



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

GUÍA DOCENTE DE MINERÍA DE TEXTO

La presente guía docente corresponde a la asignatura Minería de Texto (TEXT), aprobada para el curso lectivo 2017-2018 en Junta de Centro y publicada en su versión definitiva en la página web de la Escuela Politécnica Superior. La guía docente de TEXT aprobada y publicada antes del periodo de matrícula tiene el carácter de contrato con el estudiante.



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

ASIGNATURA

MINERÍA DE TEXTO (TEXT)

1.1. Código

33087 del Máster en Bioinformática y Biología computacional

1.2. Materia

Ciencias de la Computación e Inteligencia Artificial

1.3. Tipo

Obligatoria

1.4. Nivel

Máster

1.5. Curso

1º

1.6. Semestre

1º

1.7. Número de créditos

3 ECTS

1.8. Requisitos previos

Para un seguimiento óptimo de la asignatura se precisa tener conocimientos a nivel básico en técnicas de programación.

1.9. Requisitos mínimos de asistencia a las sesiones presenciales

La asistencia es obligatoria al menos en un 80%.



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

1.10. Datos del equipo docente

Profesores:

Dr. Martin Krallinger

Unidad de Minería de Textos Biológica
Centro Nacional de Investigaciones Biológicas
Teléfono: +34 917328000. Ext. 3014
Correo electrónico: mkrallinger
Horario de tutorías: Petición de cita previa en clase o por correo electrónico.
Nota: se debe añadir @cnio.es a la dirección de correo electrónico

Dr. Manuel Sánchez-Montañés Isla (coordinador)

Departamento de Ingeniería Informática
Escuela Politécnica Superior
Despacho: B-303
Teléfono: +34 914972290
e-mail: manuel.smontanes
Horario de tutorías: Petición de cita previa en clase o por correo electrónico.
Nota: se debe añadir @uam.es a la dirección de correo electrónico

1.11. Objetivos del curso

Esta asignatura presenta los aspectos básicos de las técnicas de procesamiento de lenguaje natural (NLP) aplicadas a la literatura biomédica y a la minería de texto. El programa incluye una breve introducción a los principales temas de la NLP, así como los aspectos básicos que caracterizan la estructura de la literatura biomédica. La asignatura también introduce el uso de algunas de las herramientas de software más importantes, desarrolladas para manejar textos de biología molecular. La sesión práctica se dedica principalmente a proporcionar una visión general sobre el desarrollo y uso de la minería de texto y aplicaciones biomédicas.

Las competencias básicas y generales que el estudiante adquiere en esta asignatura son:

- CG1 - Capacidad para comprender y aplicar métodos y técnicas de investigación en el ámbito de la Bioinformática.
- CG2 - Capacidad para proyectar, calcular y diseñar productos bioinformáticos.
- CG3 - Capacidad para trabajar en equipos multidisciplinares, comunicándose eficientemente y desarrollando su actividad de acuerdo con las buenas prácticas científicas.
- CG4 - Capacidad para la investigación, desarrollo e innovación, en empresas y centros tecnológicos, en el ámbito de la Bioinformática.



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

- CG5 - Capacidad para la aplicación de los conocimientos adquiridos y resolución de problemas en entornos nuevos o poco conocidos en el ámbito de la Bioinformática.
- CG6 - Capacidad de búsqueda, análisis y gestión de información; incluyendo la capacidad de interpretación y evaluación con un razonamiento crítico y autocrítico.
- CG7 - Capacidad de estudiar y resolver problemas biológicos y biomédicos con el soporte de herramientas computacionales.
- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida auto-dirigido o autónomo.

Las cualificaciones ubicadas en el nivel de competencias transversales que el estudiante adquirirá en esta asignatura son:

- CT1 - Capacidad para trabajar en equipo de forma colaborativa y con responsabilidad compartida en el diseño y comunicación de estrategias experimentales.
- CT2 - Capacidad de identificar fuentes de información científica solventes para fundamentar el estado de la cuestión de un problema bioinformático y poder abordar su resolución.

La competencia de tecnología específica que el estudiante adquiere en esta asignatura es:

- CE6 - Capacidad de explotar tecnologías avanzadas de aprendizaje automático y minería de textos.



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

Al final del semestre (objetivos generales), y de cada unidad (objetivos por tema) el estudiante deberá ser capaz de:

OBJETIVOS GENERALES	
G1	Comprender y aplicar métodos y técnicas de investigación de text mining en el ámbito de la Bioinformática.
G2	Proyectar, calcular y diseñar productos bioinformáticos de text mining.
G3	Trabajar en equipos multidisciplinares, comunicándose eficientemente y desarrollando su actividad de acuerdo con las buenas prácticas científicas.
G4	Investigar, desarrollar e innovar en empresas y centros tecnológicos en el ámbito de la Bioinformática.
G5	Aplicar los conocimientos adquiridos y resolver problemas en entornos nuevos o poco conocidos en el ámbito de la Bioinformática.
G6	Buscar, analizar y gestionar información siendo capaz de interpretarla y evaluarla con un razonamiento crítico y autocrítico.
G7	Estudiar y resolver problemas biológicos y biomédicos con el soporte de herramientas computacionales de text mining.

OBJETIVOS ESPECIFICOS POR TEMA	
TEMA 1.	
1.1.	Comprender la importancia de los datos textuales libres en bioinformática y biomedicina
1.2.	Identificar los repositorios de literatura más importantes en biociencias
1.3.	Comprender los aspectos más básicos de las anotaciones en bases de datos, biocuración y minería de textos.
1.4.	Identificar los elementos básicos en la formalización de datos textuales libres.
TEMA 2.-	
2.1.	Comprender las bases de los sistemas de recuperación y extracción de información en bases de datos de texto libre
2.2.	Comprender y diseñar sistemas de clustering y clasificación automática de textos
2.3.	Comprender y diseñar sistemas de respuestas a preguntas y extracción de conocimiento
TEMA 3.-	
3.1.	Identificar y comprender los aspectos esenciales de los sistemas biomédicos de recuperación de información
3.2.	Construir sistemas de extracción de relaciones en bases de datos de textos biológicos



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

1.12. Contenidos del programa

Programa

I. Literatura Biomédica

- Importancia de los datos textuales libres en bioinformática y la biomedicina
- Características básicas de la literatura biomédica
- Repositorios de literatura para las ciencias de la vida
- Anotaciones en bases de datos, curación de la literatura (biocuración) y minería de textos
- Vocabularios controlados, ontologías y terminologías jerárquicas.

II. Procesamiento del lenguaje natural

- Recuperación de información
- Agrupamiento de texto
- Clasificación de texto
- Extracción de información
- Reconocimiento de Entidades Nominales (NER)
- Respuestas a Preguntas
- Descubrimiento de Conocimiento

III. Aplicaciones de los sistemas de minería de textos en biología (Bio-PNL)

- Sistemas biomédicos de recuperación de información
- Reconocimiento y normalización de entidades biológicas: genes, proteínas, ADN, ARN, células, especies, etc.
- Sistemas de extracción de relaciones: interacciones de proteínas, recuperación de anotaciones y relaciones indirectas.
- Ordenación de genes y clasificación de texto.

1.13. Referencias de consulta

2. Métodos docentes

- Clases Teóricas apoyadas con material multimedia
- Resolución de problemas o casos prácticos en el aula
- Prácticas asistidas por ordenador



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

3. Tiempo de trabajo del estudiante

		Nº de horas	Porcentaje
Presencial	Clases teóricas	16 h	23 h (30.7%)
	Clases prácticas	4 h	
	Tutorías	3 h	
No presencial	Estudio semanal	22 h	52 h (69.3%)
	Realización de actividades prácticas	10 h	
	Preparación del examen (convocatoria ordinaria)	10 h	
	Preparación del examen (convocatoria extraordinaria)	10 h	
Carga total de horas de trabajo: 25 horas x 3 ECTS		75 h	

4. Métodos de evaluación y porcentaje en la calificación final

En el periodo ordinario de evaluación, esta será realizada de acuerdo al siguiente esquema:

- 50 % Prácticas de laboratorio
- 50 % Examen final

Para poder realizar el examen final es necesario que el estudiante haya entregado todas las prácticas y además tenga en ellas una nota mínima de 5.

Es necesario alcanzar una nota mínima de 5 tanto en el examen final como en la nota media de prácticas para aprobar el curso.

En caso de que el estudiante no haya entregado todas las prácticas recibirá la calificación “No evaluado”.

En caso de que el estudiante no apruebe la asignatura en la convocatoria ordinaria, las notas de las partes individuales se conservan para la convocatoria extraordinaria.

En el periodo extraordinario de evaluación, esta será realizada también de acuerdo al siguiente esquema:



Asignatura: Minería de Texto
Código: 33087
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 3

- 50 % Prácticas de laboratorio
- 50 % Examen final

El estudiante tiene la oportunidad, aunque haya aprobado las prácticas en la convocatoria ordinaria, de volver a entregarlas mejoradas.

Es necesario alcanzar una nota mínima de 5 tanto en el examen final como en la nota media de prácticas para aprobar el curso.

Solo en el caso de que el estudiante no presente prácticas en la convocatoria extraordinaria ni se presente al examen de esta convocatoria recibirá la calificación “No evaluado” en esta.

5. Cronograma

Semana	Contenido	Horas presenciales	Horas no presenciales
1	Literatura Biomédica	8	10
2	Procesamiento del Lenguaje Natural	7	11
3	Aplicaciones de los sistemas de minería de textos en biología	8	11
	- Preparación Examen Final convocatoria ordinaria		10
	- Preparación Examen Final convocatoria ordinaria		10
	Examen Final convocatoria ordinaria	2	