

CD-1354-16.



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Planilla a completar para presentación de Cursos de Posgrado

1.- DEPARTAMENTO de COMPUTACIÓN

2.- NOMBRE DEL CURSO: .. Sistemas híbridos de extracción de información.....

3.- DOCENTES:

RESPONSABLE/S: .. Pablo Duboue.....

COLABORADORES:

AUXILIARES:

4.- CARRERA de DOCTORADO

5.- AÑO: 2016.....

CUATRIMESTRE/S: Segundo.....

6.- PUNTAJE PROPUESTO PARA CARRERA DE DOCTORADO: ½ pto.

7.- DURACIÓN (anual, cuatrimestral, bimestral u otra): .1 Semana.....

8.- CARGA HORARIA SEMANAL:

Teóricas:

Problemas:

Laboratorio:

Seminarios:

Teórico - Práctico: ..15.....

Salida a Campo:

9.- CARGA HORARIA TOTAL:15hs.....

10.- FORMA DE EVALUACIÓN: ..Trabajo final

11.- PROGRAMA ANALÍTICO:

Módulo 1

Extracción de información. Generalidades. Competencias de extracción de información.
Pipelines de procesamiento de lenguaje natural. Anotaciones off-stand. Sistemas de tipos.
Control de flujo. Serialización de anotaciones.

Módulo 2

Entidades nombradas. Generalidades. Sistemas basados en diccionario. Sistemas basados en desambiguación de sentidos. Bootstrapping usando diccionarios de arranque. Destilado de diccionarios a partir de datos abiertos.

Módulo 3

Extracción de información basados en reglas. Métodos basados en expresiones regulares. JAPE. RefO. Métodos basados en puntos de anclaje. RuTA. Inducción de reglas a partir de ejemplos anotados. Algoritmo Whisk. Algoritmo LP2. Algoritmo KEP. Metodología de programación y evaluación.

Módulo 4

Extracción de información basada en métodos estadísticos. Metodología Begin-Inside-Outside. Sistemas basados en entropía máxima. Sistemas basados en Conditional Random Fields. Criterios de anotación y evaluación.

Módulo 5

Sistemas híbridos de extracción de información. Errores en cascada. Caso de estudio. Puesta en producción y mantenimiento. Cierre.

12.- BIBLIOGRAFÍA:

1. Cunningham, Hamish. "Information extraction, automatic." Encyclopedia of language and linguistics, (2005): 665-677. <https://gate.ac.uk/sale/ell2/ie/preprint.pdf>
2. Documentacion de UIMA: <http://uima.apache.org/d/uimaj-current/index.html>
3. Documentacion Apache Ruta: <https://uima.apache.org/d/ruta-current/tools.ruta.book.html>

Opcional

Text Mining, Weiss et al. (2005), Chapter 6

- * <https://gate.ac.uk/sale/tao/splitch8.html>
- * <http://iepy.readthedocs.org/en/latest/tutorial.html>
- * <http://www.nltk.org/book/ch07.html>
- * <http://mallet.cs.umass.edu/fst.php>