

Extracción de habilidades y competencias profesionales de los perfiles universitarios estatales en Costa Rica, mediante técnicas de procesamiento de texto

Using word processing techniques for extracting professional skills and competencies from state university profiles in Costa Rica

Extração de habilidades e competências profissionais dos perfis universitários na Costa Rica, mediante técnicas de processamento de texto

Paola Chaves Bonilla
Observatorio Laboral de Profesiones
San José, Costa Rica
pchaves@conare.ac.cr

 ORCID: <https://orcid.org/0000-0001-7017-328X>

Carlos Gamboa-Venegas
Centro Nacional de Alta Tecnología
San José, Costa Rica
cgamboa@cenat.ac.cr

 ORCID: <https://orcid.org/0000-0001-9712-0575>

Katherine Sandí Araya
Observatorio Laboral de Profesiones
San José, Costa Rica
ksandi@conare.ac.cr

 ORCID: <https://orcid.org/0000-0001-8129-3826>

Karen Corrales Bolívar
Observatorio Laboral de Profesiones
San José, Costa Rica
kcorrales@conare.ac.cr

 ORCID: <https://orcid.org/0000-0002-2979-2190>

Javier Herrera Mora
Centro Nacional de Alta Tecnología
San José, Costa Rica
javier.herreramora1@gmail.com

 ORCID: <https://orcid.org/0000-0003-2906-502X>

Recibido – Received – Recebido: 15 / 09 / 2021 Corregido – Revised – Revisado: 29 / 10 / 2021 Aceptado – Accepted – Aprovado: 12 / 11 / 2021

DOI: <https://doi.org/10.22458/ie.v24i36.3818>

URL: <https://revistas.uned.ac.cr/index.php/innovaciones/article/view/3818>

Resumen: La cuarta revolución industrial trae consigo muchos cambios en la forma y la diversidad en las que se desarrollan las actividades del mundo en general. La situación se vuelve bastante compleja en el caso de las personas graduadas universitarias, dado que esta revolución presenta retos con los nuevos conocimientos, habilidades y capacidades para desarrollar dentro de los puestos de trabajo, lo cual fomenta la aparición de grandes desafíos que involucran los procesos de la automatización. Con el objetivo de perfilar cada una de las disciplinas respecto a diferentes grupos de habilidades, la presente investigación se estructura en tres etapas metodológicas. La primera, es mediante la descripción de los perfiles académicos de las universidades estatales

costarricenses, siendo este conjunto de documentos la población de estudio. La segunda, es sobre la construcción de un diccionario de palabras y la tercera, se trata de la construcción de la base de datos que alberga competencias y habilidades de las personas graduadas universitarias de cada una de las disciplinas. Todo lo anterior fue posible, mediante el desarrollo de una serie de herramientas programadas en el lenguaje Python, para la recolección de información, extracción de habilidades y competencias, y finalmente, el cálculo de las proporciones. El principal resultado de la investigación, es la construcción y naturalización de una base de datos a nivel de carreras universitarias estatales, mediante la creación de ocho grupos de habilidades y competencias debidamente integrados por competencias asociadas a diferentes temáticas. El aporte más importante, es el análisis integral de los perfiles académicos de las carreras universitarias.

Palabras clave: Enseñanza superior, estudiante universitario, habilidades, competencias, industria, profesión, procesamiento de datos.

Abstract: The fourth industrial revolution brings with it many changes in the way as well as a diversity in which the activities of the world at large take place. Therefore, the situation becomes quite complex in the case of university graduates, for this revolution presents challenges involving new knowledge, skills, and abilities needed for the jobs, which encourages the appearance of significant challenges that are comprehended in the processes of automation. In order to outline each of the disciplines concerning different groups of abilities, this research was structured in three methodological stages. The first is a description of the academic profiles of the Costa Rican state universities- this is a set of documents representing the population studied. The second is the creation of a dictionary, and the third is when we built a database that houses the skills and abilities of university graduates from each discipline. All of the above was possible, through the development of a series of tools programmed in Python language, to collect information, extract skills and competencies, and finally, calculate proportions. The main result of the research is the construction and naturalization of a database at the level of state university careers through the creation of eight groups of skills and competencies appropriately integrated by competencies associated with different themes. The most important contribution is the comprehensive analysis of the academic profiles of university careers.

Keywords: Higher education, college student, skills, competencies, industry, profession, data processing.

Resumo: A quarta revolução industrial trouxe muitas mudanças na maneira e na diversidade na qual as atividades do mundo são desenvolvidas em geral. A situação se torna ainda mais complexa no caso das pessoas já com a graduação universitária, pois esta revolução apresenta desafios e novos conhecimentos, habilidades e capacidades para desenvolver nos locais de trabalho, o que fomenta a aparição de grandes desafios que envolvem os processo de automatização. Com o objetivo de criar um perfil para cada uma das disciplinas no tocante aos diferentes grupos e habilidades. Esta pesquisa se estrutura em três etapas metodológicas: A primeira tem a ver com a descrição dos perfis acadêmicos das universidades estaduais da Costa Rica, sendo este o conjunto de documentos da população em estudo; a segunda, lida com a construção de um dicionário de palavras; e a terceira ocupa-se da construção da base de dados que aloja competências e habilidades das pessoas com graduação universitária para cada uma das disciplinas. O anterior foi possível mediante o desenvolvimento de uma série de ferramentas usando Python como língua de programação para a colheita de informações, a extração de habilidades e competências e finalmente para o cálculo das proporções. O principal resultado da pesquisa é a construção e a naturalização de uma base de dados em nível de carreira universitária mediante a criação de oito grupos de habilidades e competências debidamente integradas por competências associadas às diferentes temáticas. A contribuição mais importante é a análise integral dos perfis acadêmicos das carreiras universitárias.

Palavras-chave: ensino superior, estudante universitário, habilidades, competências, indústria, profissão, processamento de dados..

INTRODUCCIÓN

Actualmente, a nivel mundial, se está experimentando una serie de cambios que obligan al ser humano a adecuarse de acuerdo con las exigencias que plantea el entorno y de esta manera, reinventarse, para sobrevivir en un ambiente radical y de constantes cambios. Esto es consecuente con lo que plantea Charles Darwin (1859), acerca de la capacidad del ser humano para adaptarse a sus diferentes entornos.

Por otra parte, los entornos cambiantes han sido constantes a lo largo de la historia del ser humano. Un ejemplo de esto, es la primera revolución industrial, que se caracterizó por el uso de motores impulsados a vapor; seguida por la Industria 2.0, la cual florece a principios del siglo XX, donde surge la energía eléctrica, dando paso a la producción y consumo masivo, germinando con esto a la Industria 3.0 (período comprendido entre los años 1970-2000) con un mayor nivel de automatización al ser incorporada

la electrónica y tecnologías de información, en los sistemas de producción. Sin embargo, cuando se pensó que ya se había avanzado en muchos ámbitos y que el mundo había llegado a su máxima etapa evolutiva; en el año 2010 se presenta la digitalización, la cual da paso a una industria y productos inteligentes, caracterizados por la aparición de herramientas, como los sistemas ciber-físicos, el Internet de las cosas (IOT), la hiper-conectividad y el *big data*; dando paso al nacimiento de la Industria 4.0, la cual plasma como pilar fundamental, la automatización de procesos y actividades repetitivas, haciéndolos más eficientes y eficaces. (Perasso, 2016).

Por ello, una de las principales preocupaciones en el ámbito laboral, son las repercusiones que trae consigo esta nueva revolución. De ahí, que la investigación desarrollada por Frey & Osborne (2013), se centró en el análisis de los efectos de la cuarta revolución industrial en el empleo, mediante el estudio de las competencias adquiridas en los puestos de trabajo, cuyo objetivo fue determinar el porcentaje de riesgo de automatización, a través de las características de los puestos y funciones de las distintas ocupaciones. Los autores plantean la premisa de que se debe mejorar la educación, promover la reubicación de los puestos laborales y, reducir la desigualdad, permitiendo que la inteligencia artificial construya más opciones laborales que las que destruirá.

Aunado a esto, Corrales y Sandí (2019) mostraron un primer panorama nacional de las carreras universitarias hacia la industria 4.0, y los perfiles en riesgo de sustitución. El estudio reflejó la necesidad del país en cuanto al desarrollo de herramientas en ciencia, tecnología, ingeniería y matemáticas, que brinden información oportuna a la educación superior, respecto al rediseño de mallas curriculares, la fusión de carreras o en su defecto, la modificación o finalmente, el cierre de aquellas cuyo futuro estaría amenazado por la industria 4.0.

De acuerdo con lo anterior, la cuarta revolución representa un gran reto para la academia, pues la implementación de la industria 4.0 requiere de la contratación de nuevos perfiles universitarios, procedentes de carreras bajo la metodología STEM (ciencia, tecnología, ingeniería, matemáticas; por sus siglas en inglés), con habilidades que les permitan adecuarse con mayor facilidad a la demanda tecnológica. (Santos, 2013).

En concordancia con lo anterior, el proyecto Tunnig define competencias, como: “la capacidad que todo ser humano necesita para resolver, de manera eficaz y autónoma, las situaciones de la vida. Se fundamentan en un saber profundo, no solo saber qué y saber cómo, sino saber ser persona en un mundo complejo cambiante y competitivo” (Beneitone, y otros, 2007, 35).

Otros autores como Velázquez, y Santiesteban (2018), definen que las habilidades constituyen formaciones psicológicas estructuradas en acciones y operaciones; las competencias constituyen configuraciones psicológicas estructuradas en estados dinámicos, con un valor emocional estable. Las competencias integran habilidades, pero no se reducen a ellas. Asimismo, Berrocal, García y Ramírez-Vielma (2021), mencionan que las competencias conducen a la demostración de habilidades y destrezas, lo que la hace observable, y que también representa la capacidad de transferir habilidades y destrezas de una área a otra.

Por tanto, para esta investigación, tanto las habilidades y competencias son claves, ya que estas le permiten al individuo acceder a niveles de logro progresivo que contempla el aprendizaje continuo, la toma de decisiones y participar de temas culturales y sociales. (Cervantes, 2007). Ahora bien, en el entorno nacional, existe un vacío sobre información, referente a las competencias y habilidades involucradas en las carreras universitarias entorno a la industria 4.0, esto lo respalda la investigación Corrales y Sandí (2019), que plantea la necesidad de una base de datos unificada, con las habilidades y competencias de las carreras universitarias, lo cual responde a uno de los objetivos específicos de la investigación y a su vez, se plantea el objetivo general de categorizar los perfiles académicos de las carreras universitarias estatales, por medio de herramientas computacionales de extracción de texto, para la construcción y clasificación de habilidades y competencias.

MATERIALES Y MÉTODOS

Este apartado abordará tres secciones: la primera, es la descripción de los perfiles universitarios; la segunda, trata sobre la construcción del diccionario de palabras; y la tercera, la construcción de las habilidades y competencias.

Recolección de información: perfiles profesionales y encuesta de habilidades y competencias

Se utilizó como base, la clasificación de las carreras del último estudio de seguimiento de la condición laboral, realizado en el año 2019, por el Observatorio Laboral de Profesiones (OLaP), en cuanto a los niveles de desagregación de área, disciplina y carrera; de manera que dicha clasificación aseguró que la disciplina contaba con personas graduadas. Para efectos de la investigación, la búsqueda de información quedó delimitada en 109 disciplinas, agrupadas en 10 áreas del conocimiento (Artes y Letras, Ciencias Básicas, Computación, Ciencias Económicas, Derecho, Educación, Recursos Naturales, Ingeniería y Ciencias de la Salud).

Adicionalmente, se recolectaron los perfiles académicos y los planes de estudio, como documentos digitales en varios formatos, en los sitios *web* oficiales de cada institución, escuelas o unidades académicas de las cinco universidades estatales, oficinas de orientación vocacional, y a su vez, en los colegios profesionales, tales como: Colegio Federado de Ingenieros y Arquitectos, Colegio de Enfermeras de Costa Rica, Colegio de Geólogos de Costa Rica, Colegio de Médicos y Cirujanos de Costa Rica, Colegio de Microbiólogos y Químicos Clínicos de Costa Rica, Colegio de Optometristas de Costa Rica, Colegio de Profesionales en Ciencias Políticas y Relaciones Internacionales, Colegio de Trabajadores Sociales de Costa Rica y Colegio de Profesionales en Psicología de Costa Rica. Adicionalmente, se utilizaron los dictámenes de la Oficina de Planificación de la Educación Superior (OPES) de los últimos cinco años, esto por recomendación de la jefatura de la División Académica del Conare, quienes tienen a cargo este proceso, desde 1982, fecha en que se firma el Convenio de coordinación de la educación superior universitaria estatal en Costa Rica. (Consejo Nacional de Rectores, 1982). A pesar de que dichos dictámenes de OPES concentran la mayoría de la información, no cuentan con aquellas carreras que se crearon antes de la firma del convenio; asimismo, si una carrera actualiza el plan de estudios, y esta actualización es inferior a un 25% del plan, no se actualiza el dictamen. Cabe destacar que, a nivel de las cinco universidades estatales, no existe una oficina o ente encargado de centralizar la información en cuanto a los perfiles académicos; por lo que en ocasiones, se trabaja con documentos desactualizados y con estructuras de información heterogéneas.

Particularmente, algunas fuentes de información estaban disponibles, únicamente como texto incrustado en los sitios *web* de las universidades, por lo que se utilizó un proceso conocido como *web scraping*, para recolectar el texto crudo que se refiere a la descripción de perfiles para posteriormente, procesarlo con herramientas informáticas para convertirlos en archivos de texto plano (TXT), en lo que se profundizará más adelante.

Para efectos de subsanar faltantes de información, se construyó un cuestionario, compuesto por un total de 20 interrogantes, enfocado en recolección de habilidades y competencias, que se aplicó a directores o encargados de cátedra de cada una de las unidades académicas, así como a jefaturas inmediatas de las personas graduadas, quienes conocen directamente el quehacer de la profesión. Esta información se encuentra tabulada en un archivo delimitado por comas.

Pese a lo anterior, existe una limitante en la información que se encuentra en los perfiles universitarios, ya que no se describe concretamente, si los textos indicados contemplan habilidades o competencias, o incluso tareas de la carrera.

Diccionario de palabras y construcción de habilidades y competencias

Respecto a la construcción de la base de datos, que alberga las características y habilidades y competencias de las personas graduadas universitarias, se diseñó un listado de palabras que permitió relacionarlas con un grupo de habilidades específicas.

La agrupación de palabras y posterior construcción de habilidades y competencias, está basada en el listado definido en la base de datos O* Net Resource Center (O*NET 26.0 Database, 2020). Este listado, se divide en cuatro grupos de habilidades y seis grupos de competencias. Dicha agrupación está contextualizada en un escenario estadounidense, por lo que se procedió a naturalizar la traducción al contexto nacional.

Por ese motivo, se definen ocho grupos de habilidades y competencias, específicamente las sensoriales, cognitivas, físicas y psicomotoras corresponden a habilidades y, los restantes grupos a competencias integradas por problemas complejos, sociales, sistemas y técnicas, básicas y gestión de recursos. La figura 1 presenta las definiciones para cada uno de los grupos descritos anteriormente.



Fuente: Elaboración propia.

Para seleccionar las disciplinas de las cuales se extrajeron las primeras palabras, se utilizó un muestreo simple al azar de las 129 disciplinas del estudio del OLaP. Cabe destacar, que esta primera extracción se desarrolló de forma manual. Se identificaron las palabras de interés, inmersas en los apartados de habilidades y tareas de la muestra, que se clasificaron según las habilidades anteriormente mencionadas en la figura 1. Dichas palabras conformaron la versión inicial del diccionario de palabras, que fue insumo para las herramientas informáticas creadas para el proyecto.

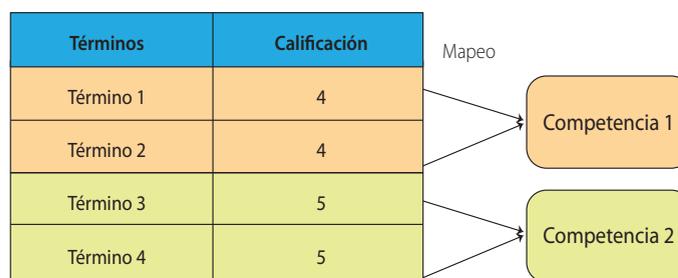
Una vez construido el diccionario con todas las palabras de interés, se realizaron dos actividades que se enumeran a continuación:

1. **La codificación**, la cual se realizó con el fin de identificar aquellos términos que, aunque tenían raíces y escrituras distintas, semánticamente significaban lo mismo. Así, todos los términos con el mismo sentido semántico tenían el mismo código, mientras que el resto de términos se agruparon con códigos únicos.

A este punto se tenían términos agrupados en cada uno de los conjuntos de habilidades y/o competencias. Sin embargo, se debía redactar una competencia que identificara a cabalidad cada uno de los términos contenidos en el grupo.

2. **El mapeo**, con el propósito de asociar cada elemento de la codificación con una competencia y/o habilidad, estas se asociaron al diccionario de palabras para que posteriormente, la herramienta *Decoder*, lograra mapearla de manera automática. La figura 2 ejemplifica el proceso de agrupación que se realizó con los términos del diccionario.

Figura 2. Codificación y mapeo de términos



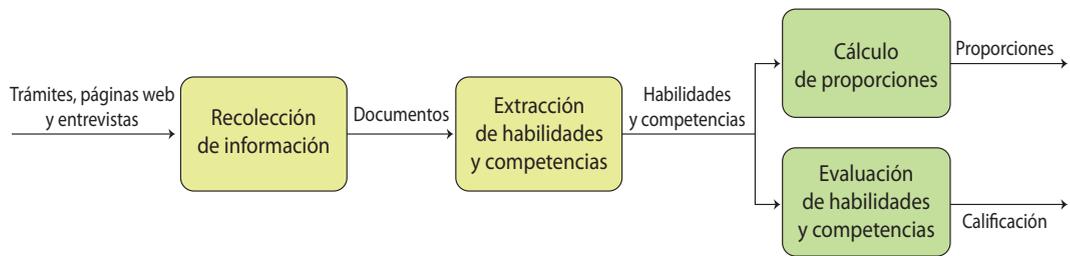
Fuente: Elaboración propia.

Implementación de la herramienta computacional

Todo el software desarrollado para el procesamiento de texto, fue programado en lenguaje *Python* versión v3.6.8 (Python, 2018). Para el etiquetado de archivos se utilizó la herramienta de comparación de texto *Jellyfish* (Turk & Stephens, 2020). En la extracción de texto de los archivos de formato pdf se utilizó *PyMuPDF* (Mackie & Liu, 2021). La manipulación de las estructuras de datos fue realizada a través de la biblioteca *Pandas* (McKinney, 2021). Concluyendo con los módulos utilizados, además de los incluidos por defecto en *python*, se tuvo la biblioteca para procesamiento de lenguaje natural *Nltk* (Bird, Steven, Edward Loper & Ewan Klein, 2021).

La figura 3 resume los tres procesos principales que realizó el algoritmo, los cuales fueron: la recolección de información, la extracción de habilidades y competencias; y finalmente, el cálculo de proporciones. A continuación, se detallarán cada una de ellas. Para muchas de estas etapas, se trabajó con la información de las carreras y no de las disciplinas como tal, debido a que los resultados por disciplina son obtenidos de agrupaciones en las etapas finales.

Figura 3. Proceso general



Fuente: Elaboración propia.

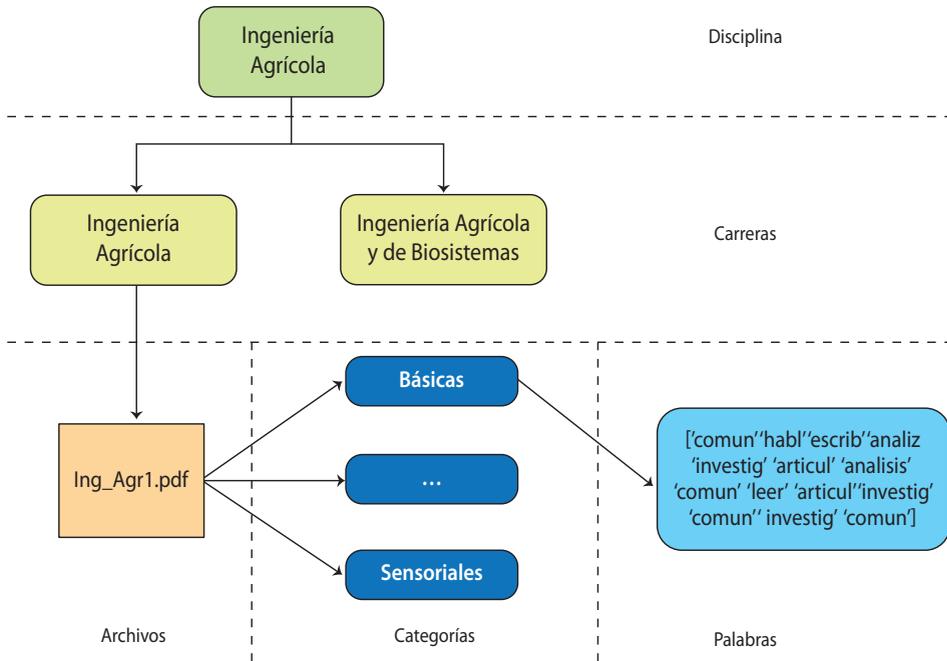
El primer proceso, considera la etapa de la recolección de información, donde se utilizaron técnicas como *web scraping* para extracción de texto de sitios web, análisis de imágenes para extracción de texto de archivos en formato *pdf* que se encontraban protegidos o escaneados y procesamiento de la base de datos de la encuesta de habilidades para industria 4.0.

El segundo proceso, fue la extracción de habilidades y competencias, llevada a cabo por las dos soluciones informáticas denominadas *Tiresias* y *Caribdis*, que se detallan a continuación:

- *Tiresias* es una herramienta informática, que permite asociar de forma semiautomática un archivo del acervo de información, con alguna de las carreras del catálogo establecido para el proyecto. El principio técnico en el que se basa *Tiresias* para obtener las carreras más probables, se llama Distancia de Levenshtein. Dicha distancia cuantifica el grado de diferencia entre dos hileras de caracteres, a partir de la cantidad de operaciones necesarias para transformar una hilera A en la hilera B. La carrera a la que se asocia un archivo del acervo, con gran probabilidad, se encuentra dentro de las 10 distancias más pequeñas. *Tiresias* permite al usuario seleccionar alguna o ninguna de las opciones que brinda la herramienta.
- *Caribdis* es la herramienta central del proyecto, dado que procesa el contenido de los archivos del acervo, extrae las habilidades y competencias, y calcula las proporciones por tipo de habilidad. Dicho algoritmo ejecuta las etapas de limpieza, búsqueda y agrupación.
 - La etapa de limpieza, es donde el texto debe ser estandarizado para aumentar las probabilidades de encontrar coincidencias; por tanto, se utilizaron diversas técnicas comunes para esto, empezando con remover los espacios extra y puntuaciones, reemplazar las mayúsculas por minúsculas, eliminar las palabras conocidas como *stopwords* para evitar diferencias por caracteres superfluos, la sustitución de todas las palabras por una versión más simplificada de las mismas, mediante el *stemming* para evitar diferencias ocasionadas por variaciones de una misma palabra y finalmente, reemplazar todos los acentos por las vocales sin acentuación. El orden de cada una de las transformaciones anteriores es relevante, ya que los resultados de una etapa dependen de la otra.
 - La segunda etapa que desarrolla *Caribdis* es la búsqueda, donde cuenta con una estructura de memoria, que contiene los datos de cada uno de los archivos del acervo por separado. Uno de los datos es el texto contenido en cada archivo, a estos textos se les aplica una serie de expresiones regulares que buscan y extraen las habilidades de interés. El conjunto de habilidades encontradas, se divide en ocho categorías que responden a las ocho habilidades establecidas anteriormente en la sección de recolección de información.
 - Finalmente, la tercera etapa corresponde a la agrupación. Esta etapa, tal como su nombre lo indica, agrupa la información respecto a diferentes condiciones o filtros, utilizando la siguiente desagregación: cada universidad tiene múltiples áreas de conocimiento, cada área tiene

múltiples disciplinas, cada disciplina tiene múltiples carreras y por cada carrera, se tiene uno o múltiples archivos, para cada archivo hay ocho categorías de búsqueda y por cada categoría hay cero, uno o múltiples hallazgos. Para ejemplificar la funcionalidad de *Caribdis*, la figura 4 describe una sección del árbol formado a partir del nivel de disciplina.

Figura 4. Ramificación parcial de la disciplina de Ingeniería Agrícola



Fuente: Elaboración propia.

Por último, también *Caribdis* obtiene la estimación de las proporciones. El algoritmo descrito anteriormente, es capaz de calcular todas las proporciones según el nivel de desagregación escogido por el usuario, mediante un parámetro proporcionado al programa, en el momento de realizar una ejecución. La proporción de cada grupo de habilidad o competencia por categoría dentro de un nivel de desagregación, está dada por la siguiente fórmula:

$$PNG = \frac{h_{CHP}}{h_G} \cdot 100$$

N: Nivel de desagregación

G: grupo de habilidad

h_{CHP} : Cantidad de habilidades del grupo *G* presentes en el nivel *N*

h_G : Suma de las habilidades de grupo *G*

La proporción P_{NG} se interpreta como el porcentaje de habilidades o competencias específicas del conjunto unión, sin duplicados del tipo *G* respecto a la suma total de las habilidades o competencias de tipo *G*; por lo tanto, señala la presencia que tienen las habilidades o competencias del conjunto unión respecto a cada grupo. Es decir, si en un nivel de desagregación se encuentran cuatro habilidades o competencias del grupo *G*, y el grupo *G* contiene un total de cien habilidades o competencias, entonces el nivel de desagregación reúne 4% del total de habilidades o competencias contenidas en el grupo *G*.

DISCUSIÓN DE RESULTADOS

En esta sección, se describen los resultados de las diferentes etapas planteadas en la metodología. Para este análisis, se consideran 109 disciplinas, las cuales se imparten en las universidades estatales; el logro principal fue la efectividad del algoritmo, al clasificar las palabras encontradas en los diferentes grupos para dar paso a la construcción de las habilidades y competencias, y finalmente, realizar el cálculo de las proporciones en los grupos.

El acervo construido cuenta con un total de 316 documentos en diferentes formatos, (*pdf, txt o html*). Dicha información se resume en el cuadro 1, en el cual, las fuentes de información principales corresponden a los sitios *web* de las escuelas y las vicerreorías de docencia de las cinco universidades estatales, reflejando de esta manera, la colaboración interinstitucional; asimismo, las oficinas de orientación vocacional de las universidades, por su naturaleza y razón de ser, cuentan con información completa de las carreras.

Finalmente, la encuesta de habilidades para industria 4.0 aplicada a las personas directoras y jefaturas inmediatas, cuenta con un total de 568 respuestas; sin embargo, este dato que se presenta en el cuadro 1 es preliminar, ya que aún se encuentra en la etapa de recolección de la información. A pesar de lo anterior, esta aplicación del cuestionario es el insumo más relevante respecto al aporte de información.

TABLA 1
Distribución absoluta de las fuentes de información

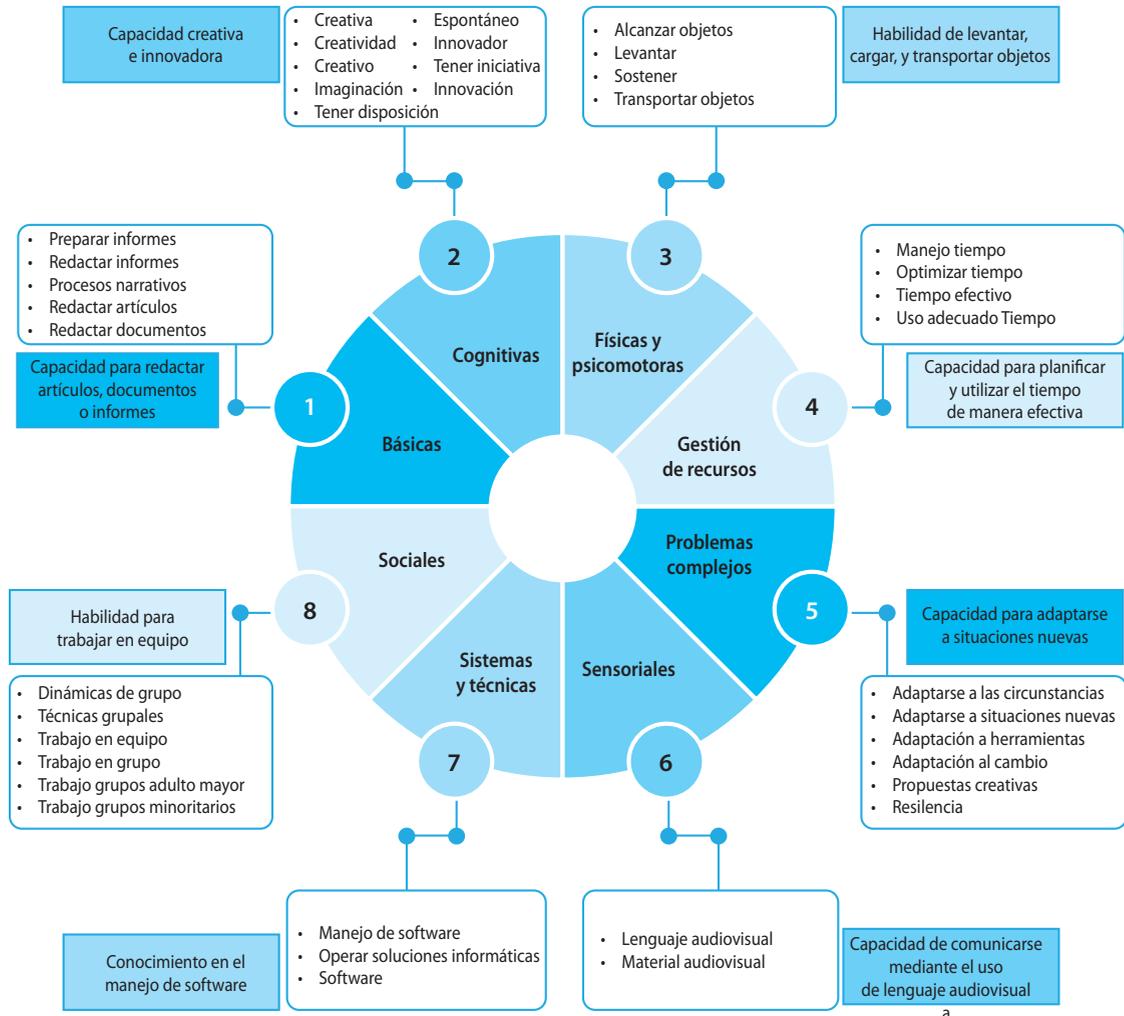
Fuente de información	Cantidad
Vicerreorías de docencia	86
Orientación vocacional	73
Sitios web escuelas	98
Dictámenes de OPES	37
Colegios profesionales	22
Encuesta de habilidades	568
Total	884

Fuente: Elaboración propia.

La figura 5 muestra un ejemplo de la construcción de habilidades, a través del resultado de análisis de texto. Dado que se tienen ocho grupos de habilidades, se visualiza un caso para cada uno de los grupos de palabras que se agrupan en una competencia. Tal como el grupo cuatro de "Gestión de recursos", donde los términos encontrados en el diccionario son manejo tiempo, optimizar tiempo, tiempo efectivo y uso adecuado del tiempo, que conforman a la capacidad para planificar y utilizar el tiempo de manera efectiva.

Respecto al grupo de habilidades "Cognitivas", se tiene que, para la competencia de capacidad creativa e innovadora, los términos correspondientes son: creativa, creatividad, creativo, imaginación, innovación, tener disposición, espontáneo, innovador y finalmente, tener iniciativa, y así sucesivamente, se va presentando un ejemplo para cada grupo de habilidad.

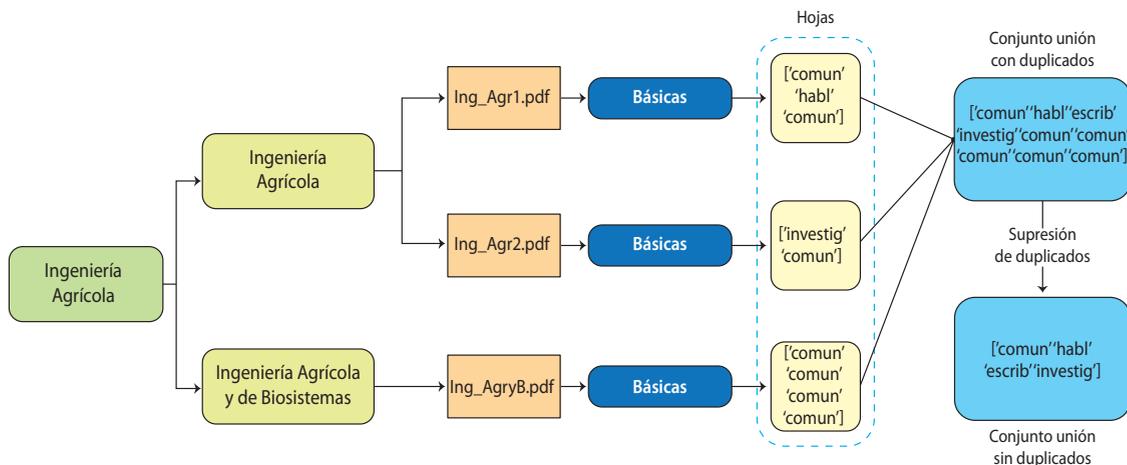
Figura 5. Ejemplo de habilidades en cada grupo



Fuente: Elaboración propia.

A continuación, considerando la cantidad de información recolectada, se da inicio al análisis de texto por medio del algoritmo de *Caribdis*, cuya principal función es la agrupación de prefijos habilidades, según los datos disponibles por disciplina. La figura 6 ejemplifica la funcionalidad de dicho algoritmo. En dicho ejemplo, el nodo de "Ingeniería Agrícola" corresponde a la disciplina, que está integrada por las carreras "Ingeniería Agrícola" e "Ingeniería Agrícola y Biosistemas". Cada una de las carreras cuenta con información en *pdf*. La carrera "Ingeniería Agrícola" incluye dos fuentes de información, mientras que la carrera de "Ingeniería Agrícola y Biosistemas", contiene únicamente una fuente. A partir de este punto, los nodos señalados como hojas, contienen los prefijos de palabras por grupo de habilidad, que particularmente para este ejemplo, se muestra solamente la agrupación de Básicas. Se procede entonces, a realizar la unión de dichos términos, para culminar con la eliminación de duplicados, dejando finalmente un único conjunto de términos sin repeticiones, que representa a las habilidades básicas encontradas para la disciplina de Ingeniería Agrícola.

Figura 6. Ejemplo de conjunto de palabras encontradas en Ingeniería Agrícola



Fuente: Elaboración propia.

La tabla 2 muestra el resultado del algoritmo *Caribdis*, la recolección de información y la estimación de las proporciones, en cada uno de los grupos de habilidades para el nivel de desagregación de área de conocimiento. También, se visualiza la columna de presencia que hace referencia a la cantidad de archivos analizados en cada área. De manera que, Educación contiene la mayor cantidad de información, ya que es el área donde se concentran más disciplinas y carreras. Esta situación es similar para Ciencias Sociales.

Las áreas de Educación, Ciencias Sociales e Ingeniería, concentran sus habilidades en el grupo de Físicas y psicomotoras, esto se debe a las actividades y labores que desarrollan diariamente, pues involucra actividades como mantenerse de pie, utilizar diferentes partes del cuerpo para desarrollar una actividad, o inclusive, la coordinación de movimientos, entre otras habilidades. A su vez, Ciencias Sociales, Ingeniería, Recursos Naturales, Ciencias de la Salud y Computación, cuentan con una proporción alta en el grupo de Gestión de recursos, conformado por competencias para el manejo y aplicación de metodologías, asignación y control eficiente de los recursos.

El área de Artes y Letras, es de los grupos cuya proporción más alta corresponde al grupo de habilidades sensoriales. Por otro lado, para el grupo de habilidades de problemas complejos, las áreas con proporciones altas corresponden a Educación, Ciencias Básicas y Ciencias Económicas; en este grupo, se consideran competencias asociadas a pensamiento crítico, capacidad de trabajar bajo presión, capacidad de adaptarse a las situaciones nuevas, entre otras.

En contraposición, el grupo habilidades sociales muestra valores más bajos en aproximadamente el 80% de las áreas analizadas, donde se incluyen competencias como la actitud colaborativa, la capacidad de establecer relaciones humanas, la capacidad de aplicar principios éticos, el trabajo en equipo, entre otras competencias. Particularmente, en el área de Educación, el grupo de competencias denominado Sistemas y Técnicas presenta el porcentaje más bajo, lo cual tiene concordancia, ya que se concentran competencias como el desarrollo de *software*, uso de programas informáticos, diseño de portales *web*, que no son de la naturaleza del área. Se refleja una mayor ausencia de competencias cognitivas, como el razonamiento deductivo, facilidad para los números, capacidad creativa e innovadora, para el área de Ciencias Sociales.

TABLA 2
 Proporción de habilidades por grupo de habilidad, según el área de conocimiento

Área de conocimiento	Proporción (%)								Frecuencia absoluta
	Físicas y psicomotoras	Cognitivas	Sociales	Básicas	Sistemas y técnicas	Gestión de recursos	Sensoriales	Problemas complejos	Presencia
Educación	83,6	44,0	40,8	61,3	38,1	59,4	65,6	80,8	205
Ciencias Sociales	80,3	45,3	54,0	62,9	52,4	78,1	59,4	73,1	134
Ingeniería	82,0	54,7	38,2	54,8	59,5	78,1	56,3	65,4	130
Ciencias Básicas	60,7	44,0	21,1	45,2	42,9	50,0	46,9	61,5	102
Artes y Letras	77,1	49,3	32,9	50,0	50,0	68,8	78,1	76,9	99
Recursos Naturales	72,1	45,3	29,0	46,8	47,6	65,6	37,5	65,4	78
Ciencias Económicas	68,9	41,3	39,5	56,5	40,5	68,8	50,0	69,2	65
Ciencias de la Salud	72,1	38,7	30,3	46,8	33,3	59,4	43,8	50,0	36
Computación	59,0	30,7	23,7	35,5	50,0	62,5	25,0	50,0	33

Fuente: Elaboración propia.

SÍNTESIS Y REFLEXIONES FINALES

En atención al objetivo de esta investigación se construyen ocho grupos de habilidades, debidamente integradas por competencias asociadas a diferentes temáticas, que permite realizar un análisis integral de los perfiles académicos de las carreras universitarias estatales, donde no solamente se consideran las competencias asociadas a la formación académica, sino a las que involucran el entorno en el que se desarrolla el profesional. Esto permite comparar los hallazgos con otros referentes internacionales, como lo es la base de O*NET 26.0 Database (2020), de naturaleza estadounidense, la cual contiene un conjunto de variables que describen las tareas y las características del puesto y las habilidades requeridas.

El aporte más importante que hace esta investigación es que con el apoyo de herramientas computacionales, se crea la primera base de datos a nivel nacional, que homogeniza la información de las competencias detectadas en los perfiles académicos de las carreras universitarias. Asimismo, la herramienta permite incluir nueva información para tener resultados actualizados y en consecuencia, vincularse con otros sistemas de información.

Un aspecto notable es que la información en los sitios *web* no necesariamente se encuentra actualizada, lo que dificulta el proceso de recolección de información, genera retrasos y se deben duplicar esfuerzos al procesar información. Asimismo, la presencia de documentos para la disciplina y la cantidad de información que contienen, posiblemente influya en el cálculo de las proporciones de las habilidades representadas. Por lo anterior, es que se decide excluir el área de Derecho, pues cuenta con poca información, producto de cambios en la facultad. Para la investigación, solo se analiza Derecho general y se excluyen sus énfasis, ya que dichos énfasis pasaron a ser un posgrado, y para este análisis, no son parte de la población de estudio.

La información que contienen los perfiles académicos, se estructura de manera heterogénea, lo que dificulta la precisión del algoritmo al momento de identificar las habilidades para industria 4.0. Se recomienda diseñar una estructura de perfil académico para las cinco universidades estatales, respaldando

el trabajo interdisciplinario que desarrolla el Conare, en el que se describen, minuciosamente, las competencias y habilidades de cada carrera universitaria.

El cálculo de proporción por área de conocimiento, es un indicador adecuado para perfilar las competencias en cada grupo de habilidad, ya que describe dónde existe una mayor o menor presencia en cada uno de ellos. También, se ve reflejada la naturaleza de la composición del área, tal como se evidenció en el área de Educación, que cuenta con mayor presencia en habilidades Físicas y psicomotoras, Problemas complejos y menor proporción en competencias de Sistemas y Técnicas.

En los resultados, se visualiza una mayor cantidad de hallazgos en habilidades físicas y psicomotoras, producto de las respuestas afirmativas, obtenidas en la encuesta de habilidades particularmente de este grupo. Esto evidencia, la necesidad de que las autoridades académicas universitarias, consideren en sus perfiles académicos, la descripción de habilidades adquiridas referentes a los grupos planteados.

Considerando el insumo que brinda esta investigación y otros estudios realizados por investigadores, como Frey & Osborne (2013), y la reinención de los puestos de trabajo que plantean Jesuthasan & Boudreau (2018), continúa la interrogante, ¿Cuál es la probabilidad de automatización de las carreras universitarias, camino a la cuarta revolución industrial?

REFERENCIAS

- BIRD, S.; LOPER, E. & KLEIN, E. (2009), *Natural Language Processing with Python*. O'Reilly Media Inc. Recuperado el 15 de setiembre del 2021 de <https://pypi.org/project/nltk>
- BENITONE, P. C.; GONZÁLEZ, J.; MALETÁ, M.; SIUFI, G. y WAGENAAR, R. (2007). *Reflexiones y perspectivas de la Educación Superior en América Latina*. España: Publicaciones de la Universidad de Deusto.
- BERROCAL, F.; GARCÍA, A. y RAMÍREZ-VIELMA, R. (2021). La elaboración de modelos de competencias técnicas y su aplicación en la detección de necesidades formativas. *Gestión y Análisis de Políticas Públicas, Nueva Época (GAPP)*, 26, 111-129. <https://doi-ñorg.ezproxy.sibdi.ucr.ac.cr/10.24965/gapp.i26.10813>.
- CERVANTES, G. (2007). Competencias y habilidades en la nueva cultura educativa. Universidad Autónoma Metropolitana. Fuentes Humanísticas. Vol. 19, n.o 35.
- CONSEJO NACIONAL DE RECTORES (1982) Convenio de coordinación de educación superior universitaria estatal en Costa Rica. 20 de marzo de 1982. Artículo 3, inciso 2.
- CORRALES, K. y SANDI, K. (2019). El futuro de las carreras universitarias costarricenses camino a la industria 4.0. San José: CONARE.
- DARWING, C. (1859). El origen de las especies. Estados Unidos: Editorial Charles Darwing.
- FREY, C. & OSBORNE, M. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, Elsevier, vol. 114(C), pp. 254-280.
- JESUTHASAN, R. & BOUDREAU, J.W. (2018). *Reinventing Jobs. A 4-step approach for applying automation to work*. Harvard Business Press. Boston, Massachusetts.
- MACKIE, J. & LIU, R. (2021). PyMuPDF (1.18.13) Adaptador de la librería MuPDF para python. [pypi.org. https://pypi.org/project/PyMuPDF/](https://pypi.org/project/PyMuPDF/)
- MCKINNEY, W. (2021). Pandas (1.2.4) [Herramienta para el análisis de datos]. [pypi.org. https://pypi.org/project/pandas/](https://pypi.org/project/pandas/)
- O*NET 26.0 DATABASE (2020). O*NET Resource Center. Recuperado el 14 de julio del 2020 de <https://www.onetcenter.org/database.html>

- PERASSO, V. (2016). ¿Qué es la cuarta revolución industrial? (¿y por qué debería preocuparnos?). Recuperado el 13 de julio del 2020 de BCC mundo: <https://www.bbc.com/mundo/noticias-37631834>
- PYTHON (2018). Lenguaje de programación. python.org. Versión 3.6.8. <https://www.python.org/downloads/release/python-368/>
- SANTOS, R. (2013). Levantarse y luchar. Editorial Conecta. Barcelona, España.
- TURK, J. T. & STEPHENS, M. (2020). Jellyfish (0.8.8) [Biblioteca para hacer pareo aproximado y fonético]. pypi.org. <https://pypi.org/project/jellyfish>
- VELÁZQUEZ ÁVILA, K. y SANTIESTEBAN NARANJO, E. (2018). Dicotomía entre habilidad y competencia. Opuntia Brava, 9(1), 40-49. Recuperado a partir de <http://opuntiabrava.ult.edu.cu/index.php/opuntiabrava/article/view/116>