

## PERTINENCIAS DE PLANES DE ESTUDIO DE CARRERAS DE INFORMATICA CON NORMATIVAS ESTABLECIDAS POR CONEAU

Lic. Laura Gutiérrez Mag. Raúl Klenzi,  
Departamento Informática (DI) / Facultad de Ciencias Exactas Físicas y Naturales  
(FCEFN) / Universidad Nacional de San Juan (UNSJ)  
Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", San Juan  
gutierrez.laura, rauloscarklenzi@gmail.com

### Resumen

Este trabajo analiza la similitud entre documentos de texto con los contenidos mínimos de los planes de estudio de las carreras Licenciatura en Ciencias de la Computación (LCC) y Licenciatura en Sistemas de Información (LSI) del Departamento de Informática (DI) de la FCEFN de UNSJ, con las pautas establecidas por CONEAU en la Resolución Ministerial N° 786/2009 con el objetivo de evaluar la pertinencia de los contenidos mínimos de los planes de estudio de las carreras con las áreas mencionadas en dicha resolución.

Se trata de determinar la afinidad o pertinencia de los planes de estudio con respecto a las áreas determinadas en la Resolución 786/2009, mediante la aplicación de la herramienta de software libre RapidMiner (RM) [7] utilizando sus módulos de modelado y minería de texto (TextMining-TM). Esta herramienta, como medida de similitud sintáctica entre documentos, permite la utilización de métricas y tareas de segmentación que serán comparadas desde el punto de vista de la calidad del resultado, con la finalidad de mejorar los planes de estudio de las carreras LCC y LSI del DI de la FCEFN de la UNSJ.

**Palabras clave:** TextMining, pertinencias, planes de estudio.

### Contexto

Esta línea de investigación se enmarca en el proyecto bianual 2011-2012 “**MINERÍA DE DATOS EN LA DETERMINACIÓN DE PATRONES DE USO Y PERFILES DE USUARIO**” código 21/E889 que se desarrolla en el ámbito de la FCEFN-UNSJ, aprobado por el Consejo de Investigaciones Científicas Técnicas y de Creación Artística (CICITCA), financiado por la propia Universidad y ajustado a evaluación externa.

Los datos sobre los que se trabaja en el proyecto, son relativos a las áreas de salud y farmacia como así también al área educación. En esta última área se trabajará con datos generados en el marco de la acreditación de las carreras del DI, del análisis de rendimiento académico de alumnos, planes de estudios de las carreras del citado departamento y datos inherentes a la Biblioteca de la FCEFN.

En este contexto, toda posible mejora en los planes de estudio de las carreras LCC y LSI del DI, será considerada positivamente. La presente propuesta de trabajo y línea de investigación se centra en la determinación de pertinencias de planes de estudio de las carreras LCC y LSI del DI que se dictan en la FCEFN de la UNSJ respecto de las pautas establecidas por CONEAU en la Resolución Ministerial N° 786/2009.

## Introducción

Conforme lo manifestado por la resolución 786/09 del Ministerio de Educación de la Nación quedan definidos los contenidos curriculares básicos correspondientes a las diferentes titulaciones del área de conocimiento informática y computación como se detalla brevemente en las primeras hojas de la mencionada resolución.

La citada resolución aprueba los contenidos curriculares básicos, carga horaria mínima, criterios de intensidad de formación práctica, estándares y nómina de actividades profesionales reservadas para las carreras correspondientes a los títulos de LICENCIADO EN CIENCIAS DE LA COMPUTACION, LICENCIADO EN SISTEMAS /SISTEMAS DE INFORMACION /ANALISIS DE SISTEMAS, LICENCIADO EN INFORMATICA, INGENIERO EN COMPUTACION e INGENIERO EN SISTEMAS DE INFORMACION /INFORMATICA.

Los contenidos curriculares básicos de las diferentes carreras abarcadas por la presente resolución, son evaluados según las definiciones explicitadas en los diferentes anexos (I, II, y III) que conforman dicha resolución.

En el presente trabajo se realiza una comparación entre **los contenidos mínimos de los planes de estudio** de las carreras LCC y LSI del DI de la FCFN de la UNSJ., vigentes hasta el año 2011, con los contenidos de las distintas **áreas** establecidas en la Resolución 786/2009. Se hace uso de técnicas de DM, TM, IR y de medidas de similitud, con el fin de encontrar **pertinencias de contenidos mínimos de planes de estudio con**

**respecto a las áreas**, para ello se utiliza el módulo Text Processing de la herramienta de software RM que permite llevar a cabo las tareas de preprocesamiento y determinación de pertinencias.

## Líneas de investigación y desarrollo

El presente trabajo se encuentra enmarcado en el proyecto “Minería de datos en la determinación de Patrones de uso y perfiles de usuarios”, las líneas de investigación que se detallan a continuación y siempre con el afán de que la investigación aplicada ayude a la toma de decisiones de la autoridad competente. Por un lado y desde el relevamiento de una encuesta asociada a un constructo que permite aproximar la capacidad de resiliencia, se ha encuestado a alumnos avanzados e ingresantes de las carreras del departamento informática con el objetivo que, mediante tareas de modelación se pueda aproximar el perfil de un alumno rezagado o posible desertor y que en función de ello las autoridades de la FCFN-UNSJ arbitren las medidas necesarias a nivel de tutorías o apoyaturas docentes y de pares que atenúen la deserción y el rezago.

Por otro y desde la aplicación de diferentes técnicas de minería de datos DM y text mining procesar títulos bibliográfico de la biblioteca midiendo la similitud sintáctica de los mismos con los contenidos de las diferentes carreras que se dictan en la FCFN-UNSJ y proponiendo una primera aproximación a la determinación de código Dewey para aquella bibliografía que no lo posee o es edición, en español, de la propia universidad.

El uso de métricas de minería de texto es también un camino seguido en la presente propuesta de trabajo y es un eslabón que pretende ir cerrando la cadena de aplicaciones que en el tema de satisfacción de usuarios de bibliotecas universitarias y desde el procesamiento de la encuesta a usuarios alumnos y docentes de la biblioteca encaró el grupo de investigadores pertenecientes al proyecto.

En la mayoría de los casos los trabajos se han llevado adelante mediante la utilización de herramientas de software libre del área de minería de datos.

En este caso los resultados que se presentan se han alcanzado mediante el uso de algoritmos de DM que posee la herramienta RapidMiner cuya última versión es la 5.3.005.

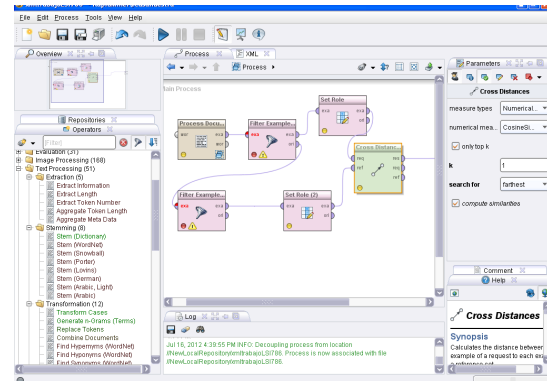
Para este trabajo se propone para contrastar documentos la técnica de la similaridad del coseno, que no es otra cosa que el coseno del ángulo que forman un vector consulta  $q$  (contenido mínimo) y un vector documento  $d_j$  (áreas de la 786/2009). [4]

A partir de una consulta se obtiene una lista de documentos ordenados por distancia (los más relevantes primeros). Luego, se procede a realizar los cálculos algebraicos para determinar la semejanza entre el vector consulta y cada uno de los vectores que representan documentos de la colección. [5]

### Preprocesamiento de datos

Al comenzar este trabajo se contaba con los contenidos mínimos de los planes de Estudio de las carreras LCC y LSI del DI en archivos .doc y la Resolución 786/2009 en un archivo .pdf. Lo primero que se realizó fue separar cada contenido mínimo de los distintos planes de estudio en archivos .txt, como así también se

llevaron a archivos .txt cada una de las áreas de la Resolución 786/2009.



Figural: Entorno de RapidMiner utilizando entre otros el módulo de Textmining.

En la Figural se observa las diferentes áreas en que se divide el entorno visual de la herramienta Rapidminer, y desde donde se pueden elegir los operadores a aplicar y definir los parámetros de los mismos.

Los documentos, que contienen los contenidos mínimos de planes de estudio de las carreras LCC y LSI (Consulta o request **req**) y los que contienen las áreas establecidas por CONEAU en la Resolución 786/2009 (Base de Datos de Referencia **ref**) son preprocesados por un módulo de RM [3]

Tras la instancia de preprocesamiento los documentos se separan en contenidos mínimos por un lado (Consulta o Request, **req**) y áreas establecidas en la Resolución 786/2009 (Referencia, **ref**) por otro. Esta separación o filtrado permite, desde el módulo (Cross Distances), la aplicación de métricas de similitud (la del coseno) entre documentos.

La métrica de similitud considerada, posee un rango de valores posibles que oscila en forma continua entre 0 o 1, cuando los documentos comparados son

sintácticamente diferentes, y 1 cuando reflejan una similitud total. [3]

En la Figura2, se observa lo dicho anteriormente.

Row No.	request	document	distanc
1	24- BASE DE DATOS AVANZADA	Ingeniería de Software Base de Datos y Sistemas de Información	0.129
2	4- INFORMACIÓN Y SISTEMA	Ingeniería de Software Base de Datos y Sistemas de Información	0.125
3	7- ASPECTOS PROFESIONALES Y SOCIALE	Aspectos profesionales y Sociales	0.122
4	21- PRINCIPIOS DE INGENIERIA DE SOFTWARE	Ingeniería de Software Base de Datos y Sistemas de Información	0.112
6	19- BASE DE DATOS	Ingeniería de Software Base de Datos y Sistemas de Información	0.112
6	25- DISEÑO DE SOFTWARE	Ingeniería de Software Base de Datos y Sistemas de Información	0.070
7	14- ESTRUCTURA Y FUNCIONAMIENTO DE LAS COMPUTADORAS I	Arquitectura, Sistemas Operativos y Red	0.061
8	27- ARQUITECTURAS AVANZADAS DE COMPUTADORA	Arquitectura, Sistemas Operativos y Red	0.056
9	29- REDES	Arquitectura, Sistemas Operativos y Red	0.047
10	3- ESTRUCTURA Y FUNCIONAMIENTO DE LAS COMPUTADORAS	Arquitectura, Sistemas Operativos y Red	0.044
11	19- INTRODUCCIÓN A LOS SISTEMAS DIGITALES I	Arquitectura, Sistemas Operativos y Red	0.031
12	8- INTRODUCCIÓN A LOS SISTEMAS DIGITALES	Arquitectura, Sistemas Operativos y Red	0.023
13	13- ESTRUCTURAS DE DATOS Y ALGORITMOS	Ingeniería de Software Base de Datos y Sistemas de Información	0.021
14	26- INTELIGENCIA ARTIFICIAL	Arquitectura, Sistemas Operativos y Red	0.019
15	9- PROGRAMACIÓN ORIENTADA A OBJETOS	Arquitectura, Sistemas Operativos y Red	0.015
16	5- PROGRAMACIÓN PROCEDURAL	Ingeniería de Software Base de Datos y Sistemas de Información	0.015
17	29- COMPILADORES	Arquitectura, Sistemas Operativos y Red	0.013
18	23- TEORÍA DE LA COMPUTACIÓN	Aspectos profesionales y Sociales	0.011
19	16- MATEMÁTICA DISCRETA	Arquitectura, Sistemas Operativos y Red	0.010
20	1- ALGORITMOS Y RESOLUCIÓN DE PROBLEMA	Ingeniería de Software Base de Datos y Sistemas de Información	0.010
21	22- INOLESI	Ingeniería de Software Base de Datos y Sistemas de Información	0.010
22	16- PARADIGMAS DE LENGUAJE	Arquitectura, Sistemas Operativos y Red	0.009
23	3- MATEMÁTICA BÁSIC	Aspectos profesionales y Sociales	0.009

Figura2 LCCcon-min LCC786

## Resultados y Objetivos

Con los objetivos iniciales de comparar los planes de estudio de las carreras LCC y LSI del DI con los contenidos de las áreas establecidas por CONEAU en la Resolución Ministerial N° 786/2009, utilizando el módulo de Text Processing de RM, y encontrar pertinencias se han logrado los siguientes resultados preliminares:

- De la utilización de la medida de similitud del Coseno se pudo determinar cuán pertinente es el contenido mínimo de los planes de estudios de las carreras LCC y LSI con las áreas establecidas en la Resolución Ministerial N° 786/2009, ya que ante documentos cuya pertinencia se aproxima a 1 indica lo cercano de los contenidos mínimos con las áreas consideradas en la Resolución Ministerial N° 786/2009, sus valores fluctúan entre 0 (sin coincidencia) y 1 (coincidencia total).
- Se logró una buena aproximación entre la tarea de segmentación que encuentra los grados de pertinencia y las áreas

consideradas por los docentes en la realización de los planes de estudio.

Se pretende desde este proyecto profundizar este análisis de pertinencias semánticamente, utilizando herramientas de textmining proporcionadas por RM, haciendo uso del módulo del Diccionario de Wordnet para sinónimos, el cual se encuentra en inglés y como nuestro dominio de trabajo es en español, se está trabajando en los pasos a seguir

## Formación de Recursos Humanos

La formación de recursos humanos es un tema de vital importancia en marco del presente proyecto de investigación. En este marco y en temas afines a las temáticas abordadas en el mismo, se están dirigiendo un conjunto de trabajos finales de grado de alumnos becarios de becas de finalización de carrera de la Agencia Nacional de Promoción Científica y Tecnológica.

En particular trabajos finales de grado y tesis de posgrado, se está trabajando con datos de títulos bibliográficos pertenecientes a la biblioteca de la FCFN-UNSJ a los que en forma automática se intenta asignar la numeración dewey asociada a la temática abordada por el material bibliográfico bajo estudio. Así mismo y desde una tesis de posgrado, se intenta encontrar automáticamente también, y mediante similitud sintáctica qué contenidos mínimos establecidos en la resolución 786/09 son contenidos y brindados por los planes de estudios de las carreras de informática de la FCFN-UNSJ

**Licenciatura en Sistemas de Información**

- *Estrategias de marketing web y minería de datos para promover tráfico de calidad hacia un sitio web.* Rafa
- *Minería web a los datos de acceso a sitios pertenecientes a centros de información académica.* Vero
- *Minería de Datos en la segmentación y clasificación de un banco de germoplasma.* Kari

**LCC**

- *Minería de Texto en la determinación automática de código Dewey.* J.Araya
- *Analítica web en centros de información.* S. Lobo
- *Desarrollo de un sistema de información Intranet para el Inst. de Ing. Química-UNSJ.* A. Sepúlveda

**LCI**

- *Técnicas de Minería de Texto en la determinación de la correlación entre valoración numérica y comentarios asociados en encuestas.*

A su vez se están dirigiendo dos tesis de posgrado en el área de Minería de Texto y el área de Inteligencia de Negocios respectivamente.

**Referencias**

- [1] Kantardzic, M (2003) "Data Mining: Concepts, Models, Methods, and Algorithms" ISBN:0471228524 John Wiley & Sons © (343 pages)
- [2] Klenzi, R. Tesis de posgrado de maestría "Aplicación de minería de datos a la gestión bibliotecaria". Biblioteca FCFN-UNSJ. 2008.
- [3] Klenzi, R, Gutiérrez L., Villafañe V. "Técnicas de recuperación de información en la determinación de pertinencias bibliográficas" 2012.
- [4] Liu B., "Web DataMining. Exploring Hyperlinks, Contents, and Usage Data" Springer-Verlag Berlin Heidelberg 2007
- [5] Manning C, Prabhakar R. Hinrich & Hinrich Schütze, "An Introduction to Information Retrieval ", Cambridge University Press. 2009.
- [6] Min, S; Yi-Fang B. Handbook of Research on Text and Web Mining Technologies -Information science reference- Editorial Advisory Board 2009.
- [7] Rapid-I. <http://rapid-i.com> ver.5.3.005 de 2013.
- [8] Tolosa G y Bordignon F., "Introducción a la Recuperación de Información. Conceptos, modelos y algoritmos básicos" UNLu, Arg. 2007.