciales: 60

Objetivos: Introducción a los principales temas y aplicaciones en Procesamiento de Lenguaje Natural y al rol de las Gramáticas formales en este contexto. Conocimiento de distintos enfoques y formalismos para la descripción sintáctico semántica del lenguaje natural. Experimentación con herramientas y Gramáticas existentes. Desarrollo de nuevas reglas o extensiones por parte del estudiante. Conocimiento de medidas de evaluación, trabajo con conjuntos de prueba.

Conocimientos previos exigidos: Lógica de predicados. Probabilidad y estadística. Gramáticas formales. Programación.

Conocimientos previos recomendados:

Metodología de enseñanza:

- Horas clase (teórico): 25
- Horas clase (práctico): 10
- Horas clase (laboratorio): 0
- Horas consulta: 22
- Horas evaluación: 3
- Subtotal horas presenciales: 60
- Horas estudio: 22
- Horas resolución ejercicios/prácticos: 15
- Horas proyecto final/monografía: 147
- Total de horas de dedicación del estudiante: 207

Forma de evaluación:

El curso se evaluará a partir de:

- un proyecto: 30%
- una prueba individual escrita: 40%
- estudio en profundidad de algún tema del curso y elaboración de un *paper*: 30%

Las tres instancias son obligatorias y se requerirá como mínimo el 60% del puntaje de la prueba individual escrita. El curso se aprueba obteniendo más del 60% de los puntos totales.

Temario:

- Introducción al área de Procesamiento de Lenguaje Natural. Motivación. Aplicaciones más comunes. Enfoques simbólicos, estadísticos, híbridos.
- Conceptos básicos de gramática. Nociones básicas de gramática del español.
- Distintos modelos para Gramáticas formales:
 - Gramáticas libres de contexto
 - Gramáticas de dependencias
 - Gramáticas categoriales
 - Teoría de Gobierno y Ligadura
 - Gramáticas de restricciones, HPSG
 - Gramáticas probabilistas
 - Reglas contextuales
- Semántica del Lenguaje Natural
- Análisis sintáctico. Métodos de parsing. Medidas de evaluación, conjuntos de prueba.
- Aplicaciones

Bibliografía:

- Bibliografía básica:
 - *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, (2nd edition), D. Jurafsky y J. H. Martin. Prentice Hall, Upper Saddle River, N.J. ISBN-10: 0131873210. 2008.
 - *Introducción a la Teoría del Gobierno y la Ligadura de Noam Chomsky*, S. Costa y M. Malcuori. Serie "Avances de Investigación", Fac. de Humanidades y Ciencias. Montevideo. 1990.
 - *Implementing Typed Feature Structure Grammars*, Ann Copestake (University of Cambridge) Stanford, CA: CSLI Publications (CSLI lecture notes, number 110). ISBN 1-57586-261-1. 2002.
 - *Syntactic Theory. A Formal Introduction*, I. Sag, T. Wasow y E. Bender, CSLI Publications, Stanford University, ISBN: 1575864002, 2003.
- Bibliografía complementaria:
 - *Handbook of natural Language Processing*, R. Dale, H. Moisl, H.Somers, ed., Marcel Dekker, New York.
 - ISBN 0824790006. 2000.
 - *The Logic of Typed Feature Structures*, B. Carpenter, Cambridge University Press, USA. ISBN 0521419328. 1992.
 - *Natural Language Understanding*, J. Allen, Benjamin/Cummings Publishing Company. ISBN 0805303340. 1995.

Durante el curso se indicará bibliografía adicional.

Asignatura N° 09 Aprendizaje Automático

Responsable: Profesor Adjunto Mag. Diego Garat

Instituto: INCO

N° de créditos: 12

Horas presenciales: 26

Objetivos: Esta unidad curricular es una Introducción a los conceptos básicos y algunos de los algoritmos y técnicas utilizados en el área de aprendizaje automático. El estudiante obtiene durante el curso las herramientas fundamentales para abordar otras técnicas más complejas dentro del área.

Conocimientos previos exigidos: Probabilidad y estadística. Lógica de predicados. Estructuras de datos y algoritmia. Programación orientada a objetos. Bases de datos.

Metodología de enseñanza:

El curso se basa en la lectura guiada de la bibliografía y resolución de ejercicios teórico-prácticos, con clases teóricas expositivas de apoyo. Las clases se complementan con cuatro a seis entregas de trabajos de laboratorio, en donde los estudiantes aplican en grupo los conocimientos teóricos adquiridos.

- Horas clase (teórico): 23
- Horas clase (práctico): 0
- Horas clase (laboratorio): 0
- Horas consulta: 0
- Horas evaluación: 3
- Subtotal horas presenciales: 26
- Horas estudio: 54
- Horas resolución ejercicios/prácticos: 100
- Horas proyecto final/monografía: 0

Total de horas de dedicación del estudiante: 180

Forma de evaluación:

El curso tiene un total de 100 puntos que se distribuyen en base a dos componentes:

1. Laboratorios: trabajos grupales por un total de 40 puntos.
2. Evaluación teórico-práctica: pruebas individuales por un total de 60 puntos

Ambas modalidades son eliminatorias: el estudiante debe obtener como mínimo un 50% del puntaje de cada parte para su aprobación.

El curso se aprueba obteniendo como mínimo 60 de los 100 puntos totales con, teniendo en cuenta el punto anterior, un mínimo de 20 puntos en los trabajos de laboratorio y un mínimo de 30 puntos en la

evaluación teórico-práctica.

Temario:

1. Introducción.
2. Aprendizaje conceptual.
3. Árboles de decisión.
4. Preprocesamiento, evaluación y selección de modelos
5. Aprendizaje Bayesiano.
6. Aprendizaje Basado en Casos.
7. Regresión lineal.
8. Regresión logística.
9. Redes neuronales
10. Aprendizaje profundo.
11. Aprendizaje por Refuerzos.
12. Aprendizaje no supervisado

Bibliografía:

Básica

1. Mitchell, Tom (1997). Machine Learning. McGraw-Hill, ISBN 0-07-042807-7.
2. Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). Deep Learning. Mit Press, ISBN 978-0262035613.
3. Jurafsky, Daniel; Martin, James (2009). Speech and Language processing (2nd ed.). Prentice Hall, ISBN 0131873210

Complementaria

1. Bishop, Christopher (2006). Pattern Recognition and Machine Learning. Springer, ISBN 0-38-731073-8.
2. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2001). The Elements of Statistical Learning, data mining, inference, and prediction. Springer, ISBN 0-387-95284-5.
3. Manning, Christopher; Schütze, Hinrich (1999). Foundations of Statistical Natural Language Processing. Mit Press, ISBN 0-262-13360-1.
4. Ng, Andrew (2006). Apuntes del curso Stanford CS229. Disponibles en: <http://cs229.stanford.edu>, último acceso: setiembre 2017.

Asignatura N° 10 Bases de Datos no Relacionales

Responsables: Profesora Adjunta Dra. Lorena Etcheverry

Instituto: INCO

N° de créditos: 10

Horas presenciales: 60

Objetivos: Durante más de 30 años, los sistemas de bases de datos basados en el modelo relacional dominaron el almacenamiento y la gestión de datos. Este modelo de datos simple y robusto, acompañado del soporte a transacciones ACID y del lenguaje de consultas SQL, se transformaron en el estándar de facto para la gestión de datos en todo tipo de aplicaciones. En la última década, y debido a múltiples razones, esta hegemonía se ha visto desafiada. En particular, las aplicaciones sobre la Internet, con billones de usuarios distribuidos mundialmente generando datos constantemente, configuran un nuevo escenario y plantean problemas que los sistemas de bases de datos relacionales no logran resolver adecuadamente. Como consecuencia, en la actualidad coexisten nuevos modelos y estrategias de gestión de datos. A la hora de diseñar una solución informática resulta imprescindible contar con información y hacer un análisis crítico de cada una de estas estrategias, que permita tomar decisiones fundadas más allá de las modas o tendencias.

El objetivo de esta unidad curricular es brindar una visión general sobre estos nuevos modelos de gestión de datos, analizando las fortalezas y debilidades de cada uno, y buscando así enriquecer la mirada y el espíritu crítico sobre estas estrategias que llegaron para quedarse. Al final del curso el estudiante contará con elementos para decidir cuál es el modelo y la estrategia de gestión de datos que mejor se ajusta a su problema.

Conocimientos previos exigidos: Modelo Relacional, Sistemas manejadores de Bases de Datos Relacionales, Manejo de transacciones, Lenguajes de Consultas, Programación

Conocimientos previos recomendados: Arquitecturas de Sistemas de Información

Metodología de enseñanza:

Se trabajará en base a instancias teórico-prácticas (4 horas semanales), donde se realizarán presentaciones teóricas y se pautarán ejercicios para realizar en máquina. Se propiciará la discusión y reflexión sobre los temas tratados. Cada alumno deberá dedicar adicionalmente un promedio de 6 horas semanales para estudio y trabajo fuera del aula.

A lo largo del curso se realizarán diferentes actividades que estimulen la investigación, la lectura de artículos y la discusión y reflexión sobre las problemáticas presentadas, buscando involucrar a los participantes en discusiones activas.

- Horas clase (teórico): 30
- Horas clase (práctico): 30
- Horas clase (laboratorio):
- Horas consulta: 5
- Horas evaluación: 3
 - Subtotal horas presenciales: 68
- Horas estudio: 30
- Horas resolución ejercicios/prácticos: 60
- Horas proyecto final/monografía:
 - Total de horas de dedicación del estudiante: 158

Forma de evaluación:

Se realizarán ejercicios prácticos (mini-proyectos) de cada una de las estrategias presentadas en el curso, los cuales deberán ser realizados en forma individual. Además, habrá una prueba escrita individual al final del curso.

A continuación se presentan los ítems que se tendrán en cuenta en la calificación del curso, junto con su peso relativo:

- (55%) Prueba escrita
- (30%) Ejercicios prácticos
- (15%) participación en clase

Para la participación en clase se tomará en consideración la asistencia, la lectura del material de referencia indicado y la frecuencia/calidad de las contribuciones.

La asignatura se aprueba con el 60% de los puntos, teniendo un mínimo del 60% en cada actividad (prueba escrita, ejercicios y participación).

Temario:

1- Introducción: perspectiva histórica y evolución de los sistemas de bases de datos.

2- Nuevas estrategias de almacenamiento y procesamiento de datos:

2.1- BigData y el ecosistema Hadoop

2.2- El movimiento NoSQL y las key-value stores

2.3- Bases de datos de documentos

2.4- Bases de datos de grafos

2.5- Bases de datos columnares

2.6 – Bases de datos en memoria

3- Aspectos técnicos

3.1 – Distribución de datos

3.2 – Modelos de consistencia

3.3 – Modelos de datos y almacenamiento

3.4 – Lenguajes de consulta

Bibliografía:

- Next Generation Databases: NoSQL, NewSQL, and Big Data. Guy Harrison, Apress, ISBN 978-1-4842-1330-8, 2015. (disponible en el Portal Timbó)
- Concise Guide to Databases: A Practical Introduction. Peter Lake and Paul Crowther, In Undergraduate Topics in Computer Science, Springer-Verlag London, ISBN 978-1-4471-5600-0, 2013. (disponible en el Portal Timbó)
- Lectura de artículos seleccionados.

Asignatura N° 11 Análisis de Datos en Redes

Responsables: Dr. Pablo Rodríguez Bocca, Gr. 4, Instituto de Computación

Instituto: INCO

N° de créditos: 10

Horas presenciales: 60

Objetivos:

Conocimientos previos exigidos:

Conocimientos previos recomendados: Probabilidad y Estadística, Análisis de algoritmos, Algoritmos y estructura de datos, Investigación Operativa

Metodología de enseñanza:

La enseñanza estará realizada fundamentalmente en modalidad de taller+proyecto, o sea, centrada en trabajos de laboratorio y actividades individuales/grupales asistidos por los docentes con el objetivo de realizar un proyecto final. Se dictarán 4 horas semanales de clase teórico-prácticas, incluyendo presentaciones teóricas, clases de práctico en computador, y monitoreo/consultas como apoyo a la realización del proyecto final. Cada alumno deberá dedicar adicionalmente un promedio de 2 horas semanales para estudio adicional y 4 horas semanales para la realización del proyecto final.

En las clases prácticas se requerirá la utilización de un computador personal para realizar los ejercicios en clase. Durante las clases y de forma domiciliaria se realizarán diferentes actividades que estimulen la investigación, la lectura de artículos y la discusión y reflexión sobre las problemáticas presentadas. En todo momento, se buscará involucrar a los participantes en discusiones activas.

- Horas clase (teórico): 32
- Horas clase (práctico): 28
- Horas clase (laboratorio): 0
- Horas consulta: 0
- Horas evaluación: 0
 - Subtotal horas presenciales: 60
- Horas estudio: 30
- Horas resolución ejercicios/prácticos: 0
- Horas proyecto final/monografía: 60
 - Total de horas de dedicación del estudiante: 150

Forma de evaluación:

Se realizará un proyecto de investigación en la temática de la asignatura. Dependiendo de la curricula de estudiantes en cada edición, este proyecto podrá ser realizado en grupos. Los estudiantes de posgrado deben realizar el proyecto en forma individual.

No hay examen final.

Los entregables del curso (y su ponderación para la nota final) consisten en:

- (60%) Informe en formato científico del proyecto de investigación (y código fuente)
- (25%) Presentación oral del proyecto
- (15%) Participación en clase

Para la participación en clase se tomará en consideración: la asistencia a clase; la lectura previa de los temas a cubrir en cada clase; y la frecuencia/calidad de las contribuciones en las clases. No se descarta la

utilización de una evaluación escrita en algunas clases sobre la lectura pautada.

La asignatura se aprueba con el 60% de los puntos, teniendo un mínimo del 60% en cada actividad (informe, presentación, y participación).

Además, durante la ejecución de la asignatura se presentarán ejercicios domiciliarios, que son fuertemente recomendados para el seguimiento del curso, pero no serán requeridos ni evaluados.

Temario:

1. Motivación:

- Utilidad del análisis de redes (análisis de datos, aprendizaje automático, datos en formato de
- redes, visualización)
- El uso de las redes en distintas disciplinas (comunicación e información, biología, física, sociología)
- La representación de grafos (nodos, enlace, matriz de adyacencia, grado de un nodo, redes de
- uno y dos modos)
- Práctico: Introducción al software para analizar redes. Analizar mi red social online (recolectar
- datos de Internet LinkedIn/Facebook/Twitter, visualizar resultados)
- Capítulos: [NE] Cap. 1,2,3,4,5 (1-107)

2. Matemática de las redes: teoría de grafos

- Grafos con peso, hiper-grafos, grafos bipartitos, árboles. grafos planares,
- Caminos, componentes, conectividad, caminantes aleatorios, etc.
- Práctico: Software para visualizar redes (uso de etiquetas y colores)
- Capítulos: [NE] Cap. 6 (107-167)

3. Medidas en redes

- Medidas de centralidad de nodos (eigenvector, closeness, betweenness, PageRank, hubs and
- authorities, etc.)
- Detección de comunidades (componentes, cliques, k-cores, etc.)
- Práctico1: Calcular e interpretar medidas de centralidad de nodo en redes reales
- (LinkedIn/Facebook/Twitter, etc.)
- Práctico2: Calcular e interpretar solapamiento y separación de comunidades en redes
- reales (colaboración científica, blogs políticos, ingredientes de cocina, etc.)
- Capítulos: [NE] Cap. 7 (168-235)

4. Modelos de redes:

- La estructura de gran escala de las redes (componentes, efecto de pequeño mundo)
- Distribución del grado de conectividad (distribuciones power-law y scale-free)
- Grafos aleatorios y aleatorios generalizados
- Creación de redes
- Práctico: Mediante software especializado (NetLogo, Gephi, R), crear grafos aleatorios, calcular
- la distribución del grado y de las componentes, calcular camino más corto, etc.

- Capítulos: [NE] Cap. 8,12,13,14 (235-272, 397-588)
- 5. Procesos sobre redes
 - Fallas de componentes
 - Difusión de información en redes (epidemias)
 - Dinámica en redes
 - Práctico: Evaluar el efecto de mundo pequeño para la difusión de epidemias
 - Capítulos: [NE] Cap. 16,17,18 (589-705)
- 6. Tendencias en el análisis de redes
 - Redes de gran escala
 - Algoritmos y software para el cómputo en redes de gran escala
 - Práctico1: Software para redes de gran escala. Análisis de una red de gran escala
 - Práctico2: Leer y discutir trabajos académicos recientes en el área, o aplicaciones novedosas.

Bibliografía:

Requerida:

- [NE] Mark Newman. Networks: An Introduction. Oxford University Press, Inc. New York, NY, USA 2010. ISBN:0199206651 9780199206650
- [SANDR] Kolaczyk, E.D. and Csardi, G. "Statistical Analysis of Network Data with R". Use R!, Springer New York, 2014. ISBN 9781493909834. (disponible en portal Timbó)

Recomendada:

- [SAND] Kolaczyk, E.D. "Statistical Analysis of Network Data: Methods and Models". Springer Series in Statistics. Springer New York, 2009. ISBN 9780387881461. (disponible en portal Timbó)
- [BA] Albert-László Barabási, Network Science. (disponible online <http://barabasilab.com/networksciencebook/>)
- [NCM] David Easley and Jon Kleinberg. Networks, Crowds, and Markets, Reasoning About a Highly Connected World. Cornell University, New York. Date Published: July 2010. ISBN: 9780521195331 (disponible online)
- [ASN] Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). Analyzing Social Networks. SAGE Publications Limited. ISBN: 9781446247419.

Asignatura Nº 12 Introducción al Procesamiento del Lenguaje Natural

Responsable: Profesor Adj. Mag. Juan José Prada

Instituto: INCO

Nº de créditos: 12

Horas presenciales: 60

Objetivos: Presentar los principales problemas del área de PLN: traducción automática, resumen automático, reconocimiento del habla, recuperación de información, etc. Dar un panorama del estado de las aplicaciones existentes hoy día.

Desarrollar los enfoques y técnicas comúnmente utilizados en el área: análisis morfológico, análisis sintáctico de superficie, análisis semántico, análisis pragmático y modelos del discurso, métodos de estado finito, aprendizaje automático, etc. Estudiar las limitaciones teóricas y materiales de los diferentes enfoques propuestos.

Presentar algunas de las herramientas y recursos más utilizados: taggers, diccionarios, ontologías, Gramáticas, corpus paralelos, etc. Utilizar estas herramientas en la resolución de un problema práctico.

Conocimientos previos exigidos: Lógica de predicados, Probabilidad y estadística, Gramáticas formales, Programación.

Metodología de enseñanza:

Se propone la siguiente metodología:

1. Presentación y cuestionamiento de artículos

Los estudiantes deberán realizar presentaciones de artículos seleccionados. Los docentes del curso ofrecerán un conjunto inicial de artículos y los estudiantes (en grupos de dos personas) deberán elegir un artículo a presentar y un artículo (que será presentado por otro grupo) a cuestionar. La dinámica será la siguiente: el grupo que presenta el artículo debe, además, entregarle al docente, el día de la presentación, un resumen de una página conteniendo los conceptos clave del artículo. El grupo que cuestiona el artículo, deberá entregarle al docente, el día de la presentación, una lista de al menos cinco preguntas a hacer al grupo que presenta el artículo. Las preguntas las formularán los estudiantes de este grupo y el docente será el moderador de la discusión (en la que puede participar el resto de la clase), y evaluará la pertinencia e interés de las preguntas.

2. Proyectos

Los estudiantes deberán realizar un proyecto a lo largo del curso. Los docentes del curso pondrán a disposición de los estudiantes las herramientas y recursos, y una lista de proyectos.

3. Prueba final

Se prevé una prueba individual escrita final, con el objetivo de evaluar los conocimientos del curso.

Se estima un total de 180 horas de trabajo por parte del estudiante, desglosadas de la siguiente forma:

- Horas clase (teórico): 52
- Horas clase (práctico):
- Horas clase (laboratorio): 6
- Horas consulta:
- Horas evaluación: 2
 - Subtotal horas presenciales: 60
- Horas estudio: 20
- Horas resolución ejercicios/prácticos: 20
- Horas proyecto final/monografía: 80
 - Total de horas de dedicación del estudiante: 180

Forma de evaluación:

El curso se evaluará a partir de:

- la presentación del artículo: 15%
- el cuestionamiento del artículo: 10%
- el proyecto: 35%
- la prueba individual escrita: 40%

Las cuatro instancias son obligatorias y se requerirá como mínimo el 60% del puntaje de la prueba

individual escrita. El curso se aprueba obteniendo más del 60% de los puntos totales.

Temario:

1. Introducción al área. Motivación. Aplicaciones más comunes. Enfoques simbólicos, estadísticos, híbridos.
2. Nociones básicas de gramática del español.
3. Análisis léxico. Tokenización. Segmentación en oraciones. Entidades con nombre. Morfología.
4. Análisis sintáctico. Gramáticas. Métodos de parsing.
5. Creación de corpus. Realización de POS tagging usando métodos estadísticos.
6. Análisis semántico. Ambigüedad. Anáforas y catáforas. Representación de conocimiento del mundo.
7. Lingüística textual. Modelos del discurso
8. Aplicaciones

Bibliografía:

Bibliografía básica:

- Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice Hall, Upper Saddle River, N.J., 2000. D. Jurafsky, James H. Martin. ISBN 0130950696.
- Foundations of Statistical Natural Language Processing, Manning & Schütze, Mit Press Cambridge, 1999. ISBN: 9780262133609.

Bibliografía complementaria:

- Handbook of natural Language Processing, R. Dale, H. Moisl, H.Somers, ed., Marcel Dekker, New York, 2000. ISBN 0824790006.
- Natural Language Understanding, J. Allen, Benjamin/Cummings Publishing Company, 1995. ISBN 0805303340.
- Finite State Morphology, Kenneth R. Beesley and Lauri Karttunen, CSLI Publications, ISBN (Paperback): 1575864347.
- Durante el curso se indicarán artículos científicos a ser presentados.

Asignatura N° 13 Estadística Multivariada Computacional

Responsables: Dr. Mathias Bourel (Gr 3, IMERL)

Instituto: IMERL

N° de créditos: 10

Horas presenciales: 65

Objetivos: Introducir los aspectos metodológicos de algunas técnicas de la estadística multivariada y sus

modernas variantes.

Aplicar dichas técnicas a conjuntos de datos reales e interpretar los resultados obtenidos.

Acercar el estudiante al empleo de los paquetes para el análisis estadístico de datos disponibles en el ambiente de

desarrollo de software libre R (<http://www.r-project.org/>).

Conocimientos previos exigidos: Introducción a la probabilidad y estadística, álgebra lineal, Cálculo matricial.

Conocimientos previos recomendados: teoría de la inferencia estadística

Metodología de enseñanza:

- Horas clase (teórico): 28
- Horas clase (práctico): 20
- Horas clase (laboratorio): 0
- Horas consulta: 15
- Horas evaluación: 2
 - Subtotal horas presenciales: 65
- Horas estudio: 30
- Horas resolución ejercicios/prácticos: 25
- Horas proyecto final/monografía: 25
 - Total de horas de dedicación del estudiante: 145

Forma de evaluación: Para la evaluación se tendrá en cuenta:

- 40% evaluación continua (entrega de ejercicios y parciales).
- 20% entrega proyecto final
- 40% defensa proyecto y examen oral (eliminadorio).

Temario:

- Introducción al análisis estadístico de datos multivariados con R.

- Regresión lineal simple y múltiple.
- Análisis de componentes principales. Análisis factorial.
- Análisis discriminante.
- Técnicas de segmentación de datos (cluster análisis).
- Técnicas avanzadas de clasificación y de regresión. Aprendizaje Automático.

Bibliografía:

- Modern Multivariate Statistical Techniques, Alan Julian Izenman, Springer, ISBN: 978-0-387-78188-4, 2008.
- Análisis de datos multivariantes, Daniel Peña, Mac Graw Hill, 2002
- Multivariate Statistics: High-Dimensional and Large-Sample Approximations, Yasunori Fujikoshi, Vladimir V.
- Ulyanov, Ryoichi Shimizu, Wiley, ISBN: 978-0-470-41169-8, 2010.
- Multivariate Statistical Inference and Applications, Alvin C. Rencher, Wiley, ISBN: 0-471-57151-2, 1998.
- A First Course in Multivariate Statistics, Bernhard Flury, Springer, ISBN: 978-0-387-98206-9, 1997.
- A Handbook of Statistical Analyses Using R, 2nd edition, Brian S. Everitt y Torsten Hothorn, Chapman & Hall/CRC, ISBN: 978-1-4200-7933-3, 2010.
- An introduction to Statistical Learning with Applications in R, G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2013.

Asignatura N° 14 Optimización con aplicación a la planificación con incertidumbre

Responsable: MSc.Ing.Alfredo Piria

Instituto: IMERL

N° de créditos: 10

Horas presenciales: 60

Objetivos: El curso incluye el estudio de problemas de optimización lineal y no lineal, en particular aquellos en que existe incertidumbre en los datos, conocidos como de programación estocástica. Se desarrollarán los modelos y herramientas algorítmicas necesarias para la resolución de algunos problemas, así como la fundamentación teórica de esos algoritmos. Como aplicación se tratarán programas usados en problemas de diseño y planificación. En el trabajo práctico se usará preferentemente el software Matlab o similar

Conocimientos previos exigidos: Los del ciclo matemático de la Facultad, incluyendo métodos numéricos.

Metodología de enseñanza:

- Horas clase (teórico): 30
- Horas clase (práctico): 21
- Horas clase (laboratorio):
- Horas consulta:
- Horas evaluación: 9
 - Subtotal horas presenciales: 60
- Horas estudio: 15
- Horas resolución ejercicios/prácticos: 75
- Horas proyecto final/monografía:
 - Total de horas de dedicación del estudiante: 150

Forma de evaluación: Entrega obligatoria de resultados de un conjunto de problemas propuestos (del orden de 10) y aprobación de un examen final, oral y eliminatorio.

Temario:

- 1) Conceptos generales de programación no lineal. Métodos cuasi-newton para problemas sin restricciones, condiciones de Kuhn-Tucker y sensibilidad, programación cuadrática secuencial en problemas con restricciones.
- 2) Algoritmos de programación lineal y dinámica. Algoritmo Simplex y métodos de punto interior en problemas lineales. Programación dinámica.
- 3) Relajación y Dualidad Relajación lagrangeana. Condiciones de dualidad débil y fuerte. Aplicaciones a problemas de ingeniería y economía.
- 4) Métodos de descomposición. Descomposición por precios y por cortes de Benders.
- 5) Programación estocástica. Modelos de escenarios, con recurso, y multietapas. Método de Benders anidado, de hedging progresivo y de splitting de variables. Programación dinámica estocástica.
- 6) Aplicación a problemas de planificación.

Bibliografía:

- Convex Analysis and Minimization Algorithms. J.B. Hiriart-Urruty, C. Lemarechal, Springer Verlag – 1993.
- Programmation Mathématique. Minoux, M. - Dunod - 1983.
- Introduction to Stochastic Programming. Birge, J y Louveaux, F. - Springer Verlag 1997.
- Practical Optimization. P. Gill, W. Murray, M. Wright, Academic Press, 1981.
- Nonlinear programming. D. Bertsekas, Athena Scientific, 1999.
- Convex Optimization. S. Boyd, L. Vanderberghe, Cambridge Univ. Press, 2004.
- Numerical Optimization: Theoretical and practical aspects. C. Lemarechal, F. Bonnans, C. Sagastiazabal, Springer Verlag, 2006

Asignatura N° 15 Introducción a la Teoría de la información

Responsable: Prof. Ing. María Simon

Instituto: IIE

N° de créditos: 10

Horas presenciales: 50

Objetivos: Al finalizar el estudiante conocerá los conceptos y fundamentos matemáticos de la Teoría de la Información, en particular los límites teóricos de compresión y codificación de datos en la comunicación de señales. Asimismo manejará las herramientas y resultados matemáticos básicos en el área para su aplicación en modelado y diseño de sistemas de comunicación de señales.

Conocimientos previos exigidos: Teoría de la Probabilidad

Metodología de enseñanza:

- Horas clase (teórico): 32
- Horas clase (práctico): 8
- Horas clase (laboratorio): 0
- Horas consulta: -
- Horas evaluación: 10
- Subtotal horas presenciales: 50
- Horas estudio: 32
- Horas resolución ejercicios/prácticos: 40
- Horas proyecto final/monografía: 30
- Total de horas de dedicación del estudiante: 152

Forma de evaluación: La asignatura se evaluará por medio de cuatro pruebas parciales, cada una sobre una parte específica del contenido del curso, y un proyecto final, que consiste en escribir un resumen y realizar una exposición oral sobre un tema en base a publicaciones del área. Para aprobar el curso deberá alcanzarse un mínimo de 60 puntos entre todas las instancias de evaluación,

donde cada prueba parcial vale un total de 15 puntos y el proyecto final vale 40 puntos. La nota de aprobación se calculará en función del puntaje total obtenido.

Temario:

1. Introducción a la Teoría de la Información
2. Definiciones y conceptos básicos
3. Asymptotic Equipartition Property (AEP)
4. Tasa de entropía de un proceso estocástico.
5. Codificación de fuentes
6. Codificación canal
7. Teorema de separabilidad fuente/canal
8. Canal con realimentación
9. El Canal Gaussiano
10. Rate distortion theory

Bibliografía:

- “Elements of Information Theory”, Thomas M. Cover, Joy A. Thomas. 2da. edición, Wiley-Interscience; 2 edition (July 18, 2006), ISBN: 0471241954 (o 1era. edición, Wiley-Interscience (August 12, 1991), ISBN: 0471062596)
- “Information Theory and Coding”, Norman Abramson, McGraw-Hill Education (December 1963), ISBN: 0070001456
- “Information Theory”, Robert B. Ash, Dover Publications; New Ed edition (November 1, 1990), ISBN: 0486665216
- Artículos relevantes del área.

Asignatura N° 16 Computación distribuida y procesamiento de grandes volúmenes de datos

Responsable: Prof. Titular Sergio Nesmachnow

Instituto: INCO

N° de créditos: 6

Horas presenciales: 35

Objetivos: Introducir al estudiante en las temáticas de computación distribuida y a las técnicas y tecnologías para resolver problemas de análisis de datos masivos.

Conocimientos previos exigidos: Fundamentos de informática.

Metodología de enseñanza:

Horas clase (teórico): 20

Horas clase (práctico): 5
Horas clase (laboratorio): 5
Horas consulta: 5
Horas evaluación: 0

Subtotal horas presenciales: 35

Horas estudio: 25
Horas resolución ejercicios/prácticos: 10
Horas proyecto final/monografía: 20

Total de horas de dedicación del estudiante: 90

Forma de evaluación: Participación oral y actividades prácticas realizadas en clases (20%). Entrega de un trabajo final (80%) que presente una aplicación de procesamiento de datos que utilice las técnicas de computación distribuida presentadas en el curso para la resolución de un problema específico.

Temario:

1. Computación distribuida y computación cloud.
2. Manejo de grandes volúmenes de datos.
3. El modelo de computación Map-Reduce.
4. El framework Hadoop y su ecosistema.
5. Almacenamiento: HDFS y HBase.
6. Aplicaciones de Map Reduce sobre Hadoop: conteo, índice invertido, filtros.
7. Procesamiento de datos en tiempo real: Apache Spark.
8. Ejemplos de aplicaciones en Spark y el lenguaje Scala.
9. Análisis de datos utilizando Spark y el lenguaje R.
10. Aplicaciones iterativas: Google Pregel y Apache Giraph..

Bibliografía:

- Kai Hwang, Jack Dongarra, and Geoffrey C. Fox. 2011. Distributed and Cloud Computing: From Parallel Processing to the Internet of Things (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN:0123858801 9780123858801.
- George Reese. 2009. Cloud Application Architectures: Building Applications and Infrastructure in the Cloud. O'Reilly Media, Inc.. ISBN:0596156367 9780596156367.
- Mario Macías, Mauro Gómez, Rubèn Tous y Jordi Torres. 2015. Introducción a Apache Spark para empezar a programar el Big Data. Editorial UOC. ISBN: 978-84-9116-037-3.

Asignatura N° 17 Calidad de Datos e Información

Responsable: Profesora Agregada Dra. Adriana Marotta

Instituto: INCO

N° de créditos: 6

Horas presenciales: 36

Objetivos: El objetivo principal de este curso es brindar al estudiante conocimientos básicos sobre calidad de datos en sistemas de información. Se espera que el estudiante conozca los principales fundamentos y técnicas existentes para poder abordar la problemática de la calidad de los datos y de la información. Además, se espera que conozca en profundidad los avances científicos más importantes en el área.

Conocimientos previos exigidos: Conocimientos de Bases de Datos y Programación.

Metodología de enseñanza:

Horas clase (teórico): 12

Horas clase (práctico):

Horas clase (laboratorio):

Horas consulta: 22

Horas evaluación: 2

Subtotal horas presenciales: 36

Horas estudio: 10

Horas resolución ejercicios/prácticos:

Horas proyecto final/monografía: 50

Total de horas de dedicación del estudiante: 96

Forma de evaluación: La evaluación se realizará a partir del informe y la presentación oral del trabajo realizado.

Temario:

- Introducción
 - Concepto de Calidad de Datos
 - Problemas y consecuencias de la mala calidad
 - Necesidades en distintas áreas de aplicación
 - Iniciativas académicas e industriales
- Dimensiones de calidad
 - Multi-dimensionalidad de la calidad
 - Estudio de algunas dimensiones
 - Relaciones entre dimensiones
- Tratamiento de la calidad en Sistemas de Información
 - Modelos de calidad
 - Evaluación de la calidad
 - Corrección y prevención
 - Tratamiento de la calidad en distintas representaciones de datos
 - Tratamiento de la calidad en sistemas de integración de datos

Bibliografía:

- Data and Information Quality. Carlo Batini, Monica Scannapieco. Springer. ISBN: 978-3-319-24104-3. 2016.
- Big Data Integration. Synthesis Lectures on Data Management. Xin Luna Dong, Divesh Srivastava. Morgan & Claypool Publishers, 2015. ISBN: 1627052240, 9781627052245