



## FACULTAD DE INFORMÁTICA

# TESINA DE LICENCIATURA

**TÍTULO:** Resumen extractivo de documentos. Un análisis comparativo de técnicas de puntuación

**AUTORA:** Julieta Pilar Corvi

**DIRECTORA:** Dra. Laura Lanzarini

**ASESOR PROFESIONAL:** Dr. Augusto Villa Monte

**CARRERA:** Licenciatura en Informática.

### Resumen

*En el presente existe una cantidad inimaginable de información digital en distintos formatos, la gran mayoría en formato texto. Esto ha incrementado el interés en el estudio de técnicas que permiten identificar de manera automática las partes relevantes y dan como resultado un resumen automático.*

*En esta tesina se desarrolló una técnica capaz de extraer las sentencias más representativas de un documento manteniendo el criterio de un usuario. El aprendizaje de ese criterio se realizó mediante una red neuronal, a partir de un conjunto de documentos previamente calificados por el usuario. La importancia de las oraciones se asigna en concordancia con la importancia que el usuario consideró para los documentos ingresados.*

### Palabras Clave

- Minería de Datos
- Minería de Texto
- Resumen automático de documentos
- Resumen extractivo
- Redes Neuronales
- Combinador lineal

### Conclusiones

- Esta tesina presenta un nuevo método capaz de generar de manera automática, a partir de resúmenes hechos por un usuario, resúmenes automáticos que se aproximen lo más posible como lo haría esa persona.
- Para verificar la metodología propuesta, se utilizaron 220 artículos científicos de la revista PLOS Medicine. Los resultados obtenidos han sido satisfactorios.

### Trabajos Realizados

- Análisis de los conceptos principales de la Minería de Datos y de la Minería de Texto.
- Descripción del proceso de extracción del texto, extracción de información del mismo y su estructuración.
- Estudio de las técnicas de generación de resúmenes automáticos, en especial, la manera extractiva.
- Definición, desarrollo e implementación de la red neuronal utilizada y análisis de los resultados.

### Trabajos Futuros

*A futuro se propone analizar si existe relación entre las métricas seleccionadas por el método y el tipo de documento o la temática. También interesa analizar la importancia del lenguaje en el que está escrito el documento así como del estilo utilizado por el autor. Por otra parte se propone ampliar las métricas utilizadas, incluyendo las de tipo de resumen abstractivo.*