



Universidad Nacional de Salta

FACULTAD DE CIENCIAS EXACTAS  
Av. Bolivia 5150 - 4400 - Salta  
Tel. (0387)425-5408 - Fax (0387)425-5449  
Republica Argentina

SALTA, 02 de junio de 2017

EXP-EXA: 8201/2017

RESCD-EXA: 230/2017

VISTO:

La Nota-exa N° 548/17 presentada por el Director del Departamento de Informática, Dr. Cristian Alejandro Martínez, mediante la cual propone el dictado del curso de extensión "Minería de la Web", a cargo de la Dra. Ana Gabriela Maguitman (UNS), en el marco de las Segundas Jornadas Informáticas de la UNSa.

CONSIDERANDO:

Que Comisión de Hacienda aconseja imputar los gastos del curso a los fondos del Departamento de Informática.

Que la Comisión de Docencia e Investigación en su despacho de fs. 35 aconseja autorizar el dictado del curso.

Que el curso en cuestión se encuadra en la Res. CS-309/00 (Reglamento para Cursos de Extensión de la Universidad) y en la RESCD-EXA N° 017/16.

POR ELLO y en uso de las atribuciones que le son propias.

EL CONSEJO DIRECTIVO DE LA FACULTAD DE CIENCIAS EXACTAS

(en su sesión ordinaria de fecha 31/05/17)

R E S U E L V E:

ARTICULO 1º: Autorizar el dictado del Curso de Extensión "Minería de la Web", a cargo de la Dra. Ana Gabriela Maguitman, con las características y requisitos que se explicitan en el Anexo I de la presente resolución.

ARTICULO 2º: Disponer que una vez finalizado el dictado del curso, la directora responsable elevará el listado de los promovidos para la confección de los certificados y/o constancias respectivos, los que serán emitidos por esta Unidad Académica de acuerdo a la reglamentación vigente.

ARTICULO 3º: Dejar aclarado que la presente resolución no acredita la concreción del curso; para ello la directora responsable del mismo deberá elevar el informe final de realización correspondiente, con los detalles que el caso amerite, dentro de los 8 (ocho) meses desde la finalización del dictado. En caso de que el curso no se pudiera dictar, la responsable deberá informar tal situación, dentro de los 30 (treinta) días de la fecha prevista para su inicio.

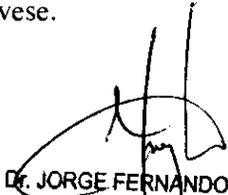
ARTICULO 4º: Imputar los gastos de pasajes, alojamiento y viáticos a favor de la Dra. Ana Gabriela Maguitman, a fondos del Departamento de Informática.

ARTICULO 5º: Hágase saber con copia a la Dra. Ana Gabriela Maguitman, al Departamento de Informática, a la Dirección Adm. Económica y Financiera, a la Dirección General Adm. Económica y al Departamento Adm. de Posgrado. Cumplido, resérvese.

mxs  
rer

  
Mag. GUSTAVO DANIEL GIL  
SECRETARIO DE EXTENSION Y BIENESTAR  
FACULTAD DE CS. EXACTAS - UNSa.



  
Dr. JORGE FERNANDO YAZLLE  
DECANO  
FACULTAD DE CS. EXACTAS - UNSa.



Universidad Nacional de Salta

FACULTAD DE CIENCIAS EXACTAS  
Av. Bolivia 5150 - 4400 - Salta  
Tel. (0387)425-5408 - Fax (0387)425-5449  
Republica Argentina

Anexo I de la RESCD-EXA: 230/2017 - EXP-EXA: 8201/2017

**Curso de Extensión: “Minería de la Web”**

**Directora Responsable del Curso:** Dra. Ana Gabriela Maguitman (UNS)

**Cuerpo Docente:** Dra. Ana G. Maguitman (UNS) y Dr. Eduardo Xamena.

**Organizado:** Dentro de las actividades de las 2das. Jornadas Informáticas de la U.N.Sa. (JIUNSa.)

**Fundamentación**

El curso está orientado a proveer una introducción a los conceptos básicos de la minería de la Web. En la primera parte del curso, se estudiarán los conceptos fundamentales de la recuperación de la información clásica para pasar luego al análisis de técnicas de crawling, indexación y búsqueda en la Web. Se presentará además una introducción a las redes sociales, la Web invisible y la búsqueda temática.

La segunda parte del curso estará orientada al estudio de aplicaciones diversas y temas especiales. Se discutirán aplicaciones de algoritmos de aprendizaje supervisado, no supervisado y semi-supervisado, sistemas de personalización y recomendación, técnicas de minería de datos en casos prácticos y sistemas para el soporte de toma de decisiones basados en información extraída de la Web.

**Objetivos**

Este curso dará la oportunidad de explorar conceptos y tecnologías actuales de minería de la Web y aplicar estos conceptos y tecnologías en el desarrollo de un proyecto de interés.

**Metodología de enseñanza**

Las clases presenciales serán teórico-prácticas y serán impartidas en laboratorio. Los contenidos teóricos del curso serán desarrollados mediante diapositivas. Las prácticas se realizarán usando computadoras a través del desarrollo de programas y uso de herramientas específicas.

A través de la modalidad virtual, se desarrollarán contenidos los cuales se discutirán a través de foros específicos.

Los trabajos finales a presentar por los asistentes del curso serán supervisados a distancia.

**Sistema de evaluación:** Para aprobar el curso, los asistentes deben presentar un trabajo final de aplicación, el cual podrá realizarse en forma grupal (máximo de 2 personas). Este consistirá de la aplicación de las técnicas tratadas en el curso a un caso de estudio. El caso de estudio y las técnicas serán coordinadas previamente con los docentes.

Se permitirá una única re-entrega, si el trabajo no cumple con las exigencias establecidas por los docentes.

**Certificación:** Se entregará los siguientes certificados:

1. Certificado de asistencia: 75% de asistencia a clases.
2. Certificado de aprobación: 75% de asistencia a clases y aprobación de Trabajo final.

**Duración del curso:** 40 horas reloj.

**Distribución horaria:** El curso en su modalidad presencial, será dictado en 3 clases de 3 horas. En modalidad virtual, 11 horas y se destina 20 horas para el desarrollo del trabajo de aplicación.

**Conocimientos previos necesarios:** Java, Python, C++ o C (nivel intermedio), Algoritmos y Estructuras de Datos, y Probabilidad y Estadística.

///...

Q

N



*Universidad Nacional de Salta*

FACULTAD DE CIENCIAS EXACTAS  
Av. Bolivia 5150 - 4400 - Salta  
Tel. (0387)425-5408 - Fax (0387)425-5449  
Republica Argentina

...///-2-

Anexo I de la RESCD-EXA: 230/2017 - EXP-EXA: 8201/2017

### **Destinatarios del Curso**

- Alumnos avanzados de la carrera de Tecnicatura Universitaria en Programación (3er año)
- Alumnos avanzados de la carrera de Licenciatura en Análisis de Sistemas (4to año en adelante)
- Egresados de la Lic. en Análisis de Sistemas
- Docentes del Departamento de Informática

### **Detalle analítico de erogaciones**

- Traslado Bahía Blanca - Salta: \$10000.
- Gastos de alojamiento y viáticos: \$10000

Las erogaciones serán afrontadas con fondos propios del Departamento de Informática.

**Arancelamiento:** Sin arancel.

**Lugar y Fecha de realización:** El curso será dictado presencialmente en las instalaciones del Departamento de Informática, del 24 al 26 de agosto de 2017.

### **Cupo**

- Cupo máximo alumnos avanzados de LAS y TUP: 10
- Cupo máximo egresados de LAS y TUP: 5
- Cupo máximo Docentes del Departamento de Informática: 10

**Inscripción:** Las inscripciones se realizarán a través del sitio oficial de las Jornadas (ji.unsa.edu.ar). La aceptación de la inscripción al curso será analizada e informada por la Comisión Organizadora JIUNSa.

### **Programa:**

1. Introducción a los conceptos básicos de la minería de la Web
  1. Historia de la Web
  - Recuperación de la información
  - Caracterización formal de modelos de recuperación de la información
  - Métodos clásicos para la evaluación de sistemas de recuperación de la información
  - Técnicas de crawling, indexación y búsqueda en la Web
  - Directorios de páginas Web
  - Análisis de contenido
  - Análisis de enlaces
  - Análisis de uso
  - Datos estructurados y no estructurados
2. Técnicas de crawling
  - Estándares y protocolos
  - Herramientas de crawling
  - Crawlers generales y por tema
  - Ingeniería de crawlers de gran escala

Q

N

///...



Anexo I de la RESCD-EXA: 230/2017 - EXP-EXA: 8201/2017

3. Técnicas de indexación y búsqueda

- Construcción de índices
- Creación de consultas
- Optimización de consultas
- Expansión de consultas
- Técnicas de ranking
- Búsquedas incrementales y basadas en contexto

4. Temas especiales y aplicaciones

- Aprendizaje mecánico basado en la Web
- La Web invisible
- Toma de decisiones basadas en información extraída de la Web
- Agentes inteligentes en la Web
- Personalización y sistemas de recomendación
- Análisis de redes sociales y comunidades en la Web
- Conceptos básicos de la Web semántica
- Búsqueda temática

**Bibliografía**

Bibliografía Básica:

1. Mining of Massive Datasets. 2nd Edition. Jure Leskovec, Anand Rajaraman, Jeffery David Ullman. Cambridge University Press, 2014.
2. Information Retrieval: Implementing and Evaluating Search Engines. Stefan Büttcher, Charles L. A. Clarke, and Gordon V. Cormack. MIT Press, 2010.
3. Modern Information Retrieval. The concepts and technology behind search. Second Edition. R. Baeza-yates and B. Ribeiro-Neto, Addison-Wesley, New York, 2011.

Bibliografía Adicional:

1. Collective Intelligence in Action. Satnam Alag. Manning Publications Co. 2008.
2. Algorithms of the Intelligent Web Haralambos Marmanis and Dmitry Babenko. Manning Publications Co. 2009.
3. Search Engines: Information Retrieval in Practice. Bruce Croft, Donald Metzler, Trevor Strohman. Addison Wesley. 2010.
4. Introduction to information retrieval. By Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze Cambridge Press. 2009.
5. Programming Collective Intelligence. Building Smart Web 2.0 Applications. Toby Segaran. O'Reilly. 2007.
6. Mining the Web: Discovering Knowledge from Hypertext Data, Soumen Charkabarti, Morgan Kaufmann, 2002.
7. Data Mining: Practical Machine Learning Tools and Techniques. Ian H. Witten, Eibe Frank Morgan Kaufmann, 2005.
8. Artículos disponibles en la Web, en particular aquellos publicados en las siguientes conferencias: Internacional World Wide Web Conference, ACM WSDM, ACM HT, ACM SIGIR, ACM KDD y ACM CIKM.

\*\*\*\*\*

Mag. GUSTAVO DANIEL GIL  
SECRETARIO DE EXTENSIÓN Y BIENESTAR  
FACULTAD DE CS. EXACTAS - UNSa.



Dr. JORGE FERNANDO YAZLLE  
DECAÑO  
FACULTAD DE CS. EXACTAS - UNSa.