

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

**FACULTAD INGENIERÍA
PROGRAMA DE SISTEMAS
BOGOTÁ D.C.**

LICENCIA CREATIVE COMMONS:

Atribución	<input checked="" type="checkbox"/>	Atribución no comercial	<input type="checkbox"/>	Atribución no comercial sin derivadas	<input type="checkbox"/>
Atribución no comercial compartir igual	<input type="checkbox"/>	Atribución sin derivadas	<input type="checkbox"/>	Atribución compartir igual	<input type="checkbox"/>

AÑO DE ELABORACIÓN: 2018

TÍTULO: Desarrollo de componente web parametrizable de clustering para análisis de datos.

AUTOR (ES): Alba Carlos, Rubio Joseph.

DIRECTOR(ES)/ASESOR(ES): Rincon Diego.

MODALIDAD:

Trabajo de investigación-

PÁGINAS:	122	TABLAS:	6	CUADROS:		FIGURAS:	17	ANEXOS:	3
-----------------	------------	----------------	----------	-----------------	--	-----------------	-----------	----------------	----------

CONTENIDO:

INTRODUCCIÓN

1. GENERALIDADES
 2. OBJETIVOS DEL PROYECTO
 3. MARCO REFERENCIAL
 4. MARCO CONCEPTUAL
 5. METODOLOGÍA
 6. DESARROLLO DEL PROYECTO
 7. RESULTADOS
 8. CONCLUSIONES
- BIBLIOGRAFÍA
ANEXOS



DESCRIPCIÓN: Se desarrolló un Componente Web de Clustering para Análisis de Datos el cual permite ingresar mediante una interfaz web y/o una API un conjunto de datos que son analizados mediante técnicas de clustering, estas técnicas permiten realizar agrupamiento de datos basados en sus características, y así arrojar un resultado que relaciona dichos datos. Lo anterior, busca proveer a la Universidad de una herramienta que permita de manera general, práctica y gratuita, realizar análisis de datos basados en clustering para cualquier estudio de investigación.

METODOLOGÍA Para desarrollar el componente de software se siguió la metodología de desarrollo XP (Extreme Programming), la cual está fundada en la agilidad y simplicidad, que a diferencia de las metodologías clásicas o tradicionales, busca reducir el número de procesos replazándolos por el mínimo de procesos cuyo esfuerzo valga la pena, con el objetivo de tener un buen desarrollo de software, más allá de una buena documentación.

En la metodología XP normalmente se diferencian cuatro fases (José Joskowicz, 2008) las cuales son:

- Fase de exploración.
- Fase de planificación.
- Fase de iteraciones.
- Fase de puesta en producción.

PALABRAS CLAVE: Análisis de Datos, Métodos de Clustering, Arquitectura de Microservicios, API, agrupamiento.

CONCLUSIONES: Sobre el desarrollo de este proyecto para el Componente Web Parametrizable de Clustering para Análisis de Datos, se puede concluir lo siguiente:

Actualmente, en el mercado no existe un componente de software que permita realizar clustering integrado a otras aplicaciones o que pueda ser ejecutado por sí solo. Así mismo, las técnicas para análisis de datos son herramientas muy efectivas al momento de extraer información relevante que permita, basado en datos, comprender y/o agrupar las características de los mismos.

El clustering es una de las técnicas más idóneas para encontrar nueva información dentro de un conjunto de datos, que permita generar un nuevo conocimiento. Para realizar el clustering se debe tener debidamente preparados y estructurados los



datos analizar, es importante tener en cuenta que no se pueden analizar datos no estructurados, ejemplo: videos, audios, imágenes, etc.

La mayoría de los algoritmos de clustering pueden ser organizados dentro de una interfaz común, ya que el conjunto de datos de entrada y sus parámetros son similares, lo que facilita el desarrollo de un componente genérico de software basado en clustering. Desacoplar la lógica de la aplicación y la lógica del clustering permite una alta reusabilidad y escalabilidad, ya que son ambas son independientes.

Usar el lenguaje de programación R es de gran ayuda en el proceso de desarrollo del software y en especial para el clustering, ya que la mayoría de las funcionalidades que se requieren para la implementación ya se encuentran creadas. En el mismo sentido, utilizar la arquitectura de micro servicios permite que el componente pueda ser usado a gran escala en aplicaciones más robustas. Finalmente, usar la metodología de desarrollo XP facilita en gran medida el cumplimiento de los compromisos estipulados en los requerimientos de software.

FUENTES:

- Andrea, A. D. ', Ferri, F., & Grifoni, P. (2015). *Approaches, Tools and Applications for Sentiment Analysis Implementation. International Journal of Computer Applications* (Vol. 125). Retrieved from <http://messenger.yahoo.com/features/emoticons>
- Andrew McAfee; Erik Brynjolfsson. (2012). *Big Data The Management Revolution*. Retrieved from <https://pdfs.semanticscholar.org/02c7/740af5540f23a2da23d1769e64a8042ec62e.pdf>
- Cathy O'Neil, R. S. (2013). *Doing Data Science_ Straight Talk from the Frontline-O'Reilly Media* (2013).
- Doug Laney. (2001). *Application Delivery Strategies*. Retrieved from <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (n.d.). *Cluster Analysis 5th addition2011*. <https://doi.org/10.1007/BF00154794>
- Garner, S. R. (n.d.). *WEKA: The Waikato Environment for Knowledge Analysis*. Retrieved from <https://www.cs.waikato.ac.nz/~ml/publications/1995/Garner95-WEKA.pdf>
- Gollapudi, S. (n.d.). *PRACTICAL MACHINE LEARNING*.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (n.d.). *The WEKA Data Mining Software: An Update*. Retrieved from https://www.kdd.org/exploration_files/p2V11n1.pdf



- Hernández-Leal, E. J., Duque-Méndez, N. D., Moreno-Cadavid, J., Hernández-Leal, E. J., Duque-Méndez, N. D., Moreno-Cadavid, J., & Big, ". (2015). *Big Data: una exploración de investigaciones, tecnologías y casos de aplicación* *Big Data: an exploration of research, technologies and application cases* (Vol. 20). Retrieved from <http://www.redalyc.org/pdf/3442/344251476001.pdf>
- IEEE. (2009). *IEEE Std 1016-2009 (Revision of IEEE Std 1016-1998), IEEE Standard for Information Technology—Systems Design—Software Design Descriptions. Middle East* (Vol. 2009).
<https://doi.org/10.1109/IEEESTD.2009.5167255>
- ISTQB. (2018). International Software Testing Qualifications Board ISTQB® Certified Tester: Foundation Level Extension Syllabus Agile Tester, 1–77.
- Jackson, N. (2017). *Building Microservices with Go*. O'Reilly Media.
- Jagla, B., Wiswedel, B., & Coppée, J.-Y. (2011). Extending KNIME for next-generation sequencing data analysis. *BIOINFORMATICS APPLICATIONS NOTE*, 27, 2907–2909. <https://doi.org/10.1093/bioinformatics/btr478>
- José Joskowicz. (2008). *Reglas y Prácticas en eXtreme Programming*. Retrieved from <https://ie.fing.edu.uy/~josej/docs/XP - Jose Joskowicz.pdf>
- Kabacoff, R. I. (2011). *R IN ACTION: Data analysis and graphics with R. Livro*.
<https://doi.org/citeulike-article-id:10054678>
- KAUSHIK, S. (2016). An Introduction to Clustering & different methods of clustering. Retrieved September 30, 2018, from
<https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>
- Lin, N. P., Chang, C. I., Chueh, H. E., Chen, H. J., & Hao, W. H. (2008). A deflected grid-based algorithm for clustering analysis. *WSEAS Transactions on Computers*, 7(3), 125–132.
- Mitchell, T. M. (1997). *Machine Learning*. (Reviews, Ed.).
- Moujahid, A. (n.d.). Clustering, 1–17.
- RapidMiner | Sistemas de Minería de Datos | Software de Minería de Datos. (n.d.). Retrieved August 31, 2018, from
<https://www.microsystem.cl/plataforma/rapidminer/>
- Requirements, Q. (2012). *International Standard Iso / Iec* (Vol. 25021).
<https://doi.org/10.1109/IEEESTD.2015.7106438>
- Richards, M. (2015). *Software Architecture Patterns*. (O. Media, Ed.).
- Software, A., & Kruchten, P. (2006). Planos Arquitectonicos : El Modelo de “ 4 + 1 ”
Vistas de la La Arquitectura L ´, 12(6), 1–16.
- Timmerman, B. (2012). Flow measurements in patient-specific conducting airways models. *16th Int Symp on Applications of Laser Techniques to Fluid Mechanics*, 2(3), 9–12. <https://doi.org/10.1109/TETC.2014.2330519>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

Una revisión de los algoritmos de partición más comunes en el análisis de conglomerados: un estudio comparativo. (2010).

LISTA DE ANEXOS:

- ANEXO 1 (SRS) documento de especificación de requerimientos de software
- ANEXO 2 (SDD) documento de diseño de software
- ANEXO 3 (SAD) documento de arquitectura de software