



Aprobado por resolución N°113 del CFi de fecha 04.07.2017



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY

## Programa de Integración de datos

### 1. NOMBRE DE LA UNIDAD CURRICULAR

Integración de Datos

### 2. CRÉDITOS

8 créditos

### 3. OBJETIVOS DE LA UNIDAD CURRICULAR

El objetivo general es capacitar al estudiante para que comprenda y maneje los problemas teóricos de la integración de datos.

Objetivos Particulares:

1. Que el estudiante comprenda y conozca técnicas para resolver los problemas de heterogeneidad y de consistencia en la integración de datos.
2. Que el estudiante pueda evaluar las diferentes metodologías de integración de datos según el contexto de la aplicación.
3. Que el estudiante pueda diseñar sistemas de integración de datos para aplicaciones particulares, considerando tanto el volumen como la complejidad de los datos

### 4. METODOLOGÍA DE ENSEÑANZA

La metodología de enseñanza se estructura a través de clases teórico-prácticas. Además de las clases teórico-prácticas se estima que cada estudiante debe dedicar 4 horas de estudio semanales. El estudiante deberá participar en foros, analizar y presentar técnicas de integración, realizar ejercicios, un proyecto y una prueba individual. Se propondrá la conformación de grupos de estudiantes para realizar tareas de forma colaborativa.

La dedicación total del estudiante puede resumirse de la siguiente manera:

- ☞ 28 horas de clases en encuentros presenciales
- ☞ 28 horas de estudio individual
- ☞ 26 hrs de actividades prácticas (foros, ejercicios, etc.)
- ☞ 40 horas elaboración y presentación del proyecto final.

## 5. TEMARIO

### Tema 1: Introducción

**Subtemas:** Areas de aplicación para la integración de datos, Tipos de integración de datos.

### Tema 2: Aspectos críticos para la Integración de datos.

**Subtemas:** Terminología, Heterogeneidades, Autonomía, Distribución, Volumen.

### Tema 3: Arquitecturas para la Integración de Datos

**Subtemas:** Sistemas federados, Wrappers/Mediadores, Peer-to-Peer, Hub Centralizado

### Tema 4: Metodologías de Integración en bases de datos

**Subtemas:** Integración de vistas, Global-As-View, Local-As-, Global-Local-As-View, SchemaMapping y Schema Matching.

### Tema 5: XML en la integración de datos.

### Tema 6: Resolución de entidades

### Tema 7: Integración en la Web Semántica

**Subtemas:** Datasets, RDF(S), LinkedData

### Tema 8: Otros aspectos a considerar en la integración de datos

**Subtemas:** Incertidumbre, Trazabilidad

## 6. BIBLIOGRAFÍA

Tema	Básica	Complementaria
Introducción	(1)	(2)
Aspectos críticos para la Integración de datos	(1)	(2),(3)
Arquitecturas para la Integración de Datos	(1)	(2),(4)
Metodologías de Integración	(1)	(2),(5)
XML en la integración de datos	(1)	(*)
Resolución de entidades	(1)	(*)
Integración en la Web Semántica	(1)	(*)
Otros aspectos a considerar en la integración de datos	(1)	(*)

### 6.1 Básica

1. AnHai Doan, Alon Halevy, and Zachary Ives (2012). Principles of Data Integration Editorial: Morgan Kaufmann.

### 6.2 Complementaria

2. Elmagarmid, M Rusinkiewicz y A. Sheth. (1999) Management of Heterogeneous and Autonomous Database Systems. Editorial: Morgan Kaufmann. (ISBN 1-55860-216-X)

3. Maurizio Lenzerini (2002) Data integration: A theoretical perspective. PODS02: 233-246
4. Özsu, M. Tamer, Valduriez, Patrick (2011) Principles of Distributed Database Systems. Springer-Verlag New York.
5. Ronald Fagin, Laura M. Haas, Mauricio A., Hernandex, Renee J. Miller Lucian Popa, Yannis Velegrakis. (2009) .Clio: Schema Mapping Creation and Data Exchange. Conceptual Modeling: Foundations and Applications, 198-236.

(\*) se complementará con artículos actuales de revistas y conferencias del área para discusión sobre las posibles aplicaciones. Algunos ejemplos posibles son:

- Tomer Sagi, Avigdor Gal, Omer Barkol, Ruth Bergman, Alexander Avram (2016). Multi-Source Uncertain Entity Resolution: Transforming Holocaust Victim Reports into People. Proceedings of the 2016 International Conference on Management of Data, 807-819.
- Victor Christen, Anika Groß, Jeffrey Fisher, Qing Wang, Peter Christen, Erhard Rahm (2017) Temporal group linkage and evolution analysis for census data En 19th International Conference on Extending Database Technology (EDBT), Venice, 2017
- D Vatsalan, P Christen (2016) Privacy-preserving matching of similar patients. Journal of biomedical informatics 59, 285-298.
- Achille Fokoue, Mohammad Sadoghi, Oktie Hassanzadeh, Ping Zhang (2016). Predicting Drug-Drug Interactions Through Large-Scale Similarity-Based Link Prediction. ESWC 2016: 774-789.
- Rui Meng, Lei Chen, Yongxin Tong, Chen Jason Zhang (2017). Knowledge Base Semantic Integration Using Crowdsourcing. En IEEE Trans. Knowl. Data Eng. 29(5): 1087-1100.
- Tanay Kumar Saha, Baichuan Zhang, Mohammad Al Hasan (2015) Name disambiguation from link data in a collaboration graph using temporal and topological features. Social Netw. Analys. Mining 5(1): 11:1-11:14.
- Patrick Arnolda, Christian Wartnerb, and Erhard Rahm (2016). Semi-Automatic Identification of Counterfeit Offers in Online Shopping Platforms Journal of Internet Commerce.

## 7. CONOCIMIENTOS PREVIOS EXIGIDOS Y RECOMENDADOS

**7.1 Conocimientos Previos Exigidos:** Se requieren conocimientos sobre diseño, consultas y administración de bases de datos relacionales. Se requieren también conocimientos de lógica de primer orden.

**7.2 Conocimientos Previos Recomendados:** Los conocimientos complementarios que pueden ayudar son sobre manejo de grafos y procesamiento de lenguaje natural.

## ANEXO A

### Para todas las Carreras

Esta primera parte del anexo incluye aspectos complementarios que son generales de la unidad curricular.

### A1) INSTITUTO

Instituto de Computación

### A2) CRONOGRAMA TENTATIVO

Semana 1	Tema 1 (2 hs de clase).
Semana 2	Tema 2 (2 hs de clase).
Semana 3	Tema 2 (2 hs de clase).
Semana 4	Tema 3 (2 hs de clase).
Semana 5	Tema 3 (2 hs de clase).
Semana 6	Tema 4 (2 hs de clase).
Semana 7	Tema 4 (2 hs de clase).
Semana 8	Tema 5 (2 hs de clase).
Semana 9	Tema 6 (2 hs de clase).
Semana 10	Tema 7 (2 hs de clase).
Semana 11	Tema 7 (2 hs de clase).
Semana 12	Tema 8 (2 hs de clase).
Semana 13	Evaluaciones Proyecto Final
Semana 14	Evaluaciones Proyecto Final

### A3) MODALIDAD DEL CURSO Y PROCEDIMIENTO DE EVALUACIÓN

La Unidad Curricular se desarrollará en base a actividades en el EVA y de encuentros presenciales, los encuentros presenciales tienen carácter de obligatorios, no pudiendo el estudiante faltar a más del 20 % de los mismos.

Procedimientos de evaluación:

- ↻ Realización en grupo de la totalidad de los ejercicios indicados como obligatorios por cada tema.
- ↻ Una prueba individual de evaluación de los conceptos adquiridos.
- ↻ Realización y presentación de un proyecto realizado en grupo.

Puntajes de evaluación total: 10% de actividades prácticas, 50% prueba de evaluación individual y 40% proyecto grupal.

La nota de aprobación corresponde al 60% de la evaluación total. Todas las instancias de evaluación son eliminatorias y requieren de un mínimo de 60%.

**A4) CALIDAD DE LIBRE**

Esta unidad curricular no adhiere a la resolución sobre la Calidad de Libre

**A5) CUPOS DE LA UNIDAD CURRICULAR**

No, tiene cupos

**ANEXO B para la carrera Ingeniería en Computación (plan 97)**

**B1) ÁREA DE FORMACIÓN**

Bases de Datos y Sistemas de Información

**B2) UNIDADES CURRICULARES PREVIAS**

Para el Curso: Examen aprobado de Fundamentos de Bases de Datos y 250 créditos de avance en la carrera

Para el Examen: No aplica

**ANEXO B para la carrera Ingeniería en Computación (plan 87)**

**B1) ÁREA DE FORMACIÓN**

No corresponde

**B2) UNIDADES CURRICULARES PREVIAS**

- Para el Curso: Previas comunes a las electivas y examen de Bases de Datos.
- × Para el Examen: No aplica

Observación: Esta unidad curricular se corresponde con una electiva

**APROB. RES. CONSEJO DE FAC. VVO.**  
Fecha 10/4/2018 Exp. 060120-000361-12