

**Método de búsqueda de información en redes
sociales de graduados sin información de contacto
de instituciones de educación superior**

Ing. Fabio Enrique Quintero DiazGranados

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría Gestión de Información
Bogotá D.C., Colombia
Enero de 2021**

Método de búsqueda de información en redes sociales de graduados sin información de contacto de instituciones de educación superior

Fabio Enrique Quintero DiazGranados

Trabajo de grado para optar al título de
Magíster en Gestión de Información

Director
Ing. Oswaldo Castillo Navetty
Decanatura Ingeniería de Sistemas

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría en Gestión de Información
Bogotá D.C., Colombia
Enero de 2021**

© Únicamente se puede usar el contenido de las publicaciones para propósitos de información. No se debe copiar, enviar, recortar, transmitir o redistribuir este material para propósitos comerciales sin la autorización de la Escuela Colombiana de Ingeniería. Cuando se use el material de la Escuela se debe incluir la siguiente nota “Derechos reservados a Escuela Colombiana de Ingeniería” en cualquier copia en un lugar visible. Y el material no se debe notificar sin el permiso de la Escuela.

Publicado en 2020 por la Escuela Colombiana de Ingeniería Julio Garavito. Avenida 13 No 205-59 Bogotá. Colombia
TEL: +57 – 1 668 36 00

Página de aceptación del jurado

El trabajo de grado de maestría titulado “Método de búsqueda de información en redes sociales de graduados sin información de contacto de instituciones de educación superior”, presentado por Fabio Enrique Quintero DiazGranados, cumple con los requisitos establecidos para optar al título de Magíster en Gestión de información.

Himelda Palacios

Olga Lucía Giraldo

Ing. Oswaldo Castillo Navetty

Bogotá, D.C., 22 de Febrero de 2021

Dedicatoria

A mis padres por haberme forjado como una persona de bien, gracias a su apoyo he logrado ser quien soy; a mi esposa quien ha tenido que afrontar desafíos en ausencia de mi colaboración y que debo destacar, ya que los ha logrado maravillosamente y, especialmente a mi hijo, quien es mi gran motivación.

Agradecimientos

Agradezco infinitamente al Ingeniero Oswaldo Castillo por dirigir este proyecto, por su paciencia y dedicación; a la Ingeniera Victoria Eugenia Ospina, sin ella lograr este proyecto no hubiese sido posible.

Resumen

Este proyecto propone un método de búsqueda de datos de graduados sin información de contacto de Instituciones de Educación Superior (IES), basado en consultas indexadas de datos públicos en Redes Sociales (RRSS) como Facebook o LinkedIn para facilitar una primera interacción y la alimentación de información de los egresados. La información consultada fue clasificada para determinar su pertinencia y veracidad como filtro previo a su almacenamiento. La metodología es de una investigación aplicada donde el análisis de acceso de información desde las redes sociales (fuente primaria) y los buscadores (fuente secundaria), determinó el desarrollo del proyecto, ya que el acceso a la información y la consecución de permisos fueron limitantes importantes que en efecto redujeron la obtención de la información básica de los graduados. Como resultados se identificó que, más allá de las dificultades de acceder a la información, el método de consecución de datos de contacto de los graduados de IES a partir de las RRSS, permitió obtener otros datos entregados por los buscadores que pueden ser sujetos para estudios posteriores como por ejemplo las empresas contratantes o en las que ha estado vinculado, sectores económicos o lugares de residencia. Como conclusión de este proyecto aplicado se identificó que las IES pueden considerar y emplear las RRSS como una herramienta adicional que les permitirá cerrar la brecha con sus graduados y así, reducir el número de graduados de quienes no tienen información de contacto y de esta forma, realimentar sus bases de datos para tener otros mecanismos de aprovechamiento de la información tanto en la oficina de graduados, como de otras instancias dentro de las IES para mejorar y enriquecer la oferta académica a partir de la retroalimentación de sus egresados.

Abstract

This Project propose an information gathering method of graduates without contact information of High Education Institutions, based on indexed searches, using social media data like Facebook or/and LinkedIn to facilitate a first interaction and feed the information of graduates, consequently, a classification method to in order to determine the pertinence and veracity of that information as first filter previous to the storage of that data.

The methodology appears up from apply research where the analysis of information access from the social media (primary source) and searchers (secondary source), determined the development from this project since the information access and the permissions achievement were very important limits which in fact reduced the basic data collection from graduates.

As result was identified that, beyond data access difficulties, the gathering method of information of graduates of High Education Institutions without contact information as of social media allowed get other data provided from the searchers which can be taken to further projects like companies or in which it had been linked, economical sector or residence places.

As a conclusion from this project has been identified that the High Education Institutions can consider and use social media as an additional tool to allow to close the gap with the graduates and then reduce the number of graduates without contact information, this way feeding back their databases to gain other mechanisms to take advantage of that information in the graduate's offices or other institution dependencies to improve the academic offer based on the graduate's feedback.

Tabla de contenido

1	Introducción	1
1.1	Objetivos	4
1.1.1	Objetivo general	4
1.1.2	Objetivos específicos	4
2	Estado del arte	5
3	Metodología	11
3.1	Fase 1. Selección de documentación base	12
3.2	Fase 2. Revisión del alcance	12
3.3	Fase 3. Implementación	13
3.4	Fase 4. Método y documentos	14
4	Resultados y contribución	15
4.1	Estado actual	15
4.1.1	Análisis de amenazas y debilidades	16
4.2	Método de búsqueda de información	17
4.3	Aplicación del método	24
4.3.1	Selección de la población objetivo	24
4.3.2	Configuración y acceso al buscador	25
4.3.3	Iterar sobre la población objetivo	25
4.3.4	Combinaciones de nombres	25
4.3.5	Desarrollo de la consulta en el buscador	26
4.3.6	Rúbrica	30
4.3.7	Evaluación de los resultados	30
4.3.8	Almacenamiento de los resultados	31
5	Conclusiones y recomendaciones	35
	Referencias	37
	Abreviaciones	39

Lista de Figuras

Figura 4-1 Método de búsqueda de información	22
Figura 4-2 Ejemplo archivo CSV de la población objetivo	23
Figura 4-3 Ejemplo código iteración sobre la población objetivo.....	25
Figura 4-4 Ejemplo combinación nombre cuatro partes	25
Figura 4-5 Ejemplo combinación nombre tres partes	265
Figura 4-6 Búsqueda programable palabras clave	26
Figura 4-7 Búsqueda programable configuraciones	27
Figura 4-8 Búsqueda programable sitios web en los que buscar.....	27
Figura 4-9 Obtenga los permisos y recursos	287
Figura 4-10 Ejemplo API Key.....	28
Figura 4-11 Ejemplo del Id del buscador	28
Figura 4-12 Ejemplo del resultado de la búsqueda	¡Error! Marcador no definido.
Figura 4-13 Código del método para evaluar los resultados.....	30
Figura 4-14 Ejemplo de la evaluación de los resultados.....	30
Figura 4-15 Ejemplo de almacenamiento de datos en una base de datos no relacional.....	31
Figura 4-16 Cantidad de resultados generados con 65% de precisión.....	32
Figura 4-17 Otras estadísticas con 65% de precisión	33
Figura 4-18 Cantidad de resultados generados con 85% de precisión.....	32
Figura 4-19 Otras estadísticas con 85% de precisión	33

Lista de Tablas

Tabla 4-1 Ejemplos de combinaciones de nombres según la cantidad de partes del nombre....	19
---	----

1 Introducción

Las Instituciones de Educación Superior (IES) deben promover la existencia de estrategias de seguimiento de los graduados para involucrarlos con la institución y realimentar los programas académicos con su experiencia y aprendizaje; y así, de esta forma, hacer que el graduado se vea directamente involucrado, fortaleciendo y participando en la misión institucional¹. Adicionalmente, con esta información las IES deberán demostrar que realizan actividades de mejoramiento de la calidad y la pertinencia de la oferta académica, como también el contar con información y herramientas para acompañar a los nuevos graduados en el proceso de inserción laboral y al retorno curricular de los programas desde su nueva experiencia. De esta forma las IES demostrarían que mediante el uso y aplicación de los mecanismos disponibles, la interacción con los graduados genera un impacto positivo en sus funciones.²

3

El reto para las IES es lograr mantener el vínculo con sus graduados usando las herramientas necesarias, aunque es difícil dar un valor certero, debido a que no se dispone de una cifra general y tampoco el ente de control cuantifica este dato en las reuniones generadas con las oficinas de graduados, de la cantidad de graduados los cuales pierden contacto con sus IES.

Sumado a esto, la experiencia de haber trabajado por algunos años en una IES, permite considerar que, en promedio, se actualiza solo un 10% la información de la población de graduados (teniendo en cuenta que en el Manual para estudios de seguimiento de graduados universitarios, se estipula que el porcentaje de recaudación este por encima del 25%)⁴. Frente a

¹ Ministerio de Educación Nacional (MEN). Decreto 1330 (25 de julio de 2019). Por el cual se sustituye el Capítulo 2 y se suprime el Capítulo 7 del Título 3 de la Parte 5 del Libro 2 del Decreto 1075 de 2015 – Único Reglamentario del Sector Educación. Bogotá, Artículo 2.5.3.2.3.1.5.

² Fondo de Desarrollo de la Educación Superior. Acuerdo 01 de 2018 - CESU. Por el cual se actualizan los lineamientos para acreditación de alta calidad de programas de pregrado. Bogotá. 2018, Capítulo 4, Artículo 6, Numeral 3.

⁴ SCHOMBURG, Harald. Manual para estudios de seguimiento de graduados universitarios, Kassel, 2004.

esto último, hay que tener en cuenta que algunas IES estarán arriba de esta media, mientras que otras se considerarán optimistas con obtener al menos el 10%.

El acceso a la información de las redes sociales (RRSS) está bastante restringido debido a los recientes acontecimientos sobre el uso que se les han dado a los datos alojados allí, ya que cualquier acceso a estos fuera de las interfaces que los proveen, está totalmente prohibido. Además, haciendo uso de las interfaces API que deben contar con los permisos necesarios y manejan una serie de pasos y aprobaciones para presentar una aplicación en las plataformas de desarrollo y alcanzar la aprobación de una red social en específico se deben cumplir con las políticas de uso y acceso a la información.

El método propuesto en esta investigación, brinda un margen de ventaja en el empleo de diferentes herramientas para la consolidación de información de los graduados, y así, que las unidades de graduados logren un primer contacto mediante las estrategias necesarias.

En general, las RRSS se presentan como una estructura interactiva de gran valor para las empresas y organizaciones, puesto que la información que se obtiene de estas redes son minas de oro para procesos de mercadeo, fidelización de clientes, entre otras ventajas. En el caso de las IES, es de destacar que desde hace mucho tiempo debieron tomar la alternativa de acoplarse a los beneficios que trae las redes sociales en temas económicos, mercadeo, publicidad, etc.; pero el contexto de saber y conocer la situación actual de sus graduados está básicamente inexplorado, tanto así que el Ministerio de Educación Nacional no cuenta con propuestas a nivel de estudios primarios al respecto, y solo se limita a recomendar mecanismos diferentes a las RRSS para la actualización de datos necesarios de esta población.

Los procesos de actualización de datos en algunos sectores económicos son vitales y valiosos ya que brindan un valor agregado a los productos, y a mediano plazo incrementan la calidad del producto o del servicio ofrecido al cliente, por ejemplo en el sector financiero la actualización de datos es importante, debido a que un producto varía de un cliente a otro a partir del riesgo que puede afrontar o no, además, para saber si un cliente llega a ser potencialmente apto o no

para un producto en particular, el cliente está ranqueado según los datos almacenados, por tanto, si el cliente presenta alguna mejora salarial, podrá obtener mayores beneficios del producto financiero. Las instituciones recomiendan que las actualizaciones de datos de sus clientes se generen entre seis meses y no sea mayor a doce meses. A nivel general, las dificultades para las organizaciones en este aspecto consisten en que los clientes no disponen del suficiente interés en realizar el proceso de actualización de datos, y como consecuencia se presentan dos problemáticas: 1) niveles bajos en los procesos de actualización de datos y 2) datos basura que se generan al momento que el cliente realiza la actualización debido a las pocas validaciones establecidas en las herramientas tecnológicas.

Si bien, por cualquiera de las problemáticas expuestas anteriormente u otras que pueden llegar a presentarse para el caso de los graduados, quienes en ocasiones pierden el contacto con la IES, como también las instituciones mediante su oficina de graduados que no puede localizar al graduado a partir de la información almacenada en su sistema; es aquí donde se reconoce la fehaciente necesidad de las IES para identificar aquellos graduados que no fue posible contactar y para quienes los procesos de actualización de datos no fueron satisfactorios, por tanto, se requiere la aplicación de otra estrategia de actualización diferente para mejorar la relación de los graduados con las IES, y en últimas, identificar las fortalezas y debilidades de la formación académica ofrecida por las IES.

Por consiguiente, la pregunta que motivó el desarrollo de esta investigación es: ¿será posible lograr un mecanismo que realmente permita incrementar la tasa de información actualizada por la oficina de graduados de las instituciones de educación superior?

Para dar respuesta a esta pregunta, el presente documento se encuentra organizado en cinco capítulos. Se inicia presentando el objetivo general y los objetivos específicos, para luego exponer el estado del arte donde se presenta un estudio sobre el uso en general de RRSS, su crecimiento a nivel global y nacional y su aplicación en el ámbito académico y en las IES como herramienta para apoyar los procesos educativos. Igualmente se aborda el uso de los buscadores y aplicaciones a las búsquedas automáticas y las dificultades en los procesos de actualización de

datos. En el tercer capítulo se expone la metodología y los pasos necesarios para realizar la actualización de datos de graduados de las IES sin información de contacto de forma satisfactoria. En el cuarto capítulo se presentan los resultados, donde se observa la aplicación del método propuesto y las diferentes opciones según las necesidades que se tengan en la IES. Finalmente, el documento cierra con un quinto capítulo donde se consignan las principales conclusiones y recomendaciones a las cuales se ha llegado a partir de la investigación realizada.

1.1 Objetivos

En el presente apartado se exponen los objetivos, tanto general como específicos con los cuales se orientó el desarrollo del presente proyecto.

1.1.1 Objetivo general

Generar un método de obtención de información pública a partir de las redes sociales (RRSS), que permita a las instituciones de educación superior (IES) recopilar un mínimo de datos de sus graduados sin información de contacto y facilite generar una primera interacción.

1.1.2 Objetivos específicos

- Determinar las características de información necesarias para establecer la población a ser consultada por intermedio de la oficina de graduados de las IES.
- Generar un algoritmo de combinaciones de nombres de graduados sin información de contacto para realizar la búsqueda.
- Automatizar la búsqueda de la población objetivo basado en la combinación de nombres.
- Evaluar las respuestas encontradas a partir de la automatización propuesta.

2 Estado del arte

Actualmente las redes sociales (RRSS), según el último estudio de We are Social⁵, cuentan con más de 3.8 billones de usuarios activos al 2020, y un crecimiento de 9.2% en relación al 2019 y de aproximadamente un 64% desde el 2015.

Facebook fue la plataforma más usada de acuerdo con los datos de usuarios activos mes a mes, las cuentas activas, las audiencias de publicidad y los visitantes con 2449 millones de usuarios. El rango de edad con mayor audiencia publicitaria está entre 24 y 36 años, siendo 19% hombres y 13% mujeres, y el segundo rango con mejor audiencia está entre 18 y 24 años, con 14% hombres y 10% mujeres; grupos que llegan a ser de mayor interés para las instituciones de educación superior (IES).

LinkedIn cuenta con alrededor de 600 millones de usuarios, cifra alcanzada a partir de su publicidad, y un crecimiento del 1.6%. La población de más alcance en esta red social está en el grupo de 25 a 34 años.

En Latinoamérica, el 67% de la población usa las RRSS, en Colombia alrededor de 35 millones de personas son usuarios activos de RRSS, cifra que tuvo un crecimiento del 11% para el 2020 siendo Facebook la segunda plataforma más usada con 95% de crecimiento y cerca de 32 millones de usuarios activos. En el caso de LinkedIn, esta es la octava red de mayor uso con un crecimiento del 37% y cerca de 7.8 millones de usuarios.

Desde la aparición de las RRSS en el ámbito académico, su uso se ha popularizado y extendido a diferentes áreas, tanto así que, estas se han impulsado como plataformas de desarrollo para los docentes y los estudiantes, logrando entender de forma inimaginable, la comunidad estudiantil a partir del contenido y análisis de los datos consignados en las RRSS⁶.

⁵ We are Social. Digital in 2020. [online]. [citado en diciembre de 2020].

⁶ ALSAIF, Abdulwahaad. Investigate The Impact of Social Media on Students. Disertación. Cardiff - Reino Unidos: Cardiff Metropolitan University, 2016.

Durante este largo camino de aprendizaje se han alcanzado grandes avances, siendo el *marketing* uno de los grandes beneficiados y benefactores de las IES, ya que en los últimos años ha tenido un auge y crecimiento exponencial. La explotación de los datos para formar prospectos y futuros aspirantes a ingresar en las universidades ha sido determinante para consolidar las metas referente al número de matriculados⁷.

El uso de las RRSS en las IES a nivel mundial no es desconocido, inclusive, se ha documentado el impacto que tiene el uso de estas en el desempeño académico de los estudiantes⁸, al igual que se ha estudiado “cuáles son las RRSS sociales más convenientes y qué herramientas de las mismas son más útiles para los estudiantes”⁹. Investigaciones que siendo convenientes y relativas al objeto de estudio de la presente investigación, sin duda, son una aproximación frente a los estudios que deben llevarse a cabo para vislumbrar el potencial de las RRSS para los estudiantes y cómo llegan a convertirse en una herramienta más, en su proceso de su formación académica y profesional.

Las búsquedas automatizadas son una herramienta con más de diez años de uso y con una amplia variedad de opciones para realizar diferentes tareas según las necesidades. Las principales búsquedas se pueden hacer usando buscadores tales como Google, Bing o Yahoo!; además, el hacer uso de sus API, permite tomar ventaja de las herramientas para la automatización.

Es importante tener en cuenta que los rastreadores de contenidos son aplicaciones que permiten identificar y alertar cuando algún sitio ha realizado cambios en sus contenidos¹⁰, mientras que otras herramientas como Google Alerts informan mediante correos electrónicos o RRSS cuando se generan noticias sobre un tema en particular.

⁷ COOPER, Jeremy. How international students use social media to choose a UK university. En: *World University Rankings* [online]. 19 de mayo de 2017; PATTERSON, M. 4 Steps for Using Social to Recruit College Students. En: *Sproutsocial*. 28 de mayo de 2015.

⁸ OSHARIVE, Peter. *social media and academic performance of the students in Universities in Lagos State*. 2015.

⁹ GÓMEZ, Marisol, ROSES, Sergio y FARIAS, Pedro. El uso académico de las redes sociales en universitarios. En: *Comunicar*, 22 de marzo de 2012. [online]. doi: 10.3916/C38-2011-03-04

¹⁰ PapelesdeInteligencia.com. Herramientas de búsqueda en Internet: ¿Cuál es la mejor solución para mí?: En: *Papeles de inteligencia* (s.f.). [online]. [citado en diciembre de 2020].

Otro tipo de herramienta es la *If this then that* (IFTTT) “Si esto, entonces eso”, la cual permite crear reglas que se ejecutan, de forma específica, cuando se realiza una acción y, por último, están las aplicaciones que permiten buscar la reputación de un término basado en la cantidad de menciones que se obtengan¹¹.

La actualización de datos de los graduados para las IES en el mundo entero, es una problemática común¹², en aras de la consecución de calidad¹³; lo cual se ha venido tratando de resolver de distintas formas, como por ejemplo es el caso del uso de encuestas, que si bien permiten tener un aproximado, por lo general, cubren muestras no satisfactorias. Además, al final de toda la puesta en marcha de una estrategia para recolectar la información, puede verse como un ejercicio de investigación válido ya que son muestras propias de la población objeto de estudio, pero finalmente para las IES, los resultados en el uso de las encuestas no brindan la veracidad en la información¹⁴.

La gestión de información con los graduados, permite a las IES hacer una autoevaluación de sus programas, detectando posibles deficiencias a partir de cómo los graduados perciben su nivel de conocimiento y la formación recibida en el ejercicio de su profesión¹⁵.

Colombia cuenta con gran experiencia en el ámbito de gestión de calidad académica en IES, pues a partir de la creación del Consejo Nacional de Acreditación (CNA) en 1992¹⁶, ha tomado fuerza la necesidad de que las IES nacionales realicen seguimiento de sus graduados mediante estrategias de acompañamiento; de igual forma otras IES han planteado la necesidad de

¹¹ BANO, Juan J. Herramientas para automatizar la búsqueda de la información. En: Formación y competencias digitales en pequeñas dosis [Blog online] 22 de abril de 2016.

¹² GARCÍA ANCIRA, Claudia, TREVINO CUBERO, Arnulfo y BANDA MUÑOZ, Fernando. Caracterización del seguimiento de egresados universitarios. En: Estudios del Desarrollo social. 2019. vol. 7, no. 1, p. 24.

¹³ GARZÓN CASTRILLÓN, Ana. Modelo para el seguimiento y acompañamiento a graduados (SAG), una visión holística de la gestión de la calidad de la educación superior. En: *Edicación*. vol. 27, no. 52, p. 201-218.

¹⁴ Ministerio Educación Nacional (MEN). Experiencias Significativas de los procesos de seguimiento de graduados. [online] 5 de noviembre de 2010.

¹⁵ SCHOMBURG, Harald. Manual para estudios de seguimiento de graduados universitarios. Alemania: Universidad de Kassel, 2004. 345 p.

¹⁶ Ministerio de Educación Nacional (MEN). Consejo Nacional de Acreditación (CNA). (s.f.) [online] [citado en diciembre de 2020]. Disponible <https://www.cna.gov.co/1741/article-186382.html>

agregarse con otras IES en otros países, como estrategia alterna de seguimiento, ejemplo de ello es la RED GRADUA2, la cual busca unificar criterios y experiencias de diferentes países acerca de cómo se hace el acompañamiento de sus graduados¹⁷. Basado en esta experiencia, el seguimiento de los graduados es arcaico y depende, estrictamente, de que los graduados gestionen su actualización de datos, frente a lo cual es real y obvio el inconveniente que llega a presentarse frente a este método, como es la no localización del graduando o que esta persona no desee actualizar su información. Las encuestas que proponen la mayoría de las unidades de graduados son la herramientas más cercanas y fidedignas para alcanzar la consecución de estos datos actualmente.

El Ministerio de Educación Nacional (MEN) propone el libro *Manual para Estudios de Seguimiento de Graduados Universitarios* de Harald Schomburg, como una guía para que las IES realicen el seguimiento de sus graduados. Este libro ofrece como base de seguimiento, el uso de encuestas tal como se realizó en la Universidad de Kassel, Alemania; donde estas se proponen al 1er, 3er y 5to año de graduación como método primario de obtención de datos y, las plataformas tecnológicas como método secundario. Es así como se pueden identificar distintas universidades nacionales, como la Universidad del Pacífico y la Universidad Antonio Nariño, entre otras, donde la estrategia de seguimiento mediante encuestas ha sido aplicada¹⁸; tabulando la información obtenida para luego procesarla de la forma más conveniente para cada IES. Otras universidades deciden usar plataformas tecnológicas como método alternativo o complementario a las encuestas, por ejemplo la Universidad Francisco José de Caldas¹⁹ ha propuesto proyectos para gestionar el seguimiento de sus graduados mediante una base de datos

¹⁷ Asociación Columbus. Manual de instrumentos y recomendaciones sobre el seguimiento de egresados. 2006. [online].

¹⁸ Universidad Antonio Nariño. Anexo. Política de egresados UAN y modelo de seguimiento. 2011 [online]; Universidad del Pacífico. Procedimiento para seguimiento a egresados. 29 de mayo de 2015. [online].

¹⁹ ACOSTA CASTRO, Tatiana Andrea y ORTIZ GÓMEZ, Libny Andrea. Modelo de base de datos de egresados para análisis y fortalecimiento de las líneas de investigación del proyecto curricular de tecnología en topografía en el periodo 2009-I al 2014-I. Trabajo de grado Topografía. Bogotá: Universidad Distrital Francisco José de Caldas. Facultad del Medio Ambiente y Recursos Naturales, 2015.

donde se recolecta información netamente investigativa que le permite perfilar los datos y emplearlos de acuerdo con sus necesidades .

Otra experiencia es la de la Universidad de Nariño²⁰, quien realizó un proyecto sobre Software Web, con el cual se buscaba centralizar la información recolectada por la institución; además, este proyecto se presentó como alternativa de solución en inteligencia de negocio, donde se estaba construyendo en la fase 1, una bodega de datos.

Para el 2016 la Universidad de Cundinamarca extensión Facatativa²¹, presentó un proyecto para el seguimiento de graduados empleando las RRSS y un proyecto de informática e inteligencia de negocios para el análisis de los datos obtenidos. Los resultados de este proyecto distan del objetivo esperado debido a que: 1) tuvo limitaciones para el acceso a los datos y 2) no fue posible consultar la red social LinkedIn, ya que el acceso a esta información se encuentra limitado por la seguridad de datos²². Por tanto, para el desarrollo de la investigación solamente se consideraron los datos de Facebook y se hizo uso de la técnica de Web Scrapping, la cual está expresamente prohibida en las políticas de uso de los datos de Facebook²³ y LinkedIn²⁴. Como resultado existieron procesos manuales que impidieron llevar a cabo por completo la automatización de la actualización de datos.

²⁰ DELIO GÓMEZ, Eduardo, ESTRADA, Luis Obeymar y ARTURO DELGADO, Paola Andrea. Seguimiento a egresados en la Universidad de Nariño. En: Revista Universidad de Nariño. 2015. [online] doi: [10.22267/rhec.161919.21](https://doi.org/10.22267/rhec.161919.21).

²¹ CAÑÓN GUZMÁN, Julián Alejandro y LINARES FERNÁNDEZ, Julieth Dayana. Plataforma web con integración de minería de datos y redes sociales para el seguimiento a graduados del programa de ingeniería de sistemas de la Universidad de Cundinamarca, extensión Facatativa. Trabajo de grado Ingeniería de Sistemas. Bogotá: Universidad de Cundinamarca, Facultad de Ingeniería, 2016.

²² CAÑÓN GUZMÁN, Julián Alejandro y LINARES FERNÁNDEZ, Julieth Dayana. Plataforma web con integración de minería de datos y redes sociales para el seguimiento a graduados del programa de ingeniería de sistemas de la Universidad de Cundinamarca, extensión Facatativa. Trabajo de grado Ingeniería de Sistemas. Bogotá: Universidad de Cundinamarca, Facultad de Ingeniería, 2016, P47

²³ Para el 2018, las políticas de uso de datos de Facebook mediante sus API, fueron drásticamente cambiadas debido a la polémica suscitada por el escándalo de Cambridge Analytics (BBC New. Cambridge Analytica: la multa récord que deberá pagar Facebook por la forma en que manejó los datos de millones de usuarios. Redacción Mundo [online] 24 de julio de 2019).

FACEBOOK. Automated Data Collection Terms. [online] 15 de abril de 2020.

²⁴ LINKEDIN. Prohibited Software and Extensions. [online] 17 de diciembre de 2017.

A partir del estudio de investigaciones realizadas previamente al del presente proyecto, se puede concluir que es pertinente llevarla a cabo este estudio orientado al proceso de actualización de datos y la búsqueda de información de contacto de graduados de IES, debido a que hasta el momento no se ha considerado aplicar el enfoque ni el método estudiado tal y como se presenta en este proyecto de investigación.

En función de lo planteado anteriormente, a continuación, se exponen los métodos y las diferentes fases establecidas en el proceso metodológico para la elaboración y desarrollo del proyecto.

3 Metodología

La metodología trabajada para este proyecto es propia de una investigación aplicada, donde el análisis de acceso de información desde las redes sociales que se tuvieron en cuenta como fuentes primarias (Facebook y LinkedIn), al igual que el buscador como fuente secundaria (Google, por ser el más utilizado), determinó el desarrollo del proyecto.

Durante el planteamiento del proyecto se esperaba obtener la información directamente de las RRSS específicamente de LinkedIn y como alternativa Facebook, el proceso metodológico fue aplicado iniciando por la selección de la documentación con el objetivo planteado, en la revisión de alcance se revisaron cuales eran las opciones y requerimientos de estas RRSS para poder obtener la información directamente, viendo como única posibilidad la creación de una aplicación la cual debía pasar un varios procesos y validaciones por parte LinkedIn o Facebook.

En la implementación una aplicación a modo de prueba fue creada, en la cual consistía en que un usuario se autentificaba en ella y autorizaba a IES al uso de la información contenida por la RRSS, el permiso otorgado por defecto solo era el correo electrónico y los nombres del usuario, estos permisos de accesos de información era menos de lo esperado y necesitado por la oficina de graduados, entonces se levantó la información requerida para poder enviar al equipo de LinkedIn la aplicación solicitando el acceso a toda la información. El proceso tomó más de tres meses en autorización, la cual al final no fue satisfactoria debido a que la aplicación no justificaba que iba a realizar la IES con aquella información, tampoco como los usuarios estarían beneficiados al permitir compartir su información.

En ese punto el objetivo de proyecto debió modificarse para incluir algún método, en ese momento por definir, que permitiría la extracción de los datos, siendo así, se debió volver al principio la investigación ajustando cada uno de los pasos de la metodología de desarrollo del proyecto al nuevo objetivo.

Es de aclarar que el bajo nivel de acceso a la información y la consecución de permisos fueron dos limitantes importantes que redujeron considerablemente la posibilidad de obtener la información básica de los graduados. Fue por ello que la muestra seleccionada para la investigación se redujo a un total de 22 graduados de la Escuela Colombiana de Ingeniería Julio Garavito.

En total el proceso investigativo se desarrolló en cuatro fases, las cuales entran a ser descritas a continuación.

3.1 Fase 1. Selección de documentación base

Esta primera fase estuvo orientada a la consolidación de material investigativo relacionado con el tema de estudio del presente proyecto, lo cual permitió establecer aspectos destacados de los mismos a tener en cuenta para este proyecto y finalmente el desarrollo del estado del arte, con el cual se permitió contextualizar el problema de estudio y la orientación que se daría al mismo. Para esto os pasos realizados fueron:

1. Selección de criterios de búsqueda sobre la problemática.
2. Búsqueda de información en bases de datos especializadas.
3. Selección de documentos destacados para revisión.
4. Revisión de documentos.
5. Elaboración del estado del arte.

3.2 Fase 2. Revisión del alcance

La segunda fase de este proyecto investigativo estuvo orientada a indagar por las fuentes primarias y secundarias, siendo en este caso las redes sociales y los motores de búsqueda

respectivamente, que fuesen mejores alternativas para obtener la información que sería materia prima para el desarrollo del proyecto investigativo.

1. Investigar herramientas para acceso a información publicada en RRSS.
2. Investigar qué información es capturada por los motores de búsqueda.

Luego de varias pruebas y revisando la efectividad que brindaba las fuentes consultadas, se determinó como buscador Google, y para el caso de obtención de información se trabajó con las redes sociales de Facebook y LinkedIn.

3.3 Fase 3. Implementación

Luego de tener claridad sobre el objetivo investigativo y las fuentes que se emplearían para la obtención de la información, se pasó a la tercera fase donde se buscó aplicar e implementar el proyecto de trabajo, donde específicamente se realizaba la aplicación de conceptos y algoritmos en programación de sistemas para optimizar los resultados de la búsqueda de información.

Es así como en este proceso se realizaron varias pruebas para probar el algoritmo establecido y determinar su margen de error y éxito, al igual que evaluar sus resultados, para lo cual las actividades contempladas fueron las siguientes:

1. Preparación de la herramienta para el acceso a RRSS.
2. Elaborar un algoritmo que permita hacer combinaciones de nombres.
3. Consultar con el motor de búsqueda de la información de graduados.
4. Analizar los resultados para determinar si los datos obtenidos tienen la calidad esperada y son viables para desarrollo del proyecto.
5. Elaborar un algoritmo que evalúe cada resultado y califique este según el nivel de coincidencia del nombre.

3.4 Fase 4. Método y documentos

Como última fase de este proyecto se aplicó el algoritmo programado para la búsqueda de información de la muestra de graduados seleccionada haciendo uso de las redes sociales como fuentes primarias y del buscador.

Una vez consolidada la información obtenida a partir del método aplicado, se hizo evaluación de los resultados, permitiendo con ello identificar el margen de éxito y error que se tuvo en el rastreo de datos para el grupo de graduados que formaron la muestra de estudio, y se determinaron las ventajas y desventajas en su desarrollo, lo cual permitió dar respuesta a los objetivos planteados para el proyecto y, por ende, consolidar la experiencia investigativa en el documento final, siendo los pasos considerados para ello los siguientes:

1. Elaboración del método de obtención de información.
2. Formalización del documento de grado.
3. Elaboración del artículo de investigación.

4 Resultados y contribución

En el presente capítulo se expone un análisis del estado actual de las IES y sus oficinas de graduados y cómo en estas se gestiona la actualización de datos de sus graduados. De igual forma, a partir de reuniones realizadas con los profesionales que atienden una oficina de graduados de una IES y la experiencia de haber trabajado en un IES, se detectaron amenazas y debilidades, lo que permitió establecer y desarrollar un método de búsqueda de información de graduados que no cuentan con información de contacto.

4.1 Estado actual

Actualmente las unidades de graduados de las IES cuentan con dos métodos para la actualización de datos de los graduados:

1. Formulario de actualización de datos: la actualización que hacen los graduados es por voluntad propia. Si bien la unidad de graduados genera campañas para esta actualización, la efectividad de las mismas es baja, por lo cual el ingreso de los graduados a este formulario para la actualización de información es un proceso que no cumple con las expectativas previstas.
2. Encuestas: acorde con los lineamientos del Ministerio de Educación Nacional se realizan jornadas para la aplicación de encuestas a grupos específicos de graduados; el problema de este proceso es que los datos actualizados pierden validez con el transcurrir del tiempo.

4.1.1 Análisis de amenazas y debilidades

4.1.1.1 Amenazas

- Existe la posibilidad que otra organización realice los procesos de actualización de datos de forma eficiente, información que posteriormente sea vendida a las IES, sin importar si los datos son suficientes o contienen la información requerida.
- Las empresas de las RRSS pueden bloquear por completo a una organización en particular, el acceso a la información, de forma general o específica.
- Los graduados pueden tomar la alternativa de no permitir el acceso a su información o consentir el uso de esta.
- El MEN puede cambiar la información requerida a las IES, de tal forma que solicite más información o la ya requerida, cambie de forma o contenido.

4.1.1.2 Debilidades

- Actualización de datos de la población de graduados ineficiente o insuficiente. Puede llegar a ser entre el 25% y 50% de la población y algunas veces inferior al 25%²⁵.
- Las IES no retroalimentan las hojas de vida de los estudiantes próximos a graduarse con los logros académicos
- Las IES no ofrecen oportunidades de vinculación de las redes sociales de los estudiantes con la formación académica.
- Las IES no comparten a los estudiantes la importancia que tiene los graduados en los estudiantes y como pueden mejorar los programas académicos, siendo así, que solo una vez graduados se les tienen en cuenta para este proceso.
- El factor tiempo es causal de que la información actualizada de los graduados sea gradualmente obsoleta y necesite un reproceso.

²⁵ SCHOMBURG, Op. cit.

4.2 Método de búsqueda de información

Para iniciar con el método se debe hacer una selección de la población objetivo consultando con la oficina de graduados o la dependencia encargada en la IES, de tal forma que se identifiquen los graduados de quienes no se tiene información de contacto y se necesita la actualización de los datos, bien sea porque son graduados aptos para la presentación de información en los momento cero (M0), momento uno (M1) o momento cinco (M5), o bien porque estos datos son de interés estadístico para la IES, pues le permite conocer el estado de sus graduados.

En primer lugar, es importante definir en qué formato se recibe el listado de nombres y apellidos de los graduados, los campos de información que serán suministrados en la lista y cuál es la información mínima que se espera recibir de cada graduando para que pueda obtenerse búsquedas efectivas a partir de la información disponible.

Una vez identificada la población objetivo y los anteriores aspectos, se debe materializar el listado digital de los datos en un formato definido, dónde:

- 1) Quede establecida la posición de los nombre y apellidos.
- 2) Se defina cómo serán identificados los graduados con tres o más nombres o con apellidos compuestos.
- 3) Se debe acordar la existencia de otros campos comunes que puedan llegar a tener en las IES y que faciliten la búsqueda de información de la población determinada.

En paralelo a la última actividad descrita con relación al listado digital, y de acuerdo con la disposición de recursos para el alcance de los objetivos, se puede seleccionar el buscador adecuado que permita hacer consultas personalizadas, filtradas y detalladas de cada uno de los integrantes de la población objetivo; además, el buscador debe contar con la suficiente capacidad de poder integrar toda la información requerida, pues depende la efectividad de búsqueda a partir de la cantidad de información que este pueda rastrear y encontrar en las diferentes páginas.

El buscador debe ser configurado para que solo retorne resultados de las páginas de las RRSS deseadas u otros sitios definidos previamente; se debe tener en cuenta que la información que retorne el buscador tenga algún patrón de resultados, ya que esto permitirá realizar el análisis de la información encontrada. A su vez, el buscador debe contar con la seguridad suficiente para que la información que se busque a través de este, emplee los métodos indicados sin llegar a quebrantar las políticas de uso de la información de las RRSS.

Algunas recomendaciones que se brinda con respecto a la configuración son:

- Un aspecto importante es definir palabras clave en el motor de búsqueda, de tal forma que los resultados comiencen a ser relacionados directamente con la IES, por lo cual es recomendable usar el nombre completo de la institución, el nombre común por el cual es más reconocido esta institución en el ámbito académico, partes de su nombre y también sus siglas.
- Deshabilitar la búsqueda de imágenes.
- Que la región a buscar sea global o se cubra el mayor número de regiones posibles.
- El idioma principal, en lo posible, debe ser español. Si es factible adicionar un idioma secundario, que sea el inglés.
- Es fundamental que la búsqueda no sea sobre toda la web, por lo que se requiere configurar los sitios sobre los cuáles se desea enfocar la búsqueda; de esta forma se puede concentrar la búsqueda de los egresados en las RRSS dónde se considera que se llegue a encontrar la información de contacto requerida.

Una vez se tenga definida la población objetivo, se deberá iterar sobre esta población e ir, graduado por graduado, para correr el proceso adecuada y controladamente. Se debe asegurar que el formato del nombre contenga al menos tres partes: 1) dos nombres, 2) un apellido o 3) un nombre y dos apellidos); y así, poder realizar la búsqueda de una forma eficiente y con la menor cantidad de homónimos posibles.

Se debe marcar el nombre inicial como el “nombre base” ya que será necesario tanto para los procesos siguientes, como para un mejor control durante estos. Si se realizan varias búsquedas de la misma población objetivo, es necesario identificar si el nombre a buscar ya había sido procesado previamente, de esta forma no se desperdiciarán recursos en reprocesos.

El intentar encontrar a un graduado en las RRSS por sus nombres completos, como aparece escritos en el documento de identidad sería un desperdicio de recursos, ya que, por lo general, las personas acortan los nombres o incluso usan seudónimos. Debido a estas limitantes de información y no contar con los nombres alternativos o seudónimos, solo es posible realizar las combinaciones necesarias de nombres para encontrar a los graduados que usan sus nombres completos o combinaciones de los mismos en las RRSS. La combinación de nombres está basada en el nombre inicial o completo y, a partir de allí, se crean copias del nombre omitiendo partes de este (basado en la cantidad de partes que tenga el nombre) de la siguiente manera:

- Tres partes: no se genera combinación de Nombres, ya que es lo suficientemente corto para realizar una búsqueda satisfactoria del graduado.
- Cuatro partes: dos Nombres y dos Apellidos. Existe las siguientes posibilidades:
 - Nombre 1 Nombre 2 Apellido 1
 - Nombre 1 Nombre 2 Apellido 2
 - Nombre 1 Apellido 1 Apellido 2
 - Nombre 2 Apellido 1 Apellido 2
- Cinco partes: tiene un nombre, 3er nombre o un apellido compuesto.

Para el caso del 3er nombre, se manejan combinaciones de cuatro partes según las siguientes posibilidades:

- Nombre 1 Nombre 3 Apellido 1
- Nombre 1 Nombre 3 Apellido 2
- Nombre 2 Nombre 3 Apellido 1
- Nombre 2 Nombre 3 Apellido 2

Para el caso de apellido compuesto, se requiere unir el 3er apellido al 2do apellido y crear las combinaciones de cuatro partes y las siguientes posibilidades:

- Nombre 1 Nombre 2 Apellido 1 Apellido 2
- Nombre 1 Nombre 2 Apellido 2 Apellido 3
- Nombre 1 Nombre 2 Apellido 1 Apellido 3
- Nombre 1 Apellido 1 Apellido 2 Apellido 3
- Nombre 2 Apellido 1 Apellido 2 Apellido 3
- Seis partes: es la combinación de 5i y 5ii. (Tabla 4-1)
- Siete o más partes: para este proyecto no se contemplaron las combinaciones de estos nombres.

La tabla 4-1 ejemplifica como se podrían realizar las combinaciones de nombres, de acuerdo con la cantidad de partes que tenga un nombre.

Tabla 4-1 Ejemplos de combinaciones de nombres según la cantidad de partes del nombre

Partes	Nombre 1	Nombre 2	Nombre 3	Apellido 1	Apellido 2	Apellido 3
3-4	x	x		x		
3-4- 5i.5ii	x	x			x	
4-5i.5ii	x			x	x	
4-5i- 5ii		x		x	x	
5i	x		x	x		
5i	x		x		x	
5i		x	x	x		
5i		x	x		x	
5ii	x	x	x	x		

5ii	x	x			x	x
5ii	x	x		x		x
5ii	x			x	x	x
5ii		x		x	x	x

Fuente: elaboración propia a partir del proceso investigativo.

Al realizar dicha combinación de nombres debe ejecutarse la búsqueda, para ello debe pasársele al buscador cada uno de los nombres generados. Es necesario tener claro que dicho buscador ya debe estar configurado y parametrizado para que realice la búsqueda sobre las RRSS definidas y no sobre toda la web, igualmente debe realizar un filtro con un conjunto de palabras clave que faciliten y especifiquen la información a obtener.

En cuanto a los resultados obtenidos, se debe realizar una evaluación primaria para determinar si la información obtenida cuenta con la relación necesaria respecto al graduado que se está buscando, cabe recordar que al buscar en RRSS algunas respuestas corresponden a contactos del graduado en la RRSS y no al graduado en sí, de tal forma que se espera contar con un método que califique dicho evento y evite almacenar los datos de contacto del graduado. Igualmente, debido a que algunos nombres son cortos, debe mitigarse la posibilidad de los homónimos, para ello se plantea el uso de una rúbrica que califique cada campo retornado de la búsqueda. La siguiente es una calificación inicial²⁶:

- Si contiene el primer nombre, recibe 30%.
- Si contiene el segundo nombre, recibe 15%.
- Si contiene el primer apellido, recibe 30%.
- Si contiene el segundo apellido, recibe 25%.

Primer total 100%

²⁶ Esta calificación se puede modificar para agregar o filtrar más los resultados.

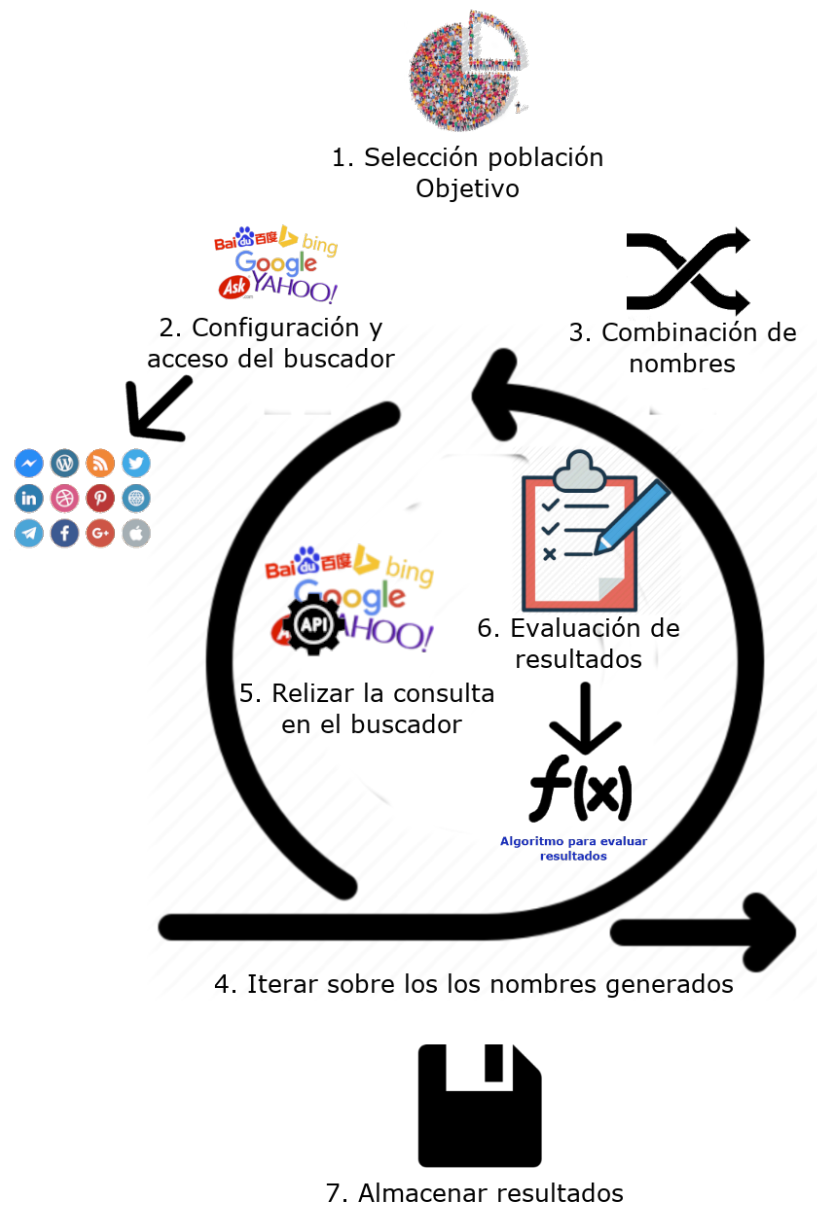
Esta calificación inicial se debe evaluar para cada una de los campos retornadas en la búsqueda que se pueden identificar como categorías ya que a lo largo del ejercicio son constantes e identifican para cada red social un grupo de datos específicos, teniendo en cuenta que podrán retornarse categorías como títulos, descripciones, etc.; lo cual se da según el buscador o la configuración de la búsqueda programable. Cada “categoría” tendrá solo un 20% (asumiendo que tiene que evaluar cinco “categorías”) de ponderación.

El algoritmo y método que permite la calificación de resultados debe catalogarse como una herramienta que puede ser modificada de forma y fondo, según las necesidades y alcance del proyecto. Pero debe quedar claro que es necesaria la evaluación de la información obtenida de cada uno de los graduados y sus combinaciones de nombres con el algoritmo propuesto u otro de objetivo similar, para que así se pueda cuantificar cada resultado y cada uno de estos, sea válido y candidato para ser almacenado.

El candidato resultante deberá ser almacenado, usando la metodología y las herramientas necesarias para tal fin.

Finalmente, para la organización de la información se debe determinar la forma más apropiada de almacenarla. Para este proceso, existe un amplio espectro de posibilidades como son las bases de datos locales o en la nube, las bases de datos relaciones o las bases de datos no relaciones. El uso de estas posibilidades queda a consideración de la IES según sus necesidades, expectativas y proyectos a futuro con la información obtenida en la búsqueda.

La figura 4-1 da cuenta del método de búsqueda descrito previamente.



Convenciones:
 -Pasos del método
 -Anotaciones

Figura 4-1 Método de búsqueda de información

4.3 Aplicación del método

Los pasos que se contemplaron para aplicar el método establecido en el presente proyecto se describen a continuación.

4.3.1 Selección de la población objetivo

Para este ejercicio se tomó una pequeña población objetivo de 22 graduados con el fin de mostrar cómo se aplicaba el método.

La figura 4-2 es ejemplo de un archivo separado por comas (CSV), con los nombres de algunos graduados de la Escuela Colombiana de Ingeniería Julio Garavito.

graduados_csv

Fabio Enrique	Quintero DiazGranados
Felipe	Vargas Beltran
Sandra Viviana	Pachón Bermúdez
Esteban Alejandro	Mejia Ramos
Johanna	Sandoval Angarita
Carolina	Sandoval Angarita
Maria Carolina	Botero Galvis
Paula Andrea	Castellanos Medina
Angie Paola	Cordoba Rodas
Maria Alejandra	Torres Avellaneda
Maria Jose	Velandia Zambrano
Ana María	Cortés Murcia

Figura 4-2 Ejemplo archivo CSV de la población objetivo

4.3.2 Configuración y acceso al buscador

Se seleccionó a Google como el buscador debido a que es el más utilizado²⁷.

4.3.3 Iterar sobre la población objetivo

La figura 4-3 es un ejemplo del código utilizado para la iteración sobre la población objetivo

```
foreach (var nameDTO in nameCombinations)
{
    Console.WriteLine(nameDTO.ToString());
    var list = searcherRepository.SearchPerson(nameDTO);
    //5. get the X number of results
    results.UnionWith(list);
}
```

Figura 4-3 Ejemplo código iteración sobre la población objetivo

4.3.4 Combinaciones de nombres

Por cada nombre de graduado, se hizo la respectiva combinación de nombres (figuras 4-4 y 4-5).



nameCombinations	Count = 5
[0]	{Fabio Enrique Quintero DiazGranados}
[1]	{Fabio Enrique Quintero}
[2]	{Fabio Quintero DiazGranados}
[3]	{Enrique Quintero DiazGranados}
[4]	{Fabio Enrique DiazGranados}

Figura 4-4 Ejemplo combinación nombre cuatro partes

²⁷ COPPOLA, María. Los 17 buscadores más utilizados en 2020. [online]. 16 de marzo de 2020 [actualizado el 6 de junio de 2020].

▼ nameCombinations	Count = 1
▶ [0]	{Felipe Vargas Beltran}

Figura 4-5 Ejemplo combinación nombre tres partes

4.3.5 Desarrollo de la consulta en el buscador

El buscador para ejecutar la consulta debe estar previamente configurado, con las palabras clave y demás configuraciones que a continuación se describen, también se deben generar las llaves necesarias para usar el recurso del API y por último la ejecución de la consulta en el buscador.

4.3.5.1 Definición de la búsqueda programable en el buscador

En el siguiente servicio de “Google cloud”

<https://programmablesearchengine.google.com/cse/all>

Se creó una búsqueda programable llamada “Social Media” con las siguientes características:

- Palabras clave (figura 4-6).
- Configuración de diferentes aspectos (idioma, región a cubrir, criterios de seguridad, restricción de imágenes, etc.) (figura 4-7).
- Configuración de las RRSS en las que el buscador realizaría la búsqueda (figura 4-8).

Palabras clave del motor de búsqueda ?

"ECI" "Escuela Colombiana de Ingenieria" "Escuela Colombiana" "Julio Garavito"

Figura 4-6 Búsqueda programable palabras clave

Búsqueda de imágenes ? **NO**

Búsqueda Segura ? **NO**

Región ? ⌵

Idioma ?

⌵

Figura 4-7 Búsqueda programable configuraciones

Sitios web en los que buscar

1 - 3 de 3

<input type="checkbox"/> Sitio	Etiqueta	Disponible en la API Site Restricted JSON ?
<input type="checkbox"/> *.facebook.com		✓
<input type="checkbox"/> *.linkedin.com		✓
<input type="checkbox"/> *.twiter.com		✓

[Opciones avanzadas](#)

Enviar solicitudes de indexación y eliminación a través de [Google Search Console](#). [Más información](#)

Buscar en toda la Web **NO**

Añade resultados generales de la Búsqueda web para aumentar los tuyos.
 Los resultados generales de la Búsqueda web no están disponibles en la API Site Restricted JSON.

Figura 4-8 Búsqueda programable sitios web en los que buscar

4.3.5.2 Obtención de los permisos y recursos

Se deben establecer las configuraciones necesarias para poder acceder a los recursos propuestos con los permisos requeridos y garantizados.

<https://developers.google.com/custom-search/v1/overview>

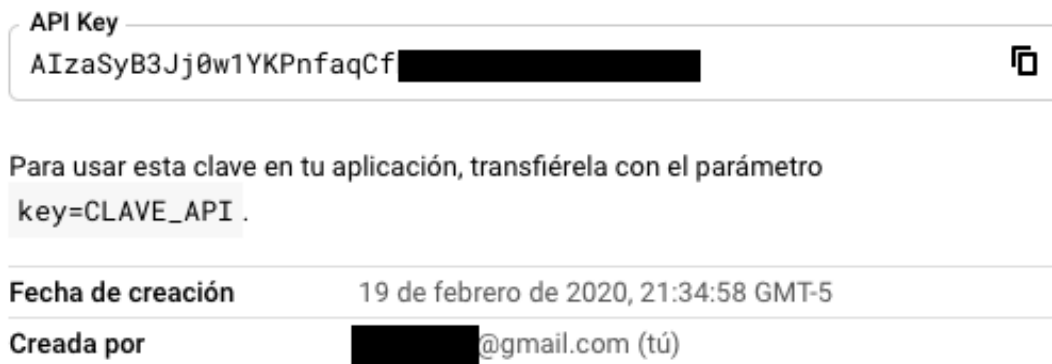
API key

Custom Search JSON API requires the use of an API key.

Get a Key

Figura 4-9 Obtenga los permisos y recursos

<https://console.developers.google.com/>



API Key

AIzaSyB3Jj0w1YKPnfaqCf [REDACTED]

Para usar esta clave en tu aplicación, transfírela con el parámetro `key=CLAVE_API`.

Fecha de creación	19 de febrero de 2020, 21:34:58 GMT-5
Creada por	[REDACTED]@gmail.com (tú)

Figura 4-10 Ejemplo API Key

El id del buscador definido en la búsqueda programable, será empleado en la consulta (figura 4-11).



Edición
Estándar

Obtener código

ID de buscador
0017336180614 [REDACTED]

Copiar al portapapeles

Figura 4-11 Ejemplo del Id del buscador

Usando la llave generada en la consulta:

```
https://www.googleapis.com/customsearch/v1?start={start}&&key={llaveGenerada}&{idD  
elBuscador}&q=%22{nombre_del_graduado}%22
```

Se obtiene un resultado de la búsqueda, como el ejemplo de la figura 4-12.

```
▼ items:  
  ▼ 0:  
    kind: "customsearch#result"  
    ▼ title: "Maria Jose Velandia Zambrano - EAE Business School - Colombia ..."  
    ▼ htmlTitle: "<b>Maria Jose Velandia Zambrano</b> - EAE Business School - Colombia ..."  
    ▼ link: "https://www.linkedin.com/in/maria-jose-velandia-zambrano-8019784b"  
    displayLink: "www.linkedin.com"  
    ▼ snippet: "Ve el perfil de Maria Jose Velandia Zambrano en LinkedIn, la mayor red \nprofesional  
    ▼ htmlSnippet: "Ve el perfil de <b>Maria Jose Velandia Zambrano</b> en LinkedIn, la mayor red <br>\n  
    ▼ formattedUrl: "https://www.linkedin.com/in/maria-jose-velandia-zambrano-8019784b"  
    ▼ htmlFormattedUrl: "https://www.linkedin.com/in/<b>maria-jose-velandia-zambrano</b>-8019784b"  
    ▼ pagemap:  
      ▼ cse_thumbnail:  
        ▼ 0:  
          ▼ src: "https://encrypted-tbn3.gstatic.com/images?q=tbn:ANd9GcQhAwrjyLExLN0bnvfcd-T8tRVWZLn-  
            width: "259"  
            height: "194"  
        ▼ metatags:  
          ▼ 0:  
            ▼ og:image: "https://media-exp1.licdn.com/dms/image/C4E03AQFKXH7gsRSBfQ/profile-displayphoto-shri  
              og:type: "profile"  
              twitter:card: "summary"  
            ▼ twitter:title: "Maria Jose Velandia Zambrano - EAE Business School - Colombia | LinkedIn"  
            ▼ og:title: "Maria Jose Velandia Zambrano - EAE Business School - Colombia | LinkedIn"  
              pagekey: "public_profile_v3_mobile"  
              locale: "es_ES"  
            ▼ og:description: "Ve el perfil de Maria Jose Velandia Zambrano en LinkedIn, la mayor red profesional d  
            ▼ twitter:image: "https://media-exp1.licdn.com/dms/image/C4E03AQFKXH7gsRSBfQ/profile-displayphoto-shri  
              profile:last_name: "Velandia Zambrano"  
              twitter:site: "@LinkedIn"  
              viewport: "width=device-width, initial-scale=1.0"  
              litmsprofilename: "public-profile-frontend"  
            ▼ twitter:description: "Ve el perfil de Maria Jose Velandia Zambrano en LinkedIn, la mayor red profesional d  
              profile:first_name: "Maria Jose"  
            ▼ og:url: "https://co.linkedin.com/in/maria-jose-velandia-zambrano"  
          ▼ cse_image:  
            ▼ 0:  
              - src: "https://media-exp1.licdn.com/dms/image/C4E03AQFKXH7gsRSBfQ/profile-displayphoto-shri-8019784b"
```

Figura 4-12 Ejemplo del resultado de la búsqueda

4.3.6 Rúbrica

La rúbrica se usa como método de filtrado, para evaluar los resultados obtenido en la búsqueda (figura 4-13).

```
public static float Evaluate(FullNameDTO fullName, ItemInfoDTO item)
{
    float totalAmount = 0F;
    var properties = typeof(ItemInfoDTO).GetProperties();
    foreach (var property in properties)
    {
        float amount = 0F;
        if (property.Name.Contains("CSE_image") || property.Name.Contains("SocialMediaInfo"))
        {
            continue;
        }
        var itemPropertyValue = ((string)property.GetValue(item, null)).ToLower();
        if (itemPropertyValue.Contains(fullName.FirstName.ToLower()))
        {
            amount += 0.3F;
        }
        if (!string.IsNullOrEmpty(fullName.MiddleName) && itemPropertyValue.Contains(fullName.MiddleName.ToLower()))
        {
            amount += 0.15F;
        }
        if (itemPropertyValue.Contains(fullName.Surname.ToLower()))
        {
            amount += 0.3F;
        }
        if (!string.IsNullOrEmpty(fullName.SecondSurname) && itemPropertyValue.Contains(fullName.SecondSurname.ToLower()))
        {
            amount += 0.25F;
        }
        totalAmount += amount * 0.2F;
    }
    return totalAmount;
}
```

Figura 4-13 Código del método para evaluar los resultados

4.3.7 Evaluación de los resultados

Usando la rúbrica propuesta en la figura 4-13, se evalúa cada resultado obtenido de la búsqueda y se compara para determinar si el resultado tiene el valor de confianza o de calidad necesario para ser tenido en cuenta para almacenarlo (figura 4-14).

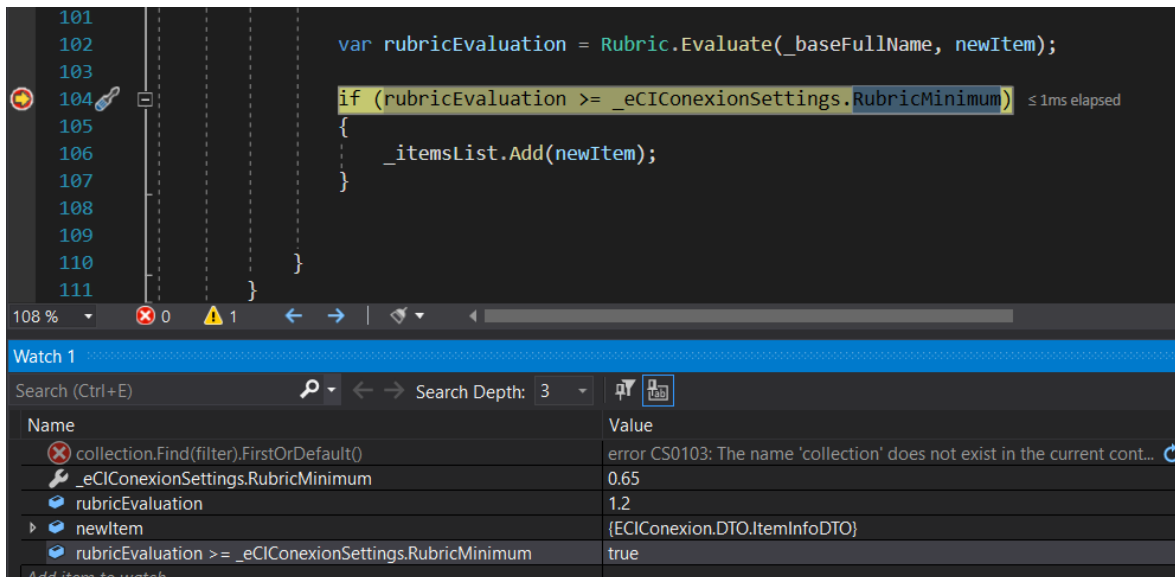


Figura 4-14 Ejemplo de la evaluación de los resultados

En el ejemplo de la figura 4-14, la evaluación dio una calificación de 1.2 y se espera que, al menos, tenga un puntaje de 0.65, como mínimo, para ser candidato a almacenar.

4.3.8 Almacenamiento de los resultados

Se usó una base de datos no relacional para almacenar los resultados, al obtener los resultados del buscador en formato JSON, se presentó una base de datos que permitiera almacenar bajo el mismo formato (figura 4-15).

```

_id: ObjectId("5fbae188818c0736a40d1fbd")
kind: "customsearch#result"
title: "Carolina Sandoval - Technical Director - Quality Engineering Studio ....."
htmlTitle: "Carolina Sandoval - Technical Director - Quality Engineering Studio ....."
link: "https://co.linkedin.com/in/carolina-sandoval-angarita"
displayLink: "co.linkedin.com"
snippet: "Carolina Sandoval. Technical Director en Globant. GlobantEscuela Colomb..."
htmlSnippet: "Carolina Sandoval. Technical Director en Globant. Globant<b>Escuela Co..."
formattedUrl: "https://co.linkedin.com/in/carolina-sandoval-angarita"
htmlFormattedUrl: "https://co.linkedin.com/in/<b>carolina-sandoval-angarita</b>"
> pagemap: Object

```

```

_id: ObjectId("5fbae188818c0736a40d1fbe")
kind: "customsearch#result"
title: "100+ Carolina Botero profiles | LinkedIn"
htmlTitle: "100+ Carolina Botero profiles | LinkedIn"
link: "https://ca.linkedin.com/pub/dir/Carolina/Botero+?trk=public_profile_sa..."
displayLink: "ca.linkedin.com"
snippet: "... at Banco de Bogotá,... Education, Universidad Camilo José Cela, Es..."
htmlSnippet: "... at Banco de Bogotá,... Education, Universidad Camilo José Cela, <b..."
formattedUrl: "https://ca.linkedin.com/pub/dir/Carolina/Botero+?trk=public_profile..."
htmlFormattedUrl: "https://ca.linkedin.com/pub/dir/Carolina/Botero+?trk=public_profile..."
> pagemap: Object

```

```

_id: ObjectId("5fbae188818c0736a40d1fbf")
kind: "customsearch#result"
title: "200+ \"Andrea Castellanos\" profiles | LinkedIn"
htmlTitle: "200+ \"Andrea Castellanos\" profiles | LinkedIn"
link: "https://au.linkedin.com/pub/dir/Andrea/Castellanos/co-0-Colombia"
displayLink: "au.linkedin.com"
snippet: "Viceministry of water, +15 more. Escuela Colombiana de Ingeniería 'Jul..."
htmlSnippet: "Viceministry of water, +15 more. <b>Escuela Colombiana</b> de Ingenier..."
formattedUrl: "https://au.linkedin.com/pub/dir/Andrea/Castellanos/co-0-Colombia"
htmlFormattedUrl: "https://au.linkedin.com/pub/dir/Andrea/Castellanos/co-0-Colombia"
> pagemap: Object

```

Figura 4-15 Ejemplo de almacenamiento de datos en una base de datos no relacional

Después de aplicar el método de búsqueda de información en RRSS a la población sin información de contacto seleccionada, aplicando el proceso dos veces variando el porcentaje de precisión de los datos, en cada una de las ejecuciones del proceso se presentó la siguiente información:

- La primera ronda fue configurada con el 65% de precisión.
Se generaron 197 resultados almacenados, alrededor de 8.9 resultados por graduado, con información de los graduados obtenidos de las RRSS (figura 4-16).

	TOTAL SIZE	AVG. SIZE
DOCUMENTS 197	414.7KB	2.1KB

Figura 4-16 Cantidad de resultados generados con 65% de precisión

Los nombres procesados fueron 74 y la ejecución se hizo en un tiempo de 211.75 segundos (figura 4-17).

```

Statistics
Total read names : 22
Total processed names : 74
Total processing time (Seg) : 211.750395
Press any key to finish

```

Figura 4-17 Otras estadísticas con 65% de precisión

- La segunda ronda fue configurada con el 85% de precisión.
Se generaron 103 resultados almacenados, alrededor de 4.7 resultados por graduado, con información de los graduados obtenidos de las RRSS.
Como era de esperarse, al aumentar el porcentaje de precisión de los datos, en la evaluación de resultados se realiza un filtro de la información con mayor intensidad (figura 4-18).

	TOTAL SIZE	AVG. SIZE
DOCUMENTS 103	233.6KB	2.3KB

Figura 4-18 Cantidad de resultados generados con 85% de precisión

Los nombres procesados en esta segunda ronda, igualmente fueron 74. Al respecto, no se esperaba una variación y la ejecución se hizo en un tiempo de 104.6 segundos (figura 4-19).

```
Statistics
Total read names : 22
Total processed names : 74
Total processing time (Seg) : 104.590058
Press any key to finish
```

Figura 4-19 Otras estadísticas con 85% de precisión

5 Conclusiones y recomendaciones

En esta oportunidad, el uso dado a las redes sociales (RRSS) y a los buscadores Puede considerarse como una oportunidad apropiada para lograr que las instituciones de educación superior (IES) puedan mitigar el problema que enfrentan con la pérdida de contacto con sus graduados.

Hacer uso de la herramienta descrita anteriormente, permitiría cerrar la brecha generada por los graduados desaparecidos y que no aportan o realimentan sus datos y los programas académicos de las instituciones.

Si bien durante el desarrollo del proyecto, parte de la investigación era usar las RRSS como fuente de información, la tala generada por las restricciones de acceso a la información fue un impedimento que no permitió hacer uso directo de los datos de las RRSS como fuente primaria, pero que finalmente permitiera a las instituciones de educación superior generar proyectos donde las aplicaciones autorizadas por las RRSS se conviertan en aliados de la comunidad estudiantil y apalanquen un vínculo con las instituciones, entonces una vez se generen nuevos grupos de graduados que permitan obtener el uso de los datos mediante el uso de una aplicación.

Queda a disposición de cada IES, determinar cuál debe ser la naturaleza, aplicación e información objetivo de dicha aplicación, para continuar pro de la misión institucional, no obstante, es importante tener en cuenta que conseguir los permisos de LinkedIn o Facebook es una tarea a largo plazo que debe incluirse a partir del diseño de una aplicación en la que se demuestre el beneficio que los usuarios pueden tener al vincularse a dicha aplicación. Además, debe discriminarse claramente cómo se usaría la información obtenida, cuáles serían las políticas de uso de dicha información, cómo se garantizará que la información obtenida pueda ser borrada una vez el usuario solicite su eliminación y, no menos importante, cómo la aplicación interactuaría con el perfil del usuario de la red social; por tanto, debe contemplarse cómo parte de la estrategia institucional, crear una aplicación que logre dar respuesta a los anteriores planteamientos, como a los objetivos necesarios.

Para ello se recomienda crear dicha aplicación pensada en los estudiantes de últimos semestres, quienes en corto tiempo se convertirán en graduados, con un alto porcentaje de poder obtener fácilmente los permisos para obtener la información necesaria para alimentar la base de datos.

Finalmente, como se mostró en la implementación del método se puede determinar que este es un mecanismo que logra, con gran esfuerzo, la obtención de información de contacto de una población objetivo de graduados, pero que a mediano plazo se le puede implementar opciones de mejora en la evaluación de resultados, como también está abierta a la posibilidad de que se le implemente mecanismos de inteligencia artificial que permitan alcanzar resultados más certeros.

Una recomendación a futuro para este tipo de proyectos es continuar mejorando el método de clasificación y filtrado de los resultados obtenidos, lo cual se podría lograr usando la información obtenida para identificar patrones que permitan entrenar modelos de inteligencia artificial y realizar de forma óptima este mismo proceso propuesto en esta investigación si se desea, pero, sobre todo, con más información valiosa y menos datos innecesarios.

Sumado a lo anterior es de tener presente que, la información recolectada y que es almacenada puede ser sujeta a otros mecanismos de análisis, como la minería de datos y obtener otros datos importantes para la oficina de graduados, para mercadeo o cualquier otra instancia que le sea productivo el uso de este tipo de información.

Referencias

- ACOSTA CASTRO, Tatiana Andrea y ORTIZ GÓMEZ, Libny Andrea. Modelo de base de datos de egresados para análisis y fortalecimiento de las líneas de investigación del proyecto curricular de tecnología en topografía en el periodo 2009-I al 2014-I. Trabajo de grado Topografía. Bogotá: Universidad Distrital Francisco José de Caldas. Facultad del Medio Ambiente y Recursos Naturales, 2015, 61 p.
- ALSAIF, Abdulwahaad. Investigate The Impact of Social Media on Students. Disertación. Cardiff - Reino Unido: Cardiff Metropolitan University, 2016, 57 p.
- Asociación Columbus. Manual de instrumentos y recomendaciones sobre el seguimiento de egresados. 2006. [online] [citado en diciembre de 2020]. Disponible en https://www.almalaurea.it/sites/almalaurea.it/files/docs/universita/altro/red_gradua_2.pdf
- BANO, Juan J. Herramientas para automatizar la búsqueda de la información. En: Formación y competencias digitales en pequeñas dosis [Blog online] 22 de abril de 2016. Disponible en <http://juanjbano.blogspot.com/2016/04/herramientas-para-automatizar-la.html>
- BBC New. Cambridge Analytica: la multa récord que deberá pagar Facebook por la forma en que manejó los datos de millones de usuarios. Redacción Mundo [online] 24 de julio de 2019. Disponible en <https://www.bbc.com/mundo/noticias-49093124>
- CAÑÓN GUZMÁN, Julián Alejandro y LINARES FERNÁNDEZ, Julieth Dayana. Plataforma web con integración de minería de datos y redes sociales para el seguimiento a graduados del programa de ingeniería de sistemas de la Universidad de Cundinamarca, extensión Facatativá. Trabajo de grado Ingeniería de Sistemas. Bogotá: Universidad de Cundinamarca, Facultad de Ingeniería, 2016. 51 p.
- COOPER, Jeremy. How international students use social media to choose a UK university. En *World University Rankings* [online]. 19 de mayo de 2017. Disponible en <https://www.timeshighereducation.com/student/blogs/how-international-students-use-social-media-choose-uk-university>
- COPPOLA, María. Los 17 buscadores más utilizados en 2020. [online]. 16 de marzo de 2020 [actualizado el 6 de junio de 2020]. Disponible en <https://blog.hubspot.es/marketing/buscadores-mas-utilizados>
- DELIO GÓMEZ, Eduardo, ESTRADA, Luis Obeymar y ARTURO DELGADO, Paola Andrea. Seguimiento a egresados en la Universidad de Nariño. En: Revista Universidad

de Nariño. 2015. [online] doi: [10.22267/rhec.161919.21](https://doi.org/10.22267/rhec.161919.21). Disponible en <https://revistas.udenar.edu.co/index.php/rhec/article/download/4136/5678?inline=1>

FACEBOOK. Automated Data Collection Terms. [online] 15 de abril de 2020. Disponible en https://www.facebook.com/apps/site_scraping_tos_terms.php

Fondo de Desarrollo de la Educación Superior. Acuerdo 01 de 2018 - CESU. Por el cual se actualizan los lineamientos para acreditación de alta calidad de programas de pregrado. Bogotá. 2018. Disponible en <https://fodeseop.gov.co/index.php/noticias/863-acuerdo-01-de-2018-cesu>

GARCÍA ANCIRA, Claudia, TREVINO CUBERO, Arnulfo y BANDA MUÑOZ, Fernando. Caracterización del seguimiento de egresados universitarios. En: Estudios del Desarrollo social. 2019. vol. 7, no. 1, p. 23-38.

GARZÓN CASTRILLÓN, Ana. Modelo para el seguimiento y acompañamiento a graduados (SAG), una visión holística de la gestión de la calidad de la educación superior. *En: Edicación*. vol. 27, no. 52, p. 201-218.

GÓMEZ, Marisol, ROSES, Sergio y FARIAS, Pedro. El uso académico de las redes sociales en universitarios. *En: Comunicar*, 22 de marzo de 2012. [online]. doi: 10.3916/C38-2011-03-04. Disponible en <https://www.revistacomunicar.com/pdf/preprint/38/14-PRE-13426.pdf>

LINKEDIN. Prohibited Software and Extensions. [online] 17 de diciembre de 2017. Disponible en Retrieved from <https://www.linkedin.com/help/linkedin/answer/56347/prohibited-software-and-extensions?lang=en>

Ministerio de Educación Nacional (MEN). Consejo Nacional de Acreditación (CNA). (s.f.) [online] [citado en diciembre de 2020]. Disponible <https://www.cna.gov.co/1741/article-186382.html>

Ministerio Educación Nacional (MEN). Experiencias Significativas de los procesos de seguimiento de graduados. [online] 5 de noviembre de 2010. Disponible en <https://es.slideshare.net/cardenasjhc/experiencias-significativas-29211713>

Ministerio de Educación Nacional (MEN). Decreto 1330 (25 de julio de 2019). Por el cual se sustituye el Capítulo 2 y se suprime el Capítulo 7 del Título 3 de la Parte 5 del Libro 2 del Decreto 1075 de 2015 – Único Reglamentario del Sector Educación. Bogotá. 2019. Disponible en <https://www.mineducacion.gov.co/1759/w3-article-387348.html?noredirect=1>

Abreviaciones

API	Interfase de programación de aplicaciones (application programming interfase)
CNA	Consejo Nacional de Acreditación
IES	Instituciones de Educación Superior
IFTTT	Si esto, entonces eso (If this then that)
MEN	Ministerio de Educación Nacional
OLE	Observatorio laboral para la educación
RRSS	Redes Sociales
SNA	Sistema Nacional de Acreditación en Colombia