Fecha de aprobación:

Departamento de Sistemas

PROGRAMA ANALÍTICO

Nivel MAESTRIA				Unidad de enseñanza-aprendizaje **TEMAS SELECTO DE COMPUTACIÓN III (Procesamiento de Lenguaje Natural con Java y Python)**	
Clave 1158038					
4.5	Horas teoría	0	Horas práctica	Seriación	Créditos 9

OBJETIVOS:

Al finalizar el curso el alumno será capaz de

1. Aplicar los principales conceptos de procesamiento de lenguaje natural (PLN) en dominios específicos.
2. Desarrollar programas y aplicaciones para procesamiento de lenguaje natural en el dominio académico y salud.
3. Describir y manejar librería para el PLN en español e inglés en Java y Python.
4. Implementar las tareas de PLN como clasificación de textos, reconocimiento de voz y sistemas de pregunta respuesta en Java y Pyhton.

CONTENIDO SINTÉTICO:

1. Introducción al Procesamiento de Lenguaje Natural (PLN) y pre-procesamiento de textos
2. Texto a características (Ingeniería de características en datos textuales)
3. Reconocimiento de la voz y sistemas de pregunta respuesta
4. Clasificación de textos
5. Sistema de pregunta respuesta en español
6. Clasificación de textos en inglés usando vectores de textos (*Word embedding*)

TEMA 1. Introducción al Procesamiento de Lenguaje Natural y pre-procesamiento de textos

OBJETIVOS ESPECÍFICOS:

1.1. Explicar los conceptos fundamentales del Procesamiento del lenguaje natural.
1.2. Explicar la importancia del PLN en los sistemas de información y sus aplicaciones actuales.

CONTENIDO:

1.2.1. Concepto de PLN

1.2.2. Tareas de PLN

1.2.2.1 Reconocimiento de voz

1.2.2.2 Sistemas de pregunta-respuesta

1.2.2.3 Clasificación de textos

1.2.2.4 Generación de resúmenes, explicaciones, similitud de textos, entre otros.

1.2.3. Pre-procesamiento de textos

1.2.3.1. Eliminación de ruido

1.2.3.2. Normalización de textos

1.2.3.3. Lematización

1.2.3.4. Stemming

1.2.3.5. Estandarización de textos

1.2.4. Aplicaciones actuales del PLN

REFERENCIAS:

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015

- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

HORAS DE CLASE:

4.5

OBSERVACIONES:

TEMA 2. Texto a características (Ingeniería de características en datos textuales)

OBJETIVOS ESPECÍFICOS:

1. Aplicar las técnicas de extracción de características en datos textuales.
2. Explicar e implementar modelos de representación de características.

CONTENIDO:

- 2.1 Concepto de características a partir de textos
- 2.2 Características estadísticas
  - 2.2.1 N-gramas como características
  - 2.2.2 Pesado booleano
  - 2.2.3 Frecuencia de términos
  - 2.2.4 TF-IDF
- 2.3 Características sintácticas
  - 2.3.1 Etiquetado POS
  - 2.3.2 Árbol de dependencias
- 2.4 Características semánticas
  - 2.4.1 Reconocimiento de entidades nombradas (NER)
  - 2.4.2 Modelado de temas
- 2.5 Vectores de textos (*Word embedding*) para grandes volúmenes de textos.
  - 2.5.1 Word2Vec
  - 2.5.2 GloVe
- 2.6 Características para sistemas de pregunta-respuesta
  - 2.6.1 Características de fragmentación (*chunking-like features*)
    - 2.6.1.1 Unigramas
    - 2.6.1.2 Bigramas
    - 2.6.1.3 Trigramas
  - 2.6.2 Características de preguntas de palabras (*question words*)

REFERENCIAS:

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015
- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

HORAS DE CLASE:

9.0

OBSERVACIONES:

TEMA 3. Reconocimiento de voz y sistemas de pregunta-respuesta

OBJETIVOS ESPECÍFICOS:

1. Identificar el concepto de reconocimiento de voz.
2. Identificar y describir los elementos de un sistema de pregunta respuesta
3. Usar librerías para el reconocimiento de voz en Java.

CONTENIDO:

3.1 Reconocimiento de voz
- 3.1.1 Reconocimiento de voz en español
- 3.1.2 Diseño de sistemas de reconocimiento de voz
- 3.1.3 Uso y aplicaciones de sistemas de reconocimiento de voz

3.2 Sistemas de pregunta respuesta
- 3.2.1 Concepto de sistemas de preguntas respuesta
- 3.2.2 Arquitectura de un sistema de pregunta-respuesta
  - 3.2.2.1 Análisis de la pregunta
  - 3.2.2.2 Recuperación de la información
  - 3.2.2.3 Extracción/formación de la respuesta
- 3.2.3 Diseño de sistemas de preguntas-respuestas

3.3 Librerías para el reconocimiento de voz en Java

REFERENCIAS:

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015
- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

HORAS DE CLASE:

4.5

OBSERVACIONES:

TEMA 4. Clasificación de textos

OBJETIVOS ESPECÍFICOS:

1. Explicar el concepto de clasificación de textos.
2. Identificar y describir los elementos de un sistema de clasificación de textos.
3. Usar librerías para la clasificación de textos en Java y Python.

CONTENIDO:

4.1 Clasificación de textos

4.2 Ingeniería de características para la clasificación de textos

4.3 Algoritmos de clasificación de textos

4.4 Evaluación de la tarea de clasificación de textos

4.5 Librerías para apoyo en la clasificación de textos en Pyhton

REFERENCIAS:

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015

- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

HORAS DE CLASE:

4.5

OBSERVACIONES:

TEMA 5. Sistema de pregunta-respuesta en español

OBJETIVOS ESPECÍFICOS:

1. Diseñar un sistema de pregunta respuesta en español para el dominio académico.
2. Implementar una primera aproximación de un sistemas de pregunta-respuesta.

CONTENIDO:

5.1 Diseño de sistemas de preguntas respuesta en el dominio académico.
- 5.1.1 Diseño de un sistema de preguntas respuesta en el dominio académico.
- 5.1.2 Módulos de un sistema de pregunta-respuesta en español para el dominio académico.
- 5.1.3 Características de un sistema de pregunta-respuesta en español para el dominio académico.

5.2 Primera aproximación de la implementación de un sistema de preguntas respuesta en el dominio académico
- 5.2.1 Traducción de voz a texto y viceversa
- 5.2.2 Ingeniería de características en el sistema de pregunta-respuesta
- 5.2.3 Análisis de la pregunta
- 5.2.4 Extracción de información
- 5.2.5 Formación de la respuesta

REFERENCIAS:

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015
- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

HORAS DE CLASE:

4.5

OBSERVACIONES:

TEMA 6. Clasificación de textos clínicos en inglés usando vectores de texto (*Word Embedding*)

OBJETIVOS ESPECÍFICOS:

1. Diseñar un clasificador de textos clínicos en inglés usando vectores de textos (Word Embedding).
2. Implementar una primera aproximación de un un clasificador de textos clínicos en inglés usando vectores de textos (Word Embedding).

CONTENIDO:

6.1 Diseño de clasificador de textos clínicos en inglés.

- 6.1.1 Módulos del clasificador de textos clínicos en inglés.
- 6.1.2 Características del clasificador de textos clínicos en inglés.

6.2 Primera aproximación de la implementación del clasificador de textos clínicos en inglés.

- 6.2.1 Pre-procesado de los textos clínicos
- 6.2.2 Ingeniería de características en los textos clínicos usando Word Embedding
- 6.2.3 Algoritmo de clasificación
- 6.2.4 Evaluación de la clasificación

REFERENCIAS:

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015
- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

HORAS DE CLASE:

4.5

OBSERVACIONES:

MODALIDADES DE CONDUCCIÓN DEL PROCESO DE ENSEÑANZA-APRENDIZAJE

Clase teórico-práctica con participación activa del alumno.
Como parte de las modalidades de conducción del proceso de enseñanza-aprendizaje será requisito que los alumnos con apoyo del profesor, participen en la revisión y análisis de al menos un texto técnico, científico o de difusión escrito en idioma inglés y que contribuya a alcanzar los objetivos del programa de estudios.

Se procurará que como parte de las modalidades de conducción del proceso de enseñanza-aprendizaje los alumnos participen en la presentación oral de sus trabajos, tareas u otras actividades académicas desarrolladas durante el curso.

INFORMACIÓN ADICIONAL

MODALIDADES DE EVALUACIÓN

Realizar evaluaciones periódicas y una evaluación terminal, consistentes en preguntas conceptuales y problemas escritos. La evaluación terminal podrá exentarse (a juicio del profesor).

No admite evaluación de recuperación.

No requiere inscripción previa.

Examen (Temas 1-4):	**20 %**
Proyecto* :	**30 %**
Exposiciones :	**20 %**
Artículo :	**30 %**

La escala de calificación es la siguiente:

0 - 5.9	→	NA
6 - 7.4	→	S
7.5 - 8.7	→	B
8.8 - 10	→	MB

BIBLIOGRAFÍA NECESARIA O RECOMENDABLE

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015
- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.

BIBLIOGRAFÍA ADICIONAL

- Natural Language Processing with Java. Richard M. Reese. Packt, open source. 2015
- Natural Language Processing with Pyhton (*Analyzing text with NLTK*), Steven Bird, Ewan Klein and Edward Loper, 2nd Edition, O'Reilly, 2016.