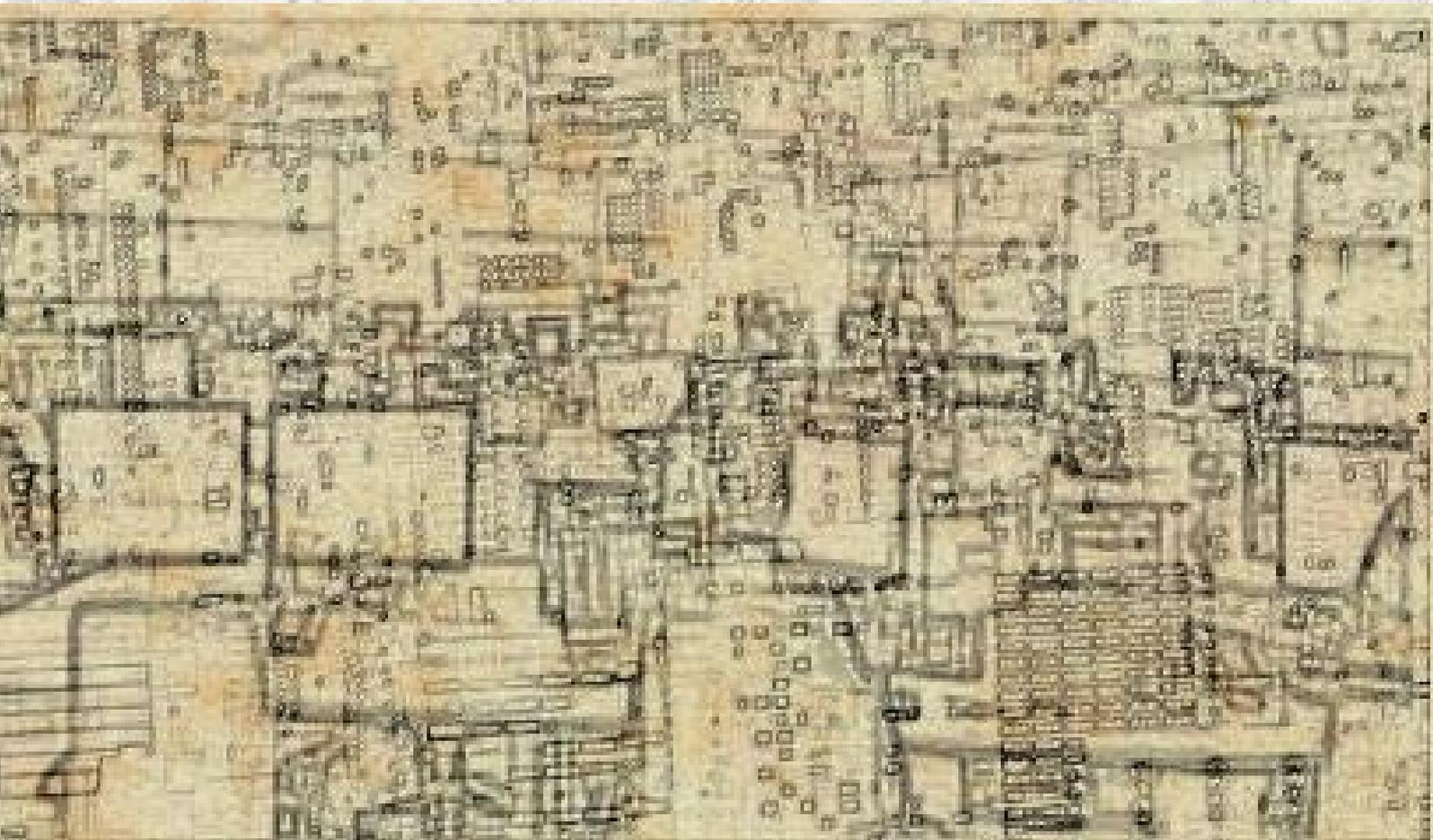


# Análisis de perfiles de profesionales y formación TIC utilizando Internet como fuente de datos

Abril 2020



GOBIERNO  
DE ESPAÑA

VICEPRESIDENCIA  
TERCERA DEL GOBIERNO  
MINISTERIO  
DE ASUNTOS ECONÓMICOS  
Y TRANSFORMACIÓN DIGITAL

ontsi  
red.es

observatorio  
nacional de las  
telecomunicaciones  
y de la SI

DIRECCIÓN Y COORDINACIÓN:

Luis Muñoz López  
Pedro Antón Martínez  
Red.es

Jerónimo Arenas García  
Universidad Carlos III de Madrid

EDICIÓN:

Jesús Cid-Sueiro  
Ángel Navia Vázquez  
Jerónimo Arenas García  
Vanessa Gómez Verdejo  
Saúl Blanco Fortes



©2020 Red.es

Queda prohibida cualquier reproducción total o parcial de este libro,  
almacenamiento en un sistema informático, transmisión en cualquier forma o por  
cualquier medio (electrónico, fotocopia u otros métodos)

<b>ÍNDICE .....</b>	<b>3</b>
<b>INTRODUCCIÓN.....</b>	<b>4</b>
<b>1 ANÁLISIS DE DEMANDA DE PROFESIONALES TICC.....</b>	<b>8</b>
1.1 PROPÓSITO DEL ESTUDIO .....	8
1.2 METODOLOGÍA .....	9
1.2.1 Descarga de las ofertas de empleo.....	9
1.2.2 Preprocesado de las ofertas de empleo.....	10
1.2.3 Identificación de perfiles.....	12
1.2.4 Visualización de resultados .....	12
1.3 FUENTES DE DATOS. ....	13
1.4 RESULTADOS DEL ANÁLISIS .....	14
1.4.1 Perfilado estático en portales de empleo. ....	14
1.4.2 Perfilado dinámico en portales de empleo. ....	23
1.4.3 Perfilado jerárquico en portales de empleo basado en CNOs .....	29
1.5 APÉNDICE TÉCNICO 1: METODOLOGÍA ML. ....	39
1.5.1 Proceso de clasificación de ofertas de empleo según CNOs. ....	39
<b>2 ANÁLISIS DE PROGRAMAS FORMATIVOS TICC.....</b>	<b>44</b>
2.1 PROPÓSITO DEL ESTUDIO .....	44
2.2 FUENTES DE DATOS. ....	44
2.2.1 Fuentes para el análisis de títulos universitarios.....	44
2.2.2 Fuentes para el análisis de cualificaciones profesionales.....	46
2.3 METODOLOGÍA .....	47
2.4 RESULTADOS DEL ANÁLISIS .....	47
2.4.1 Análisis de títulos universitarios.....	48
2.4.2 Análisis de cualificaciones profesionales.....	49

## Introducción

El presente informe es el resultado de las actividades del contrato titulado "Mantenimiento Evolutivo de Recopilación de Información en Internet como Fuente de Datos sobre Perfiles Profesionales y sobre Características de la Transformación Digital en las Empresas" (Exp. 060/18-OT) (al que, en lo sucesivo, nos referiremos con la denominación abreviada "Proyecto IaD2019"). La expresión "Internet como Fuente de Datos" (IaD, Internet As a Data source), hace referencia al uso de técnicas avanzadas de análisis de datos que pueden servir de Internet como fuente complementaria o sustitutiva de fuentes tradicionales de datos estadísticos. Lo que distingue a estos métodos de otras técnicas también orientadas a Internet, como las encuestas online, es su carácter automático, no intrusivo (pues no requiere la participación activa de terceros) y exhaustivo (pues permite la exploración completa de las fuentes sin recurrir a métodos de submuestreo). Además, su carácter completamente automático conlleva una mayor inmediatez en la obtención de resultados, y la posibilidad de repetir el análisis con la frecuencia que se desee.

Este proyecto puede considerarse evolución del Proyecto Piloto sobre Viabilidad de Utilizar Internet como Fuente de Datos realizado en 2014 (al que, en lo sucesivo, nos referiremos como "**Proyecto IaD2014**") y del Servicio de Mantenimiento Evolutivo para el Proyecto Piloto sobre Viabilidad de Utilizar Internet como Fuente de Datos", realizado en 2015<sup>1</sup> ("**Proyecto IaD2015**"). El propósito de este nuevo contrato ha sido doble:

- Continuar las tareas de análisis y visualización de datos, pero con datos nuevos tomados de procesos de captura realizados en 2017, 2018, 2019 y 2020 en diferentes períodos de tiempo.
- Crear una nueva versión de la aplicación software que permita mejorar la calidad y fiabilidad de los resultados del análisis, incorporando nueva funcionalidad que permita sacar mayor provecho de las fuentes de datos y del análisis.

Las actividades del proyecto se han organizado en dos grandes subproyectos. El primero ha estado orientado a la aplicación de técnicas de clasificación automática para detectar y caracterizar la transformación digital en las empresas en España. La finalidad del segundo ha sido la aplicación de técnicas de análisis de datos y

---

<sup>1</sup> Los resultados del análisis de datos de este proyecto están accesibles a través de la visualización interactiva en <http://iad.ontsi.es/perfilestic2020/cno.html>

aprendizaje automático para detectar y caracterizar la oferta y la demanda de profesionales de las TICC (*Tecnologías de la Información, las Comunicaciones y los Contenidos*) en España. En ambos casos, la automatización pretende evitar o minimizar las tareas de exploración o anotación manual de sitios web (de empresas, portales de empleo o catálogos oficiales de titulaciones).

Este documento describe las actividades del segundo subproyecto, sobre la detección y caracterización de ofertas de empleo. En general, se ha intentado centrar la exposición sobre los resultados del análisis de datos, omitiendo los detalles más técnicos, que en todo caso pueden consultarse en los apéndices al final de cada capítulo.





# 1

## ANÁLISIS DE DEMANDA DE PROFESIONALES TICC



VICEPRESIDENCIA  
TERCERA DEL GOBIERNO  
MINISTERIO  
DE ASUNTOS ECONÓMICOS  
Y TRANSFORMACIÓN DIGITAL

ontsi  
observatorio  
nacional de las  
telecomunicaciones  
y de la SI  
red.es

# 1 Análisis de demanda de profesionales TICC

## 1.1 Propósito del estudio

### **OBJETIVO**

---

**ANALIZAR** automáticamente, mediante exploración web y análisis de datos, la **OFERTA DE EMPLEO** y los **PLANES FORMATIVOS** en el **SECTOR TICC**

En este capítulo analizaremos los perfiles demandados en el sector TICC en España en el período comprendido desde julio de 2016 hasta marzo de 2020. Para ello, se han actualizado las herramientas disponibles en los proyectos IaD2014 e IaD2015, extendiendo el estudio en las siguientes direcciones:

- Estudio trimestral de los perfiles de empleo, utilizando para ello el histórico de datos disponible, así como estudio para el intervalo completo.
- Nuevo procedimiento para el etiquetado automático de los perfiles obtenidos, que ha sido complementado con supervisión humana.
- Clasificación automática de las ofertas de empleo según los Códigos Nacionales de Ocupación (CNO) y las Categorías Europeas de Competencias y Habilidades (ESCO). El procedimiento implementado permite asignar cada oferta de empleo a uno o varios códigos CNO y/o ESCO.
- Generación de modelos de tópicos basados exclusivamente en tecnologías y competencias técnicas identificadas de forma explícita en las ofertas de empleo publicadas. Dicha lista de tecnologías ha sido depurada de forma manual para fusionar términos equivalentes y eliminar variantes redundantes.
- Implementación de una herramienta de Inteligencia de Negocio que permite explotar todos los metadatos disponibles en los portales de empleo, y visualizar las relaciones entre los perfiles detectados, las categorías de ocupación (códigos CNO y ESCO), rangos salariales, información geográfica y temporal, etc.
- Identificación de ofertas que mencionan explícitamente que están dirigidas a hombres y mujeres.
- Nuevo procedimiento de perfilado jerárquico, consistente en buscar ofertas de empleo / titulaciones similares, y caracterización temática de cada uno de los grupos de ofertas / titulaciones identificados.
- Análisis de la evolución temporal de los perfiles en el marco temporal mencionado anteriormente, mediante la utilización de un novedoso método de análisis temporal que permite aglutinar los modelos entrenados para distintos trimestres, y comparar los perfiles de cada ventana temporal con las de un modelo único.

## 1.2 Metodología

El análisis de la demanda de profesionales del sector TICC se basa fundamentalmente en la detección automática de los perfiles más demandados por las empresas, a partir de la información disponible principalmente en portales de empleo. A estos efectos, consideramos que un perfil se caracteriza a partir de un conjunto de términos (habitualmente técnicas o conocimientos específicos) que suelen aparecer conjuntamente en las ofertas de empleo. A modo de ejemplo:

- Programador Java, caracterizado por los términos Java, J2ee, etc.
- Analista de datos, caracterizado por los términos big\_data, business\_intelligence, python, etc.

Es importante remarcar que la identificación de perfiles se realiza de forma automática a partir de las ofertas de empleo, esto es, con una mínima intervención manual.

La metodología de trabajo comprende los siguientes pasos que se detallarán en las subsecciones siguientes:

1. Descarga de las ofertas de empleo
2. Preprocesado textual de las ofertas de empleo
3. Identificación de perfiles mediante modelado de tópicos
4. Visualización de resultados

### 1.2.1 Descarga de las ofertas de empleo

Tres han sido los portales de empleo rastreados durante la realización del presente proyecto: *TecnoEmpleo*, *Infojobs*, e *Infoempleo*. Dos han sido las estrategias de descarga empleadas según las propias restricciones de acceso establecidas por los portales de empleo: 1/ Rastreo directo mediante la descarga de ofertas vía web, implementando arañas específicas que permiten parsear la información de cada portal y su carga en una base de datos relacional MySQL, y 2/ Acceso via REST API para los portales que tienen implementado y accesible dicho servicio. En general, el uso de REST API proporciona una opción más robusta y estable, en el sentido de que la importación de datos no se ve afectada por cambios en el diseño de los portales, que son habitualmente más frecuentes que los producidos en las interfaces REST y pueden dejar la araña inservible. Además, en muchos casos, los cambios en versiones de las interfaces REST se llevan a cambio garantizando la compatibilidad hacia atrás.

De los tres portales mencionados, los datos presentados en este informe se refieren exclusivamente al portal *Tecnoempleo.com*, centrado exclusivamente en publicación de ofertas del sector TICC, y por tanto el 100% de las ofertas descargadas de dicho portal resultan de interés directo para el análisis. Esto es así debido a que es para este portal para el que se dispone del conjunto más amplio de datos. El tipo de acceso implementado es de tipo CRAWLER hasta junio de 2019, disponiendo a partir de dicha fecha de un token de acceso a la interfaz REST API proporcionado por la empresa, y que ha permitido migrar la descarga de ofertas a esta interfaz más robusta. Se han implementado los procedimientos necesarios para poder utilizar de manera conjunta todos los datos disponibles independientemente del tipo de acceso.

En el caso de Infojobs, se dispone de datos temporales obtenidos vía REST hasta aproximadamente comienzos de 2019. Sin embargo, cambios implementados en la política de acceso por parte de la empresa propietaria del portal han implicado que dicho acceso haya sido deshabilitado a partir de la fecha mencionada, siendo también inviable el uso de técnicas de rastreo masivo (el portal bloquea el acceso a robots). A raíz de esta problemática se implementó la descarga vía CRAWLER de ofertas del portal InfoEmpleo, si bien no se ha considerado muy informativo la inclusión de este portal dado que se dispone de menor rango temporal, y además el número de ofertas publicadas para el sector TICC es significativamente inferior a las disponibles para *TecnoEmpleo.com*.

A pesar de que los resultados presentados en el informe se limitan al portal *Tecnoempleo.com*, hay que resaltar que la metodología de análisis utilizada es completamente independiente de la fuente de datos, por lo que queda abierta la posibilidad de replicar el análisis para una segunda fuente de datos en el momento en que esta se encuentre disponible y se disponga de un registro temporal suficientemente amplio.

### 1.2.2 Preprocesado de las ofertas de empleo

Este proceso implica la transformación de cada oferta en un conjunto de términos que se consideran más característicos de la oferta, por tanto eliminando términos comunes que se esperan puedan aparecer en todas las ofertas, así como otros elementos del texto que no aportan información semántica (puntuación, formas verbales, plurales, etc.). Concretamente, los pasos que comprende el preprocesado son los siguientes:

1. Identificación de términos, fundamentalmente siguiendo el criterio de delimitación por espacios u otros caracteres como el punto o la coma. Sin embargo, se han incluido reglas

específicas, que permiten que términos como por ejemplo asp.net se consideren de forma única, y no como dos términos diferenciados.

2. Filtrado de ofertas en lengua no española.
3. Eliminación de palabras comunes (stopwords), de acuerdo a listas predefinidas de dichos términos.
4. Identificación de n-gramas. Un n-grama es una concatenación de términos que, al aparecer conjuntamente, cobran un significado diferente a la suma de los significados individuales de cada término. El uso de n-gramas puede verificarse en los resultados de la aplicación cuando aparezcan términos que incluyen el carácter '\_' (p.ej., big\_data, bases\_de\_datos, etc).
5. Identificación de plurales. La aplicación conserva la palabra con mayor índice de aparición en la totalidad de las ofertas de empleo, con independencia de su número.
6. Obtención del vocabulario: se ha incluido un paso adicional "manual", consistente en la búsqueda de palabras equivalentes, y edición manual de stopwords a partir del entrenamiento de un primer modelo de perfiles. Para la búsqueda de equivalencias, se ha utilizado también la lista de todas las "skills" disponibles en el campo correspondiente de la base de datos, unificando tecnologías equivalentes o variantes morfológicas de la misma tecnología en una única forma común.  
Este paso adicional, ha permitido obtener un vocabulario en el que: 1/ se ha minimizado el número de expresiones no deseadas, 2/ se han unificado gran número de expresiones equivalentes, incluyendo acrónimos de uso habitual.
7. Generación de una nueva representación de cada oferta de empleo, consistente en contar cuántas veces aparecen en cada documento los términos del vocabulario.

### 1.2.3 Identificación de perfiles

La metodología empleada utiliza herramientas de **INTELIGENCIA ARTIFICIAL** concretamente **PROCESAMIENTO DE LENGUAJE NATURAL Y MACHINE LEARNING (MODELADO DE TÓPICOS)** para encontrar los conjuntos de tecnologías que son demandadas conjuntamente y que definen cada **PERFIL PROFESIONAL**

Utilizando la representación anterior, se buscan términos que ocurren con alta frecuencia en un número significativo de documentos. Para esta tarea se emplean técnicas de aprendizaje automático, concretamente de análisis de tópicos, y se obtienen fundamentalmente las salidas siguientes:

1. Listado de perfiles. En primera aproximación, podemos asumir que cada perfil se caracteriza por un conjunto de palabras altamente significativas. Con mayor precisión, un perfil asigna en realidad un peso a todas las palabras del vocabulario, por lo que todas las palabras forman en realidad parte de todos los perfiles, si bien a la hora de caracterizar un perfil es habitual recurrir únicamente al listado de las palabras más relevantes de cada perfil.
2. Explicación de cada oferta a partir de los perfiles identificados. Existirán ofertas que se asignen de forma única a un perfil concreto, si bien lo habitual es que cada oferta se asigne a varios perfiles en distinta proporción. En otras palabras, considerando que se han identificado 5 perfiles distintos, una oferta puede considerarse que pertenece de forma pura al tercero de dichos perfiles, mientras que otra oferta podría pertenecer a todos los perfiles a la vez, por ejemplo con porcentajes 70%, 4%, 20%, 3% y 3% (obsérvese que la suma de dichos porcentajes es 100).
3. Tamaño de cada perfil en el corpus. En base al grado de pertenencia de cada documento a los distintos perfiles, podemos estimar la importancia global de cada perfil.

### 1.2.4 Visualización de resultados

Finalmente, las salidas enumeradas en la sección anterior se combinan con los metadatos de cada oferta de trabajo, para generar los ficheros de resultados y otros archivos para la visualización de resultados en una herramienta web diseñada *ad-hoc* para la visualización de perfiles. La herramienta proporcionada implementa dos paneles diferenciados:

1. Visualización de perfiles profesionales. Permite explorar los perfiles más relevantes de cada modelo entrenado, mostrando para cada perfil los términos del vocabulario más relevantes, y las ofertas de empleo más características de los mismos. Esta herramienta permite, asimismo, explorar las ofertas de empleo publicadas, viendo cuáles son los perfiles asociados a cada oferta de empleo y en qué proporción.

- Herramienta de Inteligencia de Negocio (*Business Intelligence, BI*), que permite visualizar de forma comparativa los códigos de ocupaciones (CNO y ESCO), los perfiles extraídos, y los metadatos disponibles, como datos temporales, geográficos, rangos salariales, etc. Dicho panel permite filtrar según uno o varios criterios, recalculando las distribuciones del resto de variables para el subconjunto de ofertas seleccionadas por el usuario.

### 1.3 Fuentes de datos.

El GRÁFICO 1 muestra el número de ofertas descargadas del portal tecnoempleo.com durante cada trimestre analizado en el presente proyecto. Podemos observar que el número de ofertas es relativamente estable en los diversos cuatrimestres, disponiendo de un histórico no relevante en cuanto a su volumen para los trimestres anteriores al tercero de 2016. Existe un vacío de dos trimestres en 2018, en los que el proceso automatizado de descargas estuvo detenido por causas informáticas.

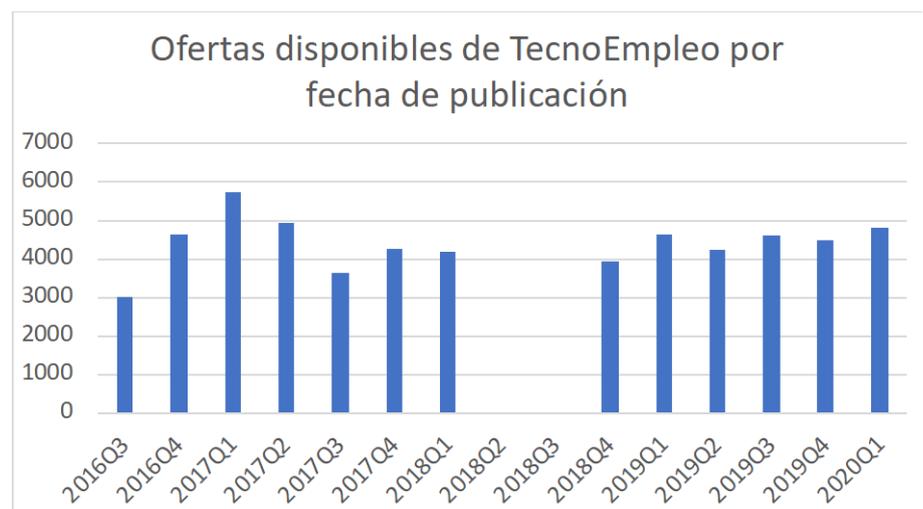
Cabe mencionar que el portal tecnoempleo.com se refiere exclusivamente al sector TICC en el que se centra el presente proyecto.

**GRÁFICO 1: EVOLUCIÓN DEL NÚMERO DE OFERTAS EN EL PORTAL TECNOEMPLO.COM PARA CADA TRIMESTRE ANALIZADO**

#### FUENTE DE DATOS

**57.026**

**OFERTAS DE EMPLEO** tomadas diariamente del portal **tecnoempleo.com** y analizadas en intervalos trimestrales



A partir de junio de 2019 la descarga se ha realizado utilizando un token para acceso por API facilitado por el portal tecnoempleo.com. Para las ofertas descargadas se ha podido obtener la siguiente

información (en algunos casos los campos pueden estar vacíos) - véase también la Tabla 1:

- Un código alfanumérico que permite identificar de manera única a cada oferta.
- Título de la oferta de empleo
- Descripción de la oferta
- Requisitos y habilidades demandadas (*skills*)
- Provincia del puesto de trabajo
- URL de publicación de la oferta de empleo
- Fecha de descarga de la oferta, y última fecha en la que la oferta seguía disponible en el portal
- Tipo de contrato ofrecido y dedicación
- Rango salarial ofrecido
- Experiencia mínima requerida

**TABLA 1: PORTAL TECNOEMPLO. METADATOS DISPONIBLES**

ofertaTecnoEmpleo	
id	varchar(30)
file	varchar(255)
published	datetime
urgent	varchar(4)
province	varchar(255)
province_extra	varchar(255)
professionalSkills	text
link	varchar(255)
country	varchar(255)
company	varchar(155)
contractType	varchar(155)
workDay	varchar(155)
experienceMin	varchar(255)
study	varchar(255)
skills	text
salaryMin	int(11) unsigned
salaryMax	int(11) unsigned
salaryPeriod	varchar(255)
description	text
city	varchar(255)
fechaDescarga	datetime
tipoDescarga	varchar(255)

## 1.4 Resultados del análisis

### 1.4.1 Perfilado estático en portales de empleo.

La herramienta web desarrollada permite un análisis exhaustivo de los perfiles de empleo detectados en cada trimestre del histórico de datos disponible, así como considerar de forma conjunta las descargas realizadas durante todos los meses. Asimismo, se han incorporado modelos obtenidos empleando únicamente el campo

“skills” proporcionado por la API de tecnoempleo.com, que contiene exclusivamente las competencias demandadas por el ofertante de empleo, separadas por comas. Al ser un campo de fácil tratamiento y carácter exclusivamente técnico, el uso exclusivo de este campo permite obtener perfiles de carácter exclusivamente técnico, y eliminar otros de carácter transversal que aparecen al utilizar la descripción completa. Además, cabe recordar que las tecnologías que aparecen listadas dentro de dicho campo *skills* han sido procesadas de manera semiautomática, fusionando en un único término variantes morfológicamente diferentes de la misma tecnología, lo que permite anticipar una mayor calidad en la descripción de los perfiles obtenidos.

Actualmente, la herramienta incorpora los siguientes conjuntos de perfiles:

- Modelos trimestrales de 20 perfiles
- Modelos trimestrales de 15 perfiles empleando únicamente el campo de “skills”
- Modelo de 19 perfiles<sup>2</sup> para las ofertas disponibles en todo el intervalo temporal
- Modelos de 14 y 21 perfiles<sup>3</sup> para las ofertas disponibles en todo el intervalo temporal, empleando únicamente el campo de “skills”.

La pantalla principal de la aplicación (GRÁFICO 2) permite seleccionar el corpus, período temporal (en su caso) y número de tópicos del modelo, mostrándose además otra información de interés, como el número total de documentos que se han analizado en cada modelo concreto.

Tras seleccionar un modelo, se puede acceder a distintas visualizaciones que ofrecen una visión global de los perfiles detectados y su relevancia (tamaño), composición de cada perfil junto a un listado de documentos que pertenecen en gran proporción a dicho perfil (GRÁFICO 3), y visualizador de documentos que posibilita acceder a cada oferta de empleo por su título, y muestra los tópicos más relevantes para dicha oferta (GRÁFICO 4).

---

<sup>2</sup> El número original de tópicos del modelo era de 25 perfiles, habiéndose reducido dicho número a 19 por medio de la herramienta de edición de tópicos incorporada en el desarrollo SW.

<sup>3</sup> Correspondientes a modelos originales de 15 y 25 tópicos, respectivamente, editados de manera manual para eliminar perfiles poco informativos o ruidosos.

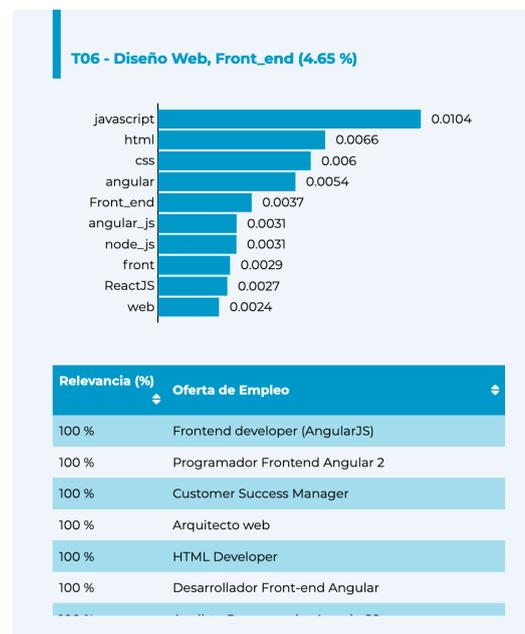
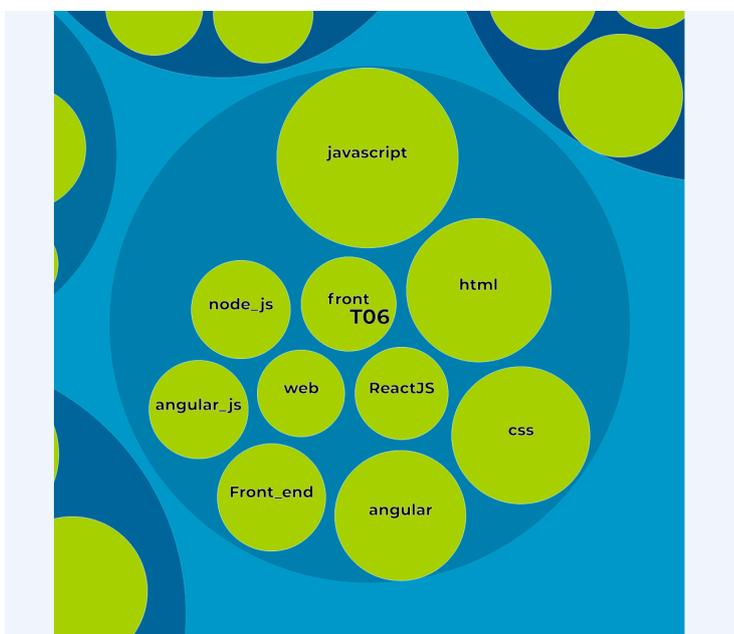
## GRÁFICO 2: INFORMACIÓN MOSTRADA POR LA PÁGINA PRINCIPAL DEL VISUALIZADOR DE PERFILES



Un análisis exhaustivo de todos los modelos de perfiles para todos los períodos considerados sería excesivamente extenso, por lo que en este informe revisaremos únicamente dos modelos que pueden considerarse representativos de todos los modelos accesibles a través de la herramienta:

- Modelo para todas las ofertas disponibles, empleando los campos de "descripción" y "skills"
- Modelo para todas las ofertas disponibles, empleando únicamente el campo de competencias ("skills").

## GRÁFICO 3: INFORMACIÓN MOSTRADA POR EL VISUALIZADOR DE PERFILES PARA CADA PERFIL CONCRETO



## GRÁFICO 4: COMPOSICIÓN TEMÁTICA DE LA OFERTA DE EMPLEO "DESARROLLADOR BIG DATA - HADOOP"



### Perfiles de tecnoempleo.com, 19 perfiles

La Tabla 2 muestra los perfiles más relevantes obtenidos al analizar las ofertas del portal tecnoempleo.com, considerando de manera conjunta el título, la descripción, y las competencias proporcionadas por el portal para cada oferta. En base a las palabras que caracterizan cada perfil se ha realizado un etiquetado semiautomático, empleando en primer lugar una herramienta basada en aprendizaje automático, y ajustando posteriormente las etiquetas propuestas, incluso corrigiéndolas en caso necesario.

En general, podemos observar que los perfiles quedan caracterizados por palabras que pueden considerarse afines, tal y como se esperaba. Asimismo, los perfiles obtenidos pueden considerarse razonablemente diferentes entre sí, si bien obviamente existen algunos perfiles concretos en los que hay cierto solape (e.g., Desarrollo Aplicaciones Web y Diseño Web Front End, dos perfiles diferentes de Soporte, etc.). No obstante, lo anterior, si tratásemos de obtener un modelo con menos perfiles, se perdería la riqueza de algunos de los perfiles detectados, por lo que la selección del número de perfiles implica un cierto compromiso entre la especificidad de los perfiles detectados, y la existencia de solapes entre perfiles.

Por otro lado, la inclusión de términos de carácter no exclusivamente técnicos tiene como consecuencia la aparición de perfiles que están asociados a competencias transversales o a beneficios del puesto del trabajo. De ellos se han eliminado 6 por considerarse ruidosos o poco relevantes (p.ej, relacionados con beneficios adicionales del puesto, etc.) manteniendo otros que, si bien son genéricos, incluyen competencias de tipo técnico (no necesariamente TICC). Por ejemplo, se ha mantenido un perfil no técnico relacionado con

habilidades para la Gestión de Proyectos y Coordinación de Equipos. De esta manera, son finalmente 19 los perfiles que se conservan en este modelo.

**TABLA 2: PORTAL TECNOEMPLEO.COM. MODELO DE 19 PERFILES**

Nº perfil	Importancia	Términos
0	7.3%	Analista Programador Java java, j2ee, spring, hibernate, maven, struts, API_REST, oracle, eclipse, JSF
1	7.1%	Programador Bases de Datos Oracle_PL/SQL, java, Analista_Programador, tecnológico, consultoría, ingeniería, multinacional, desarrollo, sistemas, programador
2	6.1%	Diseño Web, Front End javascript, html_5, angular, css, css3, jquery, html, Front_end, node_js, ReactJS
3	5.5%	Desarrollo Aplicaciones Web .net, C, MS_SQL_Server, asp_net, sharepoint, visual_studio, visual_basic, MVC, Analista_Programador, entity_framework
4	5.1%	Helpdesk cobol, helpdesk, soporte, incidencia, windows, hardware, usuario, DB2, host, CICS
5	4.8%	Gestión de Proyectos, Coordinación de Equipos negocio, comercial, liderar, cliente, Jefe_de_Proyecto, project_manager, estrategia, producto, gestión_de_proyectos, venta
6	4.4%	Ingeniero de Pruebas test, Support, English, question_answering, Agile, software, scrum, jira, security, selenium
7	4.4%	Desarrollo Web PHP - Symfony PHP, mysql, C, symfony, laravel, linux, drupal, framework, programación, git
8	3.6%	Cloud Computing, Escalabilidad, Microservicios python, AWS, docker, devops, cloud, jenkins, microservicios, git, MS_Azure, API

Nº perfil	Importancia	Términos
9	3.5%	Helpdesk incidencia, SQL, resolución, instalación, configuración, BBDD, soporte, administración, Mantenimiento, MS_SQL_Server
10	3.5%	Administración sistemas linux linux, unix, redhat, shell_script, tibco, jboss, weblogic, AIX, sistemas, Apache
11	3.5%	Analista de Datos, BI, Big Data data, BI, ETL, microstrategy, SQL, Qlikview, powercenter, SAS, Big_Data, Excel
12	3.2%	Experto SAP framework SAP, consultor, abap, SAP_FI_CO, módulo, SAP_MM, SAP_HANA, MM, SAP_SD, SD
13	3.2%	Ingeniero de Redes redes, cisco, firewall, comunicaciones, red, LAN, seguridad, routing, VPN, WAN
14	3.1%	Bases de Datos, CRM oracle, MS_CRM_Dynamics, DBA, consultor, gestión_de_proyectos, navision, trato, través, internacional, CRM
15	2.6%	Desarrollo Aplicaciones Móviles android, iOS, swift, APP, mobile, móvil, movilidad, xamarin, nativo, ionic
16	2.3%	Big Data Big_Data, TIC, social, spark, hadoop, digital, scala, tecnológico, colaborador
17	2.2%	Ciberseguridad seguridad, ITIL, ciberseguridad, consultor, riesgos, ITSM, Excel, ISO, vulnerabilidades, SIEM
18	2.1%	Experto en Marketing y Redes Sociales wordpress, web, marketing, SEO, contenido, prestashop, photoshop, UX, diseño, html

Las limitaciones que acabamos de enumerar son intrínsecas al empleo de herramientas de perfilado completamente automáticas. Como ya hemos comentado, para obtener un modelo de tópicos lo más interpretable posible, el proyecto proporciona una herramienta que permite a un usuario humano llevar a cabo las siguientes acciones:

- Fusión supervisada de tópicos
- Eliminación supervisada de tópicos
- Etiquetado automático y manual de tópicos

Sin embargo, no se aconseja el uso de la herramienta cuando el fin del análisis no sea el propio análisis del modelo de tópicos, sino obtener una representación “temática” de las ofertas de empleo.

### ***Perfiles de tecnoempleo.com basados exclusivamente en competencias, 14 perfiles***

Se muestran a continuación los perfiles obtenidos al analizar las ofertas empleando únicamente el campo de “skills” que incluye únicamente competencias técnicas. En este caso, únicamente ha sido necesario eliminar un perfil que podríamos considerar tiene carácter transversal o poco definido. El resto de perfiles queda definido por términos asociados a tecnologías o conocimientos muy concretos del ámbito TICC.

**TABLA 3. PORTAL TECNOEMPLEO.COM. MODELO DE 14 PERFILES EMPLEANDO ÚNICAMENTE COMPETENCIAS LISTADAS EN EL CAMPO “SKILLS”**

Nº perfil	Importancia	Términos
0	12%	Analista Programador Java java, j2ee, spring, hibernate, maven, struts, API_REST, JSF, JPA, eclipse
1	9.8%	Administración Sistemas Linux linux, windows, unix, helpdesk, soporte, hardware, redhat, redes, sistemas_operativos, Técnico_de_Sistemas
2	9.7%	Diseño web, front end javascript, html_5, angular, css, html, jquery, css3, Front_end, node_js, ReactJS

Nº perfil	Importancia	Términos
3	7.4%	Desarrollo de Aplicaciones Web .net, C, asp_net, MS_SQL_Server, sharepoint, visual_studio, visual_basic, MVC, entity_framework, vb_net
4	7%	Desarrollador en Banca, Seguros. Ciberseguridad seguridad, cobol, DB2, ciberseguridad, comercial, host, CICS, Jefe_de_Proyecto, banca, riesgos
5	6.7%	Bases de Datos oracle, Oracle_PL/SQL, MS_CRM_Dynamics, ERP, SQL, MS_SQL_Server, CRM, BBDD_Oracle, DBA, navision
6	5.9%	Metodologías ágiles test, question_answering, Agile, scrum, jira, API, selenium, jenkins, git, prueba
7	5.9%	Desarrollo Aplicaciones Móviles PHP, android, iOS, symfony, swift, laravel, mysql, APP, mobile, git
8	5.6%	Técnico de Sistemas Windows, Virtualización vmware, windows, Técnico_de_Sistemas, windows_server, citrix, MS_Office_365, backup, Virtualización, active_directory, Microsoft
9	5.4%	Ingeniero de Redes redes, cisco, Support, security, firewall, LAN, Excel, network, red, routing
10	4.8%	Analista de Datos, Big Data python, AWS, devops, docker, Big_Data, cloud, MS_Azure, kubernetes, ansible, spark
11	4.7%	Experto SAP framework SAP, consultor, abap, SAP_FI_CO, SAP_MM, SAP_HANA, MM, SAP_SD, SD, FI
12	4.6%	Analista de Datos, BI BI, data, ETL, microstrategy, Qlikview, powercenter, SAS, consultor, SSIS, Analytics

Nº perfil	Importancia	Términos
13	4%	Experto en Marketing y Redes Sociales wordpress, marketing, salesforce, web, APP, tibco, drupal, SEO, helpdesk, cloud

Un análisis comparativo de los dos modelos presentados permite obtener las siguientes conclusiones:

- Los perfiles técnicos que se observan en ambos casos son razonablemente parecidos, si bien existen algunos de ellos que pueden ser algo más específicos en uno u otro modelo. Por ejemplo, el perfil de programador aparece más agrupado en el modelo de 14 perfiles que en el de 19, en el que se pueden encontrar alguna especialización adicional de dicho perfil.
- Si bien el tamaño relativo de los perfiles puede considerarse relativamente estable, no podemos obtener la misma conclusión al analizar los tamaños absolutos. Esto es así porque:
  - El tamaño general de los perfiles es mayor en el modelo que tiene menos perfiles
  - Mientras que el modelo de 22 perfiles estima el tamaño de cada perfil a partir de todos los campos de texto, el modelo de 15 perfiles emplea únicamente el campo de "skills". No se puede asumir a priori que la proporción de los perfiles sea la misma al analizar un campo normalizado de pequeña longitud, o la oferta completa.
  - Por la propia naturaleza no determinista del modelo de tópicos.

Esta característica de los modelos de tópicos hace necesario que cualquier análisis comparativo entre corpus heterogéneos o intervalos temporales diferentes deba realizarse necesariamente a partir de un modelo de tópicos común. En caso contrario, las conclusiones podrían no ser válidas, y las diferencias observadas deberse únicamente a la distinta composición de perfiles en uno y otro modelo, o bien a los hiperparámetros empleados para el entrenamiento de los modelos de tópicos.

- Finalmente, podemos concluir que el vocabulario empleado parece razonablemente adecuado para el modelado de la oferta de empleo, ya que el número de términos que podríamos considerar no deseados es muy reducido, y apenas se observan términos equivalentes expresados con distintas formas en las descripciones de los tópicos obtenidos.

### 1.4.2 Perfilado dinámico en portales de empleo.

Las competencias demandadas en cualquier área de conocimiento son susceptibles de variar a lo largo del tiempo, y la detección de dichas tendencias resulta posible empleando técnicas de perfilado que permitan llevar un seguimiento de la evolución de los distintos perfiles a lo largo del tiempo. Dichas herramientas permiten, además, detectar eventos de importancia como pueden ser:

- La aparición de nuevos perfiles
- La desaparición de perfiles
- Cambios en la "composición" de cada perfil, es decir, para un mismo perfil pueden detectarse nuevas tecnologías emergentes que cobran importancia, mientras otras se hacen obsoletas.

En los estudios IaD2014 e IaD2015 el análisis dinámico se basaba en el uso de un algoritmo de detección dinámica de tópicos que presentaba algunas limitaciones:

- Mala escalabilidad para analizar conjuntos de datos de gran tamaño
- Variaciones poco significativas del tamaño de los tópicos

Para corregir estas limitaciones, se ha incorporado un nuevo modelo de dos niveles que permite caracterizar las variaciones temporales de los tópicos y permite modelar conjuntos de datos de gran tamaño. El primer nivel de dicho modelo consiste en el entrenamiento de un modelo de tópicos diferenciado para cada ventana temporal bajo consideración, en nuestro caso, intervalos de un trimestre de duración. Para poder analizar de forma conjunta dichos modelos trimestrales, se entrena un modelo adicional de "segundo nivel", que tiene como objetivo explicar todos los modelos de nivel 1. Por lo tanto, es el modelo de nivel 2 el que permite capturar las variaciones temporales de los tópicos. Por otro lado, dado que el nuevo modelo se basa en la construcción iterada de diversos modelos estáticos sobre conjuntos de datos de tamaño reducido, esta propuesta presenta unas propiedades muy adecuadas de escalabilidad.

La Tabla 4 muestra los 15 perfiles obtenidos mediante la herramienta de análisis dinámico de tópicos en orden de relevancia. Para cada tópico se muestran los términos más relevantes para la caracterización del perfil.

El perfilado dinámico permite explorar la evolución de los perfiles de empleo más demandados

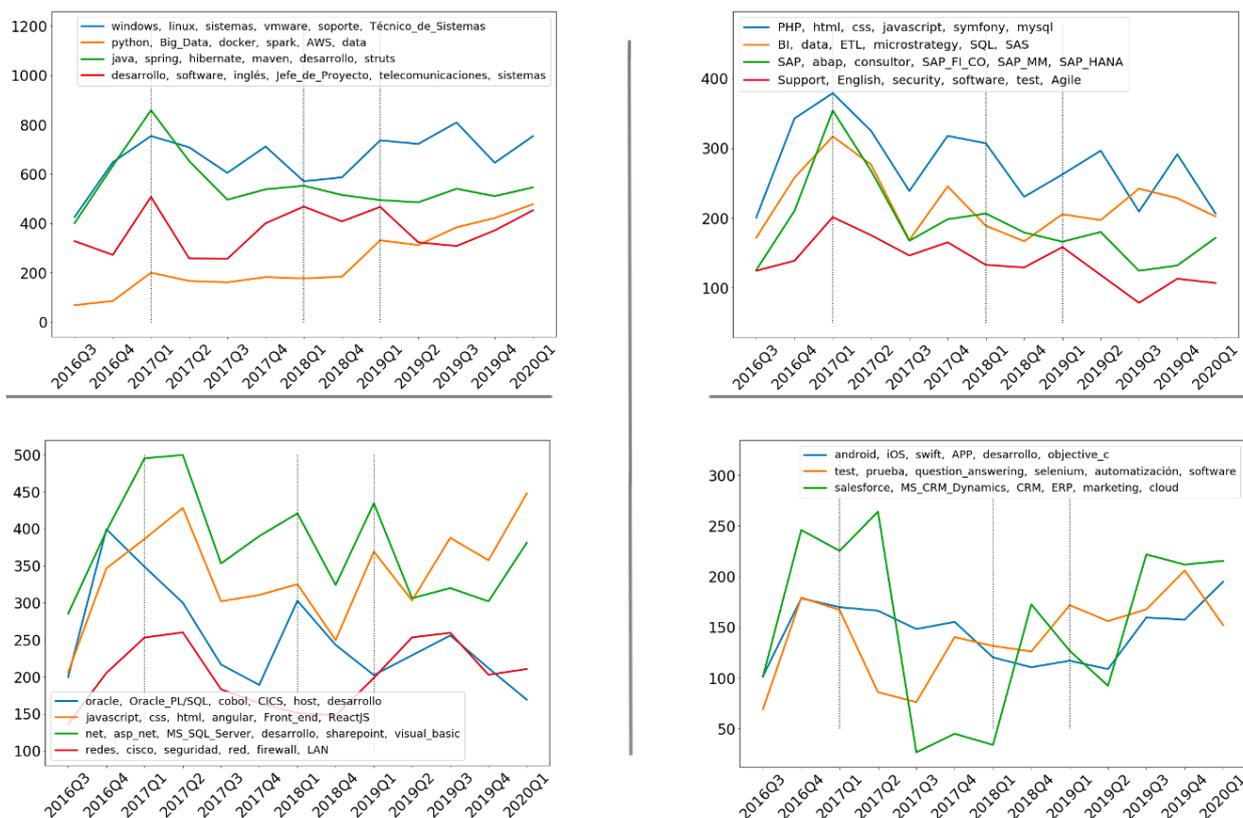
**TABLA 4: PORTAL TECNOEMPLEO.COM. MODELO DINÁMICO DE 15 PERFILES**

Nº perfil	Términos
0	windows, linux, sistemas, vmware, soporte, Técnico_de_Sistemas
1	python, Big_Data, docker, spark, AWS, data, devops, cloud, hadoop, scala, MS_Azure
2	java, spring, hibernate, maven, desarrollo, struts, Analista_Programador
3	desarrollo, software, inglés, Jefe_de_Proyecto, telecomunicaciones, sistemas, Informática
4	oracle, Oracle_PL/SQL, cobol, CICS, host, desarrollo, unix, DBA
5	javascript, css, html, angular, Front_end, ReactJS, node_js, angular_js
6	.net, asp_net, MS_SQL_Server, desarrollo, sharepoint, visual_basic
7	redes, cisco, seguridad, red, firewall, LAN, comunicaciones, ciberseguridad
8	PHP, html, css, javascript, symfony, mysql, wordpress
9	BI, data, ETL, microstrategy, SQL, SAS, powercenter, Qlikview, consultor
10	SAP, abap, consultor, SAP_FI_CO, SAP_MM, SAP_HANA, MM, SAP_SD, SD
11	Support, English, security, software, test, Agile, developer
12	android, iOS, swift, APP, desarrollo, objective_c, mobile, movilidad
13	test, prueba, question_answering, selenium, automatización, software, jira, scrum
14	salesforce, MS_CRM_Dynamics, CRM, ERP, marketing, cloud, helpdesk, navision, consultor, comercial

En la tabla anterior no se han incluido de manera intencionada los tamaños globales asignados a cada tópico dinámico. Esto es así, porque utilizar una cifra única esconde el propósito real de este tipo de modelos, que es precisamente mostrar la evolución temporal del tamaño de cada tópico. Para ello, el GRÁFICO 5 muestra una

estimación del número total de ofertas asignadas a cada perfil en cada trimestre del rango temporal analizado.

### GRÁFICO 5: EVOLUCIÓN TEMPORAL. NÚMERO DE OFERTAS PARA LOS PERFILES OBTENIDOS POR EL MODELADO DINÁMICO EN EL PORTAL TECNOEMPLEO



Las figuras muestran pocas evidencias de tendencias claras en los perfiles obtenidos, un resultado que está en línea con lo observado en IaD2014 e IaD2015. Para el primer trimestre de 2017 se observa un aumento en el número de ofertas asociadas a varios perfiles, pero esto puede deberse principalmente al elevado número de ofertas que se publicaron en dicha ventana temporal. Prestando atención fundamentalmente a los perfiles de mayor tamaño, las únicas conclusiones relevantes podrían ser:

#### ANÁLISIS TEMPORAL

Se observa un **crecimiento significativo del número de ofertas asociadas al análisis de datos y tecnologías de Inteligencia Artificial** y un ligero retroceso de otros perfiles más generalistas de programadores

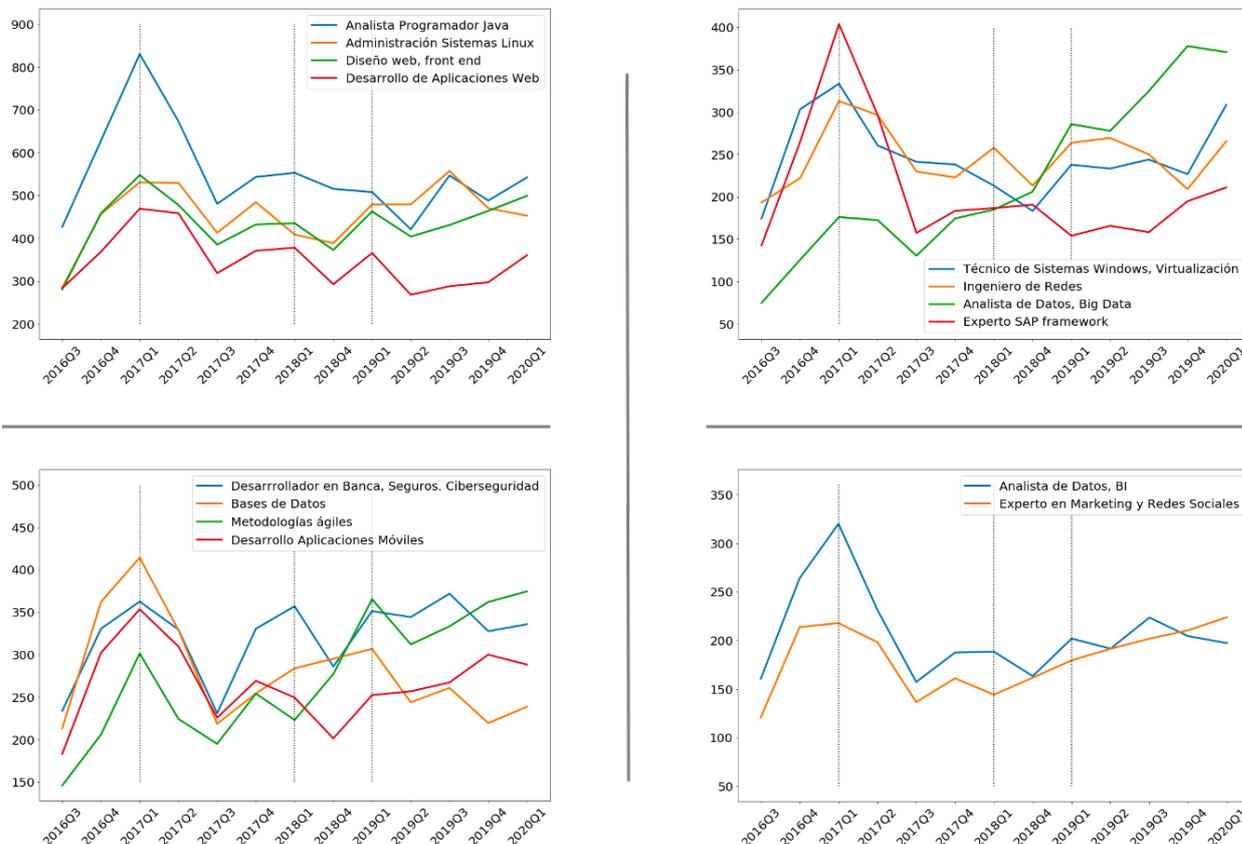
- Una tendencia al alza en el número de ofertas asociadas a un perfil de Programador – Analista de Datos (perfil “python, Big\_Data, docker, spark, AWS, data, devops, cloud, hadoop, scala, MS\_Azure ...”, con un claro crecimiento desde 2019).
- Otro perfil que tiende a incrementar su presencia, fundamentalmente en los últimos trimestres de 2019 es el caracterizado por los términos “windows, linux, sistemas, vmware, soporte, Técnico\_de\_Sistemas”.
- Un descenso moderado del número de ofertas en perfiles de programación (perfiles “java, spring, hibernate, maven” y “PHP, html, css, javascript, symfony, mysql, wordpress”, entre otros).

Una pregunta pertinente es si realmente resulta necesario disponer de un modelo dinámico para alcanzar dichas conclusiones, o si por el contrario se podría analizar la evolución temporal a partir de un modelo estático, midiendo sobre dicho modelo estático el tamaño de cada tópico en cada instante temporal. Para dar respuesta a esta pregunta, hemos repetido el análisis temporal, particularizando el modelo estático de 14 perfiles que fue descrito anteriormente, marginalizando la importancia de los tópicos para cada ventana temporal. A modo de ejemplo, para obtener las importancias relativas de los perfiles del modelo en el trimestre 2019Q1, se analiza la distribución de los tópicos empleando únicamente las ofertas de empleo asociadas a dicho intervalo temporal.

Podemos comprobar que los resultados obtenidos siguiendo esta aproximación (GRÁFICO 6) son bastante análogos a los encontrados en el caso del modelo dinámico, sobre todo en cuanto a la validez de las conclusiones alcanzadas. Este hecho, junto con la mayor conveniencia que aporta disponer de un modelo único, sugiere que podría ser aconsejable centrar el análisis en un modelo estático o de variación temporal muy lenta, sobre el que se puedan incluso inferir la distribución de tópicos de nuevas ofertas de empleo.

Una ventaja adicional de trabajar a partir de un modelo único de perfiles es que el usuario no tiene que interpretar múltiples perfiles, lo cual puede resultar incluso confuso dado el carácter no determinista de los modelos de tópicos y el carácter no exclusivo de los mismos. De esta manera, estaríamos sacrificando precisión de los modelos en aras de una mayor sencillez para el usuario.

## GRÁFICO 6: NÚMERO DE OFERTAS PARA LOS PERFILES OBTENIDOS POR EL MODELADO DE TÓPICOS EN EL PORTAL DE TECNOEMPLO. RESULTADOS OBTENIDOS A PARTIR DEL MODELO ESTÁTICO DE TÓPICOS



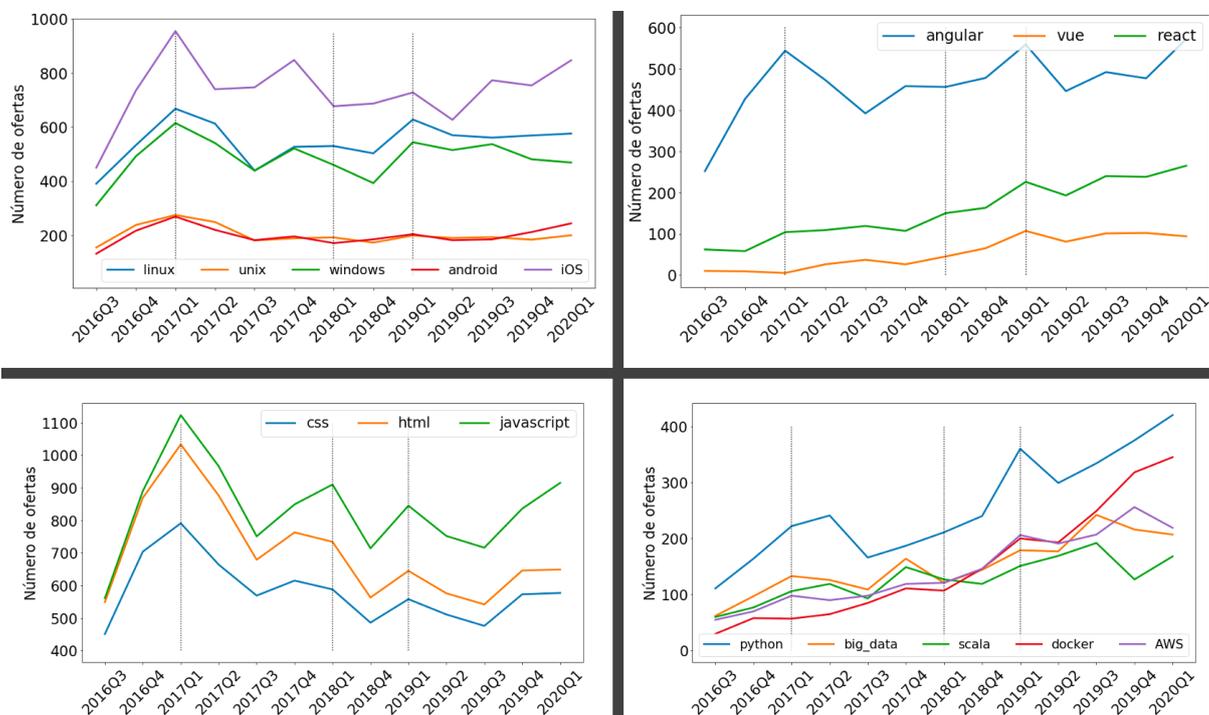
### Análisis temporal en base a la popularidad de tecnologías

La discusión previa sugiere que el análisis de tópicos puede utilizarse como herramienta para análisis comparativos. Sin embargo, si deseamos obtener una cuantificación precisa del tamaño de los diferentes perfiles, las medidas proporcionadas por un modelo de tópicos se verán afectadas por diversos parámetros, concretamente, por el número de tópicos y por los hiperparámetros del algoritmo de extracción de tópicos. Además, los resultados mostrados para la evolución del tamaño de tópicos en función del tiempo muestran una variación significativa de los tamaños de los tópicos, incluso al observar trimestres consecutivos. Cabe preguntarse si dichas diferencias están realmente presentes en los datos, o son consecuencia más bien del algoritmo de aprendizaje empleado.

Con el objetivo de responder estas preguntas, hemos realizado una revisión de la popularidad de distintas tecnologías concretas en la base de datos de tecnoempleo.com. La siguiente figura muestra el

número de ofertas que contienen los términos indicados en la leyenda, y permite observar su evolución a lo largo del tiempo.

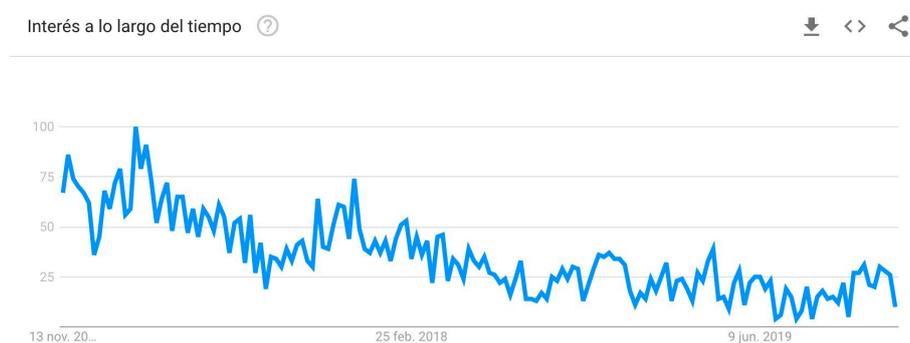
### GRÁFICO 7: NÚMERO DE OFERTAS EN LOS QUE APARECEN TECNOLOGÍAS CONCRETAS DE INTERÉS PARA LOS PERFILES DETECTADOS



Podemos observar cómo la tendencia positiva del perfil de empleo “Analista de Datos” se confirma al observar el crecimiento en el número de ofertas que incorporan términos como “python”, “big data” o “AWS”. El número de ofertas asociadas a sistemas operativos permanece estable en el tiempo. Las tecnologías asociadas al diseño de páginas web (“html”, “css”) decrecen su presencia en las ofertas publicadas. Por último, “angular” es la tecnología más extendida para el diseño de aplicaciones móviles, si bien su demanda se muestra estable frente a un crecimiento significativo y sostenido de otras tecnologías alternativas como son “vue.js” y “react”.

A modo de confirmación, se ha incluido una captura de Google Trends, que muestra una evolución decreciente de interés en las búsquedas por el término “css”, en consonancia con la observada sobre las ofertas de empleo analizadas.

## GRÁFICO 8: ESTIMACIÓN DEL INTERÉS POR EL TÉRMINO "CSS" EN BÚSQUEDAS DE INTERNET (FUENTE: GOOGLE TRENDS)



A partir de las observaciones realizadas, podemos concluir que:

- El análisis de la mención de tecnologías en ofertas de empleo proporciona una alternativa al uso directo de perfiles, y permite una cuantificación precisa del número de ofertas asociadas a una tecnología particular.
- Las tendencias observadas en la demanda de tecnologías concretas parecen más estables (menos ruidosas) que las obtenidas al analizar los perfiles de empleo.

A pesar de las ventajas anteriores, no es viable realizar un análisis de perfiles basado exclusivamente en la medición de términos tecnológicos, ya que muchas ofertas están vinculadas a varios perfiles (tecnologías no excluyentes), y a que cada perfil estará caracterizado por diversas tecnologías, no siendo la correspondencia entre perfil y tecnologías inmediata.

Una aproximación que podría resultar interesante para estimar de manera más precisa el tamaño de los perfiles, y que se propone como línea de continuación, consistiría en estimar en primer lugar el conjunto de tecnologías asociadas a un perfil concreto, bien partiendo del análisis de tópicos o partir de un agrupamiento de términos empleando Word embeddings, para a continuación estimar el tamaño de cada perfil bajo la condición de que una oferta que incluya cualquier término de un perfil dado puede asociarse a dicho perfil.

### 1.4.3 Perfilado jerárquico en portales de empleo basado en CNOs

Presentamos en esta subsección los resultados de la categorización de ofertas de trabajo según la taxonomía de la Clasificación Nacional de Ocupaciones (CNO) y según la categoría europeas de competencias y habilidades (ESCO). El método de clasificación no requiere disponer de etiquetas para el entrenamiento, sino que

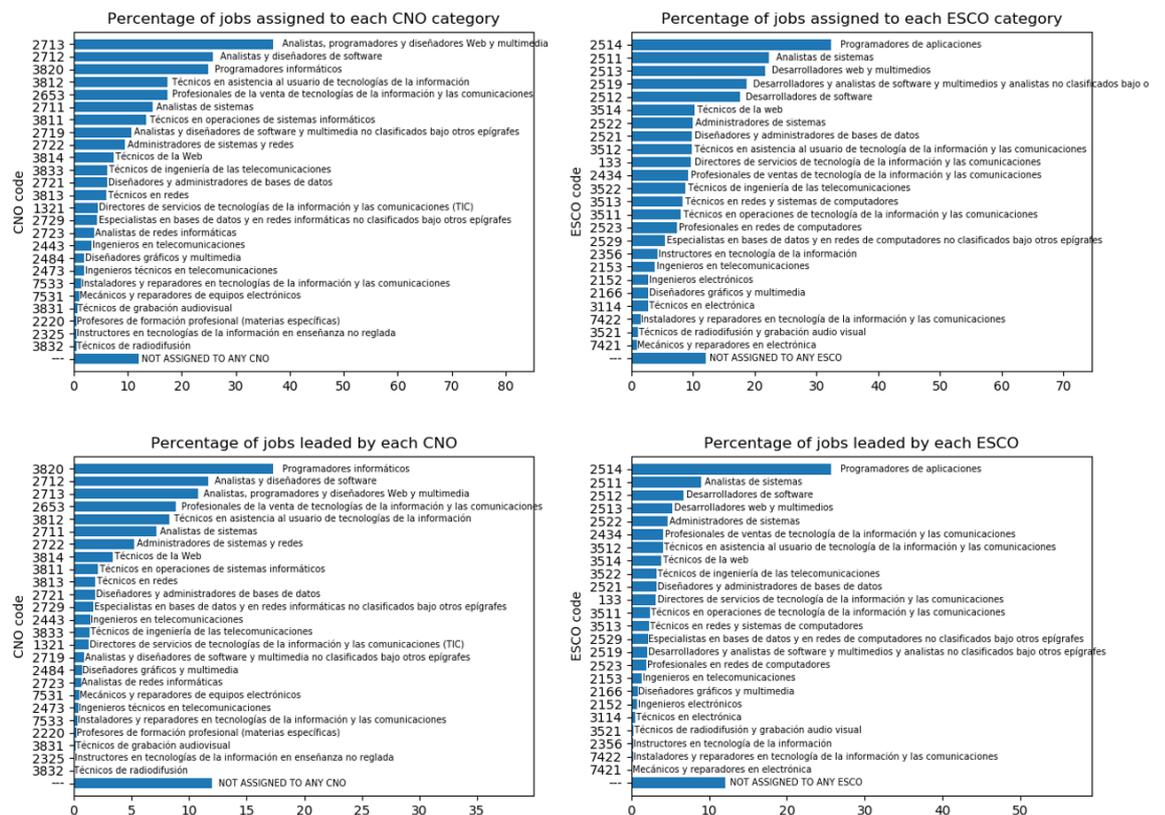
utiliza como criterio las propias definiciones de las categorías objetivo. Además, el método permite asignar varias etiquetas a cada oferta de empleo, hasta un máximo de tres etiquetas.

Una vez clasificadas las ofertas de empleo según CNO o ESCO, esta información se ha incorporado a los metadatos disponibles para las mismas, generando a partir del conjunto de información disponible un panel basado en inteligencia de negocio (BI) que permite analizar los resultados, y cruzar las categorías inferidas con el resto de información disponible para las ofertas, incluyendo los perfiles extraídos mediante el modelado de tópicos.

### Clasificación de las ofertas de tecnoempleo.com según CNO y ESCO

En primer lugar, mostramos la distribución de muestras asignadas a cada código. La Figura siguiente muestra el porcentaje de ofertas según código, tanto al considerar todas las categorías asignadas a cada oferta (hasta un máximo de tres), como al considerar únicamente la categoría más importante para cada oferta. Nótese que el número de ofertas por categoría es variable, pudiendo asignarse desde ninguna hasta 3 categorías, si bien en el caso de asignación múltiple las categorías tienen diferente importancia.

### GRÁFICO 9: CLASIFICACIÓN DE OFERTAS SEGÚN CNO Y ESCO. PORCENTAJE DE OFERTAS EN CADA CÓDIGO, Y PORCENTAJE DE OFERTAS QUE LIDERA CADA CÓDIGO

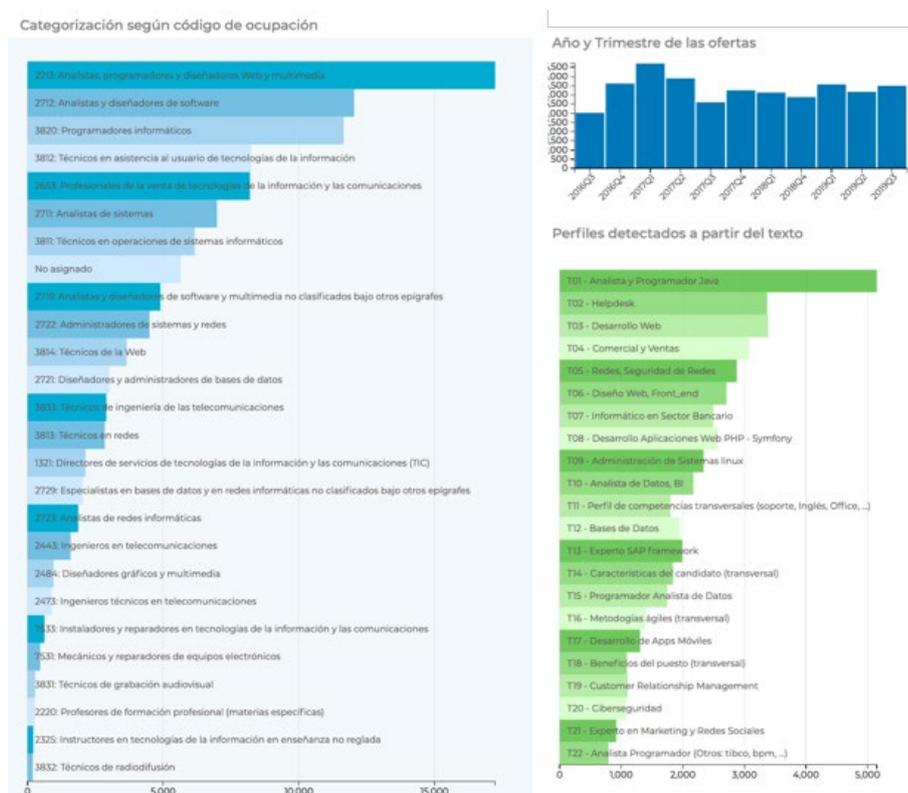


## Visualización con la herramienta de inteligencia de negocio

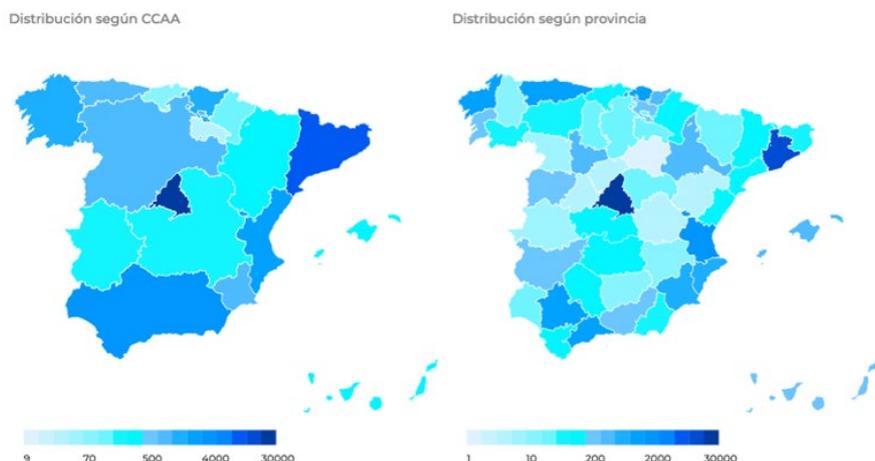
La herramienta proporcionada permite filtrar las ofertas de trabajo de acuerdo a determinados valores de los metadatos asociados, o de acuerdo a una combinación de dichos valores, estimando la distribución de las ofertas seleccionadas según CNO o ESCO. Actualmente, el panel de control implementado dispone de los siguientes paneles activables:

- Categorías (CNO o ESCO)
- Tópicos extraídos de forma automática (se pueden seleccionar varios modelos)
- Año y trimestres: permite restringir el análisis a un trimestre o conjunto de trimestres dados.
- Distribución según provincia
- Distribución según comunidad autónoma
- Experiencia mínima requerida
- Distribución salarial
- Formación mínima exigida
- Tipo de contrato
- Tiempo que la oferta permaneció publicada en el portal
- "Corrección de género": consiste en un criterio binario, que identifica si la oferta hace mención expresa al hecho de que tantos hombres como mujeres son elegibles para el puesto ofertado.

### GRÁFICO 10: MANDOS DE CONTROL PARA LA SELECCIÓN DE CNO (ESCO), PERFILES, Y VENTANA TEMPORAL



## GRÁFICO 11: MANDOS DE CONTROL PARA LA SELECCIÓN DE COMUNIDAD AUTÓNOMA Y PROVINCIA



## GRÁFICO 12: MANDOS DE CONTROL PARA LA SELECCIÓN DE OTROS METADATOS PUBLICADOS POR TECNOEMPLEO.COM, Y PARA LA SELECCIÓN SEGÚN CRITERIO DE "CORRECCIÓN DE GÉNERO"



A modo de ejemplo, mostraremos en las siguientes tablas el número de ofertas asignadas a cada CNO y ESCO (nótese que la suma de ofertas en cada CNO o ESCO excede el total de ofertas disponibles en cada período debido a la posibilidad de asignación múltiple), y su distribución según los perfiles del modelo de competencias de tecnoempleo.com. Para cada CNO o ESCO se muestran únicamente los cuatro perfiles más relevantes.

Nótese, que de igual manera que se ha extraído la información acerca de los perfiles más relevantes para el primer trimestre de 2019, la herramienta permite extraer otras distribuciones (e.g., según provincia), e incluso realizar selecciones múltiples en los distintos paneles.

**TABLA 5. PERFILES IDENTIFICADOS EN LAS OFERTAS ASIGNADAS A CADA CNO. PRIMER TRIMESTRE DE 2019**

CNO	Título	Nº de ofertas	Perfiles	Nº de ofertas
2713	Analistas, programadores y diseñadores Web y multimedia	1538	java, spring, hibernate, oracle, maven, struts	336.4
			net, android, iOS, ASP, MS_SQL_Server, MVC	276.9
			javascript, html, css, angular, Front_end, angular_js	202.7
			PHP, html, javascript, symfony, css, mysql	160.1
3820	Programadores informáticos	1039	java, spring, hibernate, oracle, maven, struts	225.5
			net, android, iOS, ASP, MS_SQL_Server, MVC	168.7
			javascript, html, css, angular, Front_end, angular_js	125.7
			java, programación, francés, software, linux, desarrollo	117.3
2712	Analistas y diseñadores de software	1018	java, spring, hibernate, oracle, maven, struts	194.0
			net, android, iOS, ASP, MS_SQL_Server, MVC	190.3
			java, programación, francés, software, linux, desarrollo	132.6
			javascript, html, css, angular, Front_end, angular_js	87.0
2653	Profesionales de la venta de tecnologías de la información y las comunicaciones	761	java, spring, hibernate, oracle, maven, struts	105.9
			java, programación, francés, software, linux, desarrollo	91.3
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	80.8
			net, android, iOS, ASP, MS_SQL_Server, MVC	79.6
3812	Técnicos en asistencia al usuario de tecnologías de la información	723	windows, helpdesk, vmware, active_directory, Microsoft, hardware	243.3
			linux, unix, shell_script, redhat, windows, oracle	76.3
			java, programación, francés, software, linux, desarrollo	70.7
			cisco, redes, seguridad, firewall, LAN, red	62.2
3811	Técnicos en operaciones de sistemas informáticos	586	java, spring, hibernate, oracle, maven, struts	86.3
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	84.6
			PHP, html, javascript, symfony, css, mysql	53.4
			net, android, iOS, ASP, MS_SQL_Server, MVC	52.6
2711	Analistas de sistemas	554	oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	63.2
			SAP, abap, consultor, SAP_FI_CO, MM, SAP_HANA	55.5
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	54.9
			java, programación, francés, software, linux, desarrollo	53.3
2722	Administradores de sistemas y redes	410	linux, unix, shell_script, redhat, windows, oracle	123.3
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	109.0
			cisco, redes, seguridad, firewall, LAN, red	64.0
			python, Big_Data, AWS, devops, docker, MS_Azure	25.9
2719	Analistas y diseñadores de software y multimedia no clasificados bajo otros epígrafes	387	java, spring, hibernate, oracle, maven, struts	67.8
			java, programación, francés, software, linux, desarrollo	63.0
			test, Support, question_answering, English, Agile, scrum	59.2
			net, android, iOS, ASP, MS_SQL_Server, MVC	56.3

CNO	Título	Nº de ofertas	Perfiles	Nº de ofertas
3814	Técnicos de la Web	346	javascript, html, css, angular, Front_end, angular_js	52.2
			net, android, iOS, ASP, MS_SQL_Server, MVC	49.1
			linux, unix, shell_script, redhat, windows, oracle	39.8
			PHP, html, javascript, symfony, css, mysql	38.6
2721	Diseñadores y administradores de bases de datos	289	oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	67.4
			linux, unix, shell_script, redhat, windows, oracle	53.8
			python, Big_Data, AWS, devops, docker, MS_Azure	35.0
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	34.7
3813	Técnicos en redes	242	windows, helpdesk, vmware, active_directory, Microsoft, hardware	78.0
			cisco, redes, seguridad, firewall, LAN, red	64.8
			linux, unix, shell_script, redhat, windows, oracle	47.3
			java, programación, francés, software, linux, desarrollo	9.9
3833	Técnicos de ingeniería de las telecomunicaciones	237	windows, helpdesk, vmware, active_directory, Microsoft, hardware	61.5
			linux, unix, shell_script, redhat, windows, oracle	29.7
			java, programación, francés, software, linux, desarrollo	27.9
			cisco, redes, seguridad, firewall, LAN, red	24.6
1321	Directores de servicios de tecnologías de la información y las comunicaciones (TIC)	223	javascript, html, css, angular, Front_end, angular_js	31.9
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	30.5
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	29.6
			java, programación, francés, software, linux, desarrollo	25.2
2729	Especialistas en bases de datos y en redes informáticas no clasificados bajo otros epígrafes	208	cisco, redes, seguridad, firewall, LAN, red	59.3
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	36.2
			python, Big_Data, AWS, devops, docker, MS_Azure	33.7
			linux, unix, shell_script, redhat, windows, oracle	21.2
2723	Analistas de redes informáticas	184	linux, unix, shell_script, redhat, windows, oracle	35.7
			cisco, redes, seguridad, firewall, LAN, red	28.5
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	27.4
			java, programación, francés, software, linux, desarrollo	16.7
2443	Ingenieros en telecomunicaciones	149	java, programación, francés, software, linux, desarrollo	36.8
			cisco, redes, seguridad, firewall, LAN, red	31.0
			linux, unix, shell_script, redhat, windows, oracle	14.1
			python, Big_Data, AWS, devops, docker, MS_Azure	11.5
2484	Diseñadores gráficos y multimedia	99	SEO, marketing, wordpress, html, photoshop, UX	32.2
			java, programación, francés, software, linux, desarrollo	14.8
			javascript, html, css, angular, Front_end, angular_js	9.0
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	7.7
2473		80	java, programación, francés, software, linux, desarrollo	19.2

CNO	Título	Nº de ofertas	Perfiles	Nº de ofertas
	Ingenieros técnicos en telecomunicaciones		cisco, redes, seguridad, firewall, LAN, red	15.7
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	9.1
			test, Support, question_answering, English, Agile, scrum	8.8
7531	Mecánicos y reparadores de equipos electrónicos	45	java, programación, francés, software, linux, desarrollo	19.3
			test, Support, question_answering, English, Agile, scrum	8.7
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	4.2
			net, android, iOS, ASP, MS_SQL_Server, MVC	2.7
7533	Instaladores y reparadores en tecnologías de la información y las comunicaciones	43	windows, helpdesk, vmware, active_directory, Microsoft, hardware	11.0
			python, Big_Data, AWS, devops, docker, MS_Azure	10.3
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	6.9
			cisco, redes, seguridad, firewall, LAN, red	5.2
2220	Profesores de formación profesional (materias específicas)	24	windows, helpdesk, vmware, active_directory, Microsoft, hardware	5.2
			java, programación, francés, software, linux, desarrollo	4.5
			net, android, iOS, ASP, MS_SQL_Server, MVC	3.2
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	2.8
3831	Técnicos de grabación audiovisual	22	windows, helpdesk, vmware, active_directory, Microsoft, hardware	4.9
			python, Big_Data, AWS, devops, docker, MS_Azure	3.4
			java, spring, hibernate, oracle, maven, struts	2.7
			linux, unix, shell_script, redhat, windows, oracle	2.2
2325	Instructores en tecnologías de la información en enseñanza no reglada	21	javascript, html, css, angular, Front_end, angular_js	4.4
			python, Big_Data, AWS, devops, docker, MS_Azure	3.5
			java, programación, francés, software, linux, desarrollo	3.1
			java, spring, hibernate, oracle, maven, struts	2.9
3832	Técnicos de radiodifusión	18	windows, helpdesk, vmware, active_directory, Microsoft, hardware	5.8
			linux, unix, shell_script, redhat, windows, oracle	5.5
			cisco, redes, seguridad, firewall, LAN, red	2.3
			java, programación, francés, software, linux, desarrollo	1.6
	No asignado	591	test, Support, question_answering, English, Agile, scrum	102.6
			python, Big_Data, AWS, devops, docker, MS_Azure	81.6
			javascript, html, css, angular, Front_end, angular_js	55.0
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	54.4

**TABLA 6. PERFILES IDENTIFICADOS EN LAS OFERTAS ASIGNADAS A CADA ESCO. PRIMER TRIMESTRE DE 2019**

ESCO	Título	Nº de ofertas	Perfiles	Nº de ofertas
2514	Programadores de aplicaciones	1311	java, spring, hibernate, oracle, maven, struts	320.5
			net, android, iOS, ASP, MS_SQL_Server, MVC	224.8
			javascript, html, css, angular, Front_end, angular_js	150.5
			PHP, html, javascript, symfony, css, mysql	128.3
2513	Desarrolladores web y multimedia	913	java, spring, hibernate, oracle, maven, struts	185.0
			net, android, iOS, ASP, MS_SQL_Server, MVC	166.8
			javascript, html, css, angular, Front_end, angular_js	144.3
			PHP, html, javascript, symfony, css, mysql	131.0
2512	Desarrolladores de software	872	net, android, iOS, ASP, MS_SQL_Server, MVC	149.4
			java, programación, francés, software, linux, desarrollo	142.0
			javascript, html, css, angular, Front_end, angular_js	117.8
			java, spring, hibernate, oracle, maven, struts	113.4
2511	Analistas de sistemas	860	java, spring, hibernate, oracle, maven, struts	121.9
			net, android, iOS, ASP, MS_SQL_Server, MVC	99.3
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	90.7
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	76.8
2519	Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes	730	java, spring, hibernate, oracle, maven, struts	137.2
			net, android, iOS, ASP, MS_SQL_Server, MVC	127.7
			java, programación, francés, software, linux, desarrollo	98.5
			test, Support, question_answering, English, Agile, scrum	76.4
2521	Diseñadores y administradores de bases de datos	477	linux, unix, shell_script, redhat, windows, oracle	82.9
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	79.5
			python, Big_Data, AWS, devops, docker, MS_Azure	77.4
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	48.8
3512	Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones	456	windows, helpdesk, vmware, active_directory, Microsoft, hardware	178.3
			java, programación, francés, software, linux, desarrollo	47.3
			linux, unix, shell_script, redhat, windows, oracle	38.6
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	33.4
133		446	java, programación, francés, software, linux, desarrollo	54.8

ESCO	Título	Nº de ofertas	Perfiles	Nº de ofertas
	Directores de servicios de tecnología de la información y las comunicaciones		java, spring, hibernate, oracle, maven, struts	54.4
			net, android, iOS, ASP, MS_SQL_Server, MVC	45.6
			javascript, html, css, angular, Front_end, angular_js	45.5
2522	Administradores de sistemas	441	linux, unix, shell_script, redhat, windows, oracle	135.3
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	108.0
			cisco, redes, seguridad, firewall, LAN, red	41.6
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	39.8
3514	Técnicos de la web	437	javascript, html, css, angular, Front_end, angular_js	80.9
			net, android, iOS, ASP, MS_SQL_Server, MVC	67.0
			java, spring, hibernate, oracle, maven, struts	65.4
			PHP, html, javascript, symfony, css, mysql	63.4
3511	Técnicos en operaciones de tecnología de la información y las comunicaciones	401	windows, helpdesk, vmware, active_directory, Microsoft, hardware	67.6
			java, spring, hibernate, oracle, maven, struts	44.9
			linux, unix, shell_script, redhat, windows, oracle	38.1
			net, android, iOS, ASP, MS_SQL_Server, MVC	35.4
2434	Profesionales de ventas de tecnología de la información y las comunicaciones	392	MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	48.6
			java, programación, francés, software, linux, desarrollo	48.5
			java, spring, hibernate, oracle, maven, struts	46.5
			net, android, iOS, ASP, MS_SQL_Server, MVC	38.3
3522	Técnicos de ingeniería de las telecomunicaciones	354	windows, helpdesk, vmware, active_directory, Microsoft, hardware	87.5
			cisco, redes, seguridad, firewall, LAN, red	53.6
			java, programación, francés, software, linux, desarrollo	41.6
			linux, unix, shell_script, redhat, windows, oracle	36.1
2523	Profesionales en redes de computadores	342	cisco, redes, seguridad, firewall, LAN, red	68.7
			linux, unix, shell_script, redhat, windows, oracle	57.2
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	41.0
			python, Big_Data, AWS, devops, docker, MS_Azure	30.8
3513	Técnicos en redes y sistemas de computadores	331	windows, helpdesk, vmware, active_directory, Microsoft, hardware	125.2
			linux, unix, shell_script, redhat, windows, oracle	69.8
			cisco, redes, seguridad, firewall, LAN, red	69.2

ESCO	Título	Nº de ofertas	Perfiles	Nº de ofertas
			java, programación, francés, software, linux, desarrollo	12.8
2529	Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes	279	python, Big_Data, AWS, devops, docker, MS_Azure	61.3
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	57.2
			cisco, redes, seguridad, firewall, LAN, red	53.2
			linux, unix, shell_script, redhat, windows, oracle	21.4
2356	Instructores en tecnología de la información	172	java, spring, hibernate, oracle, maven, struts	41.4
			net, android, iOS, ASP, MS_SQL_Server, MVC	28.5
			javascript, html, css, angular, Front_end, angular_js	22.6
			java, programación, francés, software, linux, desarrollo	17.1
2153	Ingenieros en telecomunicaciones	161	java, programación, francés, software, linux, desarrollo	47.6
			cisco, redes, seguridad, firewall, LAN, red	33.1
			java, spring, hibernate, oracle, maven, struts	13.3
			linux, unix, shell_script, redhat, windows, oracle	12.9
2152	Ingenieros electrónicos	125	java, programación, francés, software, linux, desarrollo	39.6
			cisco, redes, seguridad, firewall, LAN, red	15.4
			test, Support, question_answering, English, Agile, scrum	14.4
			java, spring, hibernate, oracle, maven, struts	10.7
3114	Técnicos en electrónica	118	java, programación, francés, software, linux, desarrollo	34.4
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	23.9
			cisco, redes, seguridad, firewall, LAN, red	10.7
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	10.0
2166	Diseñadores gráficos y multimedia	112	SEO, marketing, wordpress, html, photoshop, UX	26.7
			java, programación, francés, software, linux, desarrollo	17.2
			net, android, iOS, ASP, MS_SQL_Server, MVC	14.8
			javascript, html, css, angular, Front_end, angular_js	13.7
7422	Instaladores y reparadores en tecnología de la información y las comunicaciones	60	python, Big_Data, AWS, devops, docker, MS_Azure	16.4
			cisco, redes, seguridad, firewall, LAN, red	11.7
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	8.5
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	6.1
3521	Técnicos de radiodifusión y	48	net, android, iOS, ASP, MS_SQL_Server, MVC	13.1
			python, Big_Data, AWS, devops, docker, MS_Azure	6.2

ESCO	Título	Nº de ofertas	Perfiles	Nº de ofertas
	grabación audio visual		linux, unix, shell_script, redhat, windows, oracle	5.3
			java, programación, francés, software, linux, desarrollo	4.4
7421	Mecánicos y reparadores en electrónica	31	java, programación, francés, software, linux, desarrollo	14.4
			windows, helpdesk, vmware, active_directory, Microsoft, hardware	3.7
			test, Support, question_answering, English, Agile, scrum	2.7
			linux, unix, shell_script, redhat, windows, oracle	2.4
n.a.	No asignado	523	test, Support, question_answering, English, Agile, scrum	85.4
			MS_CRM_Dynamics, ERP, CRM, salesforce, Excel, consultor	67.3
			python, Big_Data, AWS, devops, docker, MS_Azure	57.1
			oracle, BI, Oracle_PL/SQL, ETL, microstrategy, MS_SQL_Server	43.6

## 1.5 Apéndice técnico 1: Metodología ML.

### 1.5.1 Proceso de clasificación de ofertas de empleo según CNOs.

La clasificación de ofertas de empleo según CNO ha seguido una metodología diferente a la de los otros sistemas de clasificación y detección utilizados en este proyecto. La razón principal se encuentra en la disponibilidad, para este problema, de una información para el aprendizaje supervisado de naturaleza diferente: en lugar de un conjunto de etiquetas que indiquen la categoría o categorías CNO a las que pertenece cada oferta, se dispone de un documento descriptivo, en castellano, de cada CNO, con 5 campos por cada CNO:

1. Código
2. Título
3. Descripción
4. Ejemplos (ejemplos de ocupaciones incluidos en el CNO)
5. Excluidos (ejemplos de ocupaciones que no deben incluirse en el CNO).

El clasificador automático, por tanto, no emplea un algoritmo de aprendizaje supervisado, sino técnicas de procesamiento en lenguaje natural. El proceso consta de dos fases principales:

1. Caracterización de CNOs
2. Clasificación de ofertas.

## Caracterización de CNOs.

El documento descriptivo de cada CNO se utiliza para obtener una representación en forma de bolsas de palabras. Para ello, se identifican todos los términos presentes en el documento descriptivo de cada CNO. De forma general, un término es una palabra, de la que se han eliminado variaciones de género, número o tiempo verbal, aplicando un algoritmo de *stemming*, pero con algunas excepciones:

- Los siguientes n-gramas se tratan como términos:
  - "formación profesional"
  - "base de datos"
  - "red social"
  - "administra-dor/-ción de redes"
  - "administra-dor/-ción de sistemas"
  - "sistema operativo"
  - sistema de comunicaciones
  - sistema de transmisión
  - sistema informático

Esto permite, por ejemplo, desambiguar el uso de palabras como "sistema", que adquieren un significado especialmente relevante para describir un CNO cuando se agrupan con las palabras que la acompañan.

- Las palabras "desarrollador" y "desarrollo" no se tratan como un único término ("desarroll-") sino como dos términos diferentes, dado que en general se emplean con significados diferenciados en los CNOs.

Una vez identificados los términos de cada CNO, se calcula una puntuación o score asociada a cada término y cada CNO, que es el resultado de sumar los puntos de cada aparición del término en cada uno de los campos del CNO, del modo siguiente:

- Aparición en el campo "Título": 3 puntos
- Aparición en el campo "Descripción": 1 punto
- Aparición en el campo "Examples": 2 puntos.

De este modo, el algoritmo otorga mayor puntuación (y, por tanto, mayor relevancia) a los términos que aparecen en el Título del CNO, y menor a los términos que aparecen en la descripción.

El resultado de este proceso es una representación del CNO como una lista de términos, con una puntuación asociada a cada uno de ellos. A partir de esta representación se calcula la representación TF-IDF que también se ha empleado en el análisis de textos en otros módulos del proyecto.

Obsérvese que, hasta este punto, no se ha utilizado la información del campo "Excluded". Con los términos que aparecen en este campo se construye una representación "negativa" del CNO, como sigue:

1. Extracción de términos del campo "Excluded"
2. Eliminación de todos los términos que no son exclusivos de este campo, es decir, que también están presentes en alguno de los otros campos del CNO.
3. Puntuación de cada término, de acuerdo con el número de repeticiones, y sumando dos puntos por cada aparición.
4. Cálculo de una representación TF-IDF

Como resultado de todo este proceso, se han obtenido dos representaciones vectoriales de cada CNO. La diferencia entre ambas será la que se utilice para clasificar cada CNO.

### **Clasificación.**

Para clasificar una oferta de empleo, se sigue un procedimiento similar en muchos aspectos al seguido para los CNOs, y que se resume en lo siguiente:

1. Identificación de términos.
  - a. Se ignora el término "formación" cuando no está referido a "formación profesional" (dado que, en general, el término se utiliza en las ofertas para explicar la formación ofrecida al candidato, y no para especificar la búsqueda de un candidato especialista en formación).
  - b. Se ignora el término "telef-" cuando está asociado a "agentes telefónicos", "encuestadores telefónicos" o "entrevistadores telefónicos", que no constituyen empleos para técnicos especialistas en telefonía.
  - c. Se ignora el término "programa" cuando va asociado a programas de formación / trabajo, etc, para evitar la confusión con el término "programa informático" que sí es característico de algunos CNOs.
  - d. El término "márketing" (muy frecuente en las ofertas de empleo pero inexistente en el vocabulario de los CNOs) se identifica con el término "venta" (que es la expresión utilizada en los documentos descriptivos de CNOs, y no es tan frecuente en las ofertas de empleo)
2. Eliminación de todos los términos que no aparecen en el vocabulario de los CNOs.
3. Puntuación de cada término, dependiendo del número de repeticiones del mismo en la oferta, y del lugar de aparición de los mismos, de acuerdo con el siguiente criterio:

- a. Aparición en el campo "title": 6 puntos
  - b. Aparición en el campo "description": 1 punto.
  - c. Aparición en el campo "skills": 1 punto.
4. Cálculo de la transformación TF-IDF.

Como resultado de este proceso, cada oferta de empleo queda caracterizada por un vector  $\mathbf{v}$  de la misma dimensión que los vectores que caracterizan a cada CNO. En este punto, para cada CNO se calcula una puntuación final, que es el producto escalar del vector  $\mathbf{v}$  y el vector asociado al CNO, normalizado por la raíz cuadrada del tamaño de la oferta de empleo (para eliminar el efecto de las diferencias de extensión en las ofertas de empleo). La oferta de empleo se asigna a los CNOs que obtienen la máxima puntuación, hasta un máximo de tres, y siempre que dicha puntuación sea superior a un umbral mínimo.



# 2

## ANÁLISIS DE PROGRAMAS FORMATIVOS TICC



VICEPRESIDENCIA  
TERCERA DEL GOBIERNO  
MINISTERIO  
DE ASUNTOS ECONÓMICOS  
Y TRANSFORMACIÓN DIGITAL

ontsi observatorio  
nacional de las  
telecomunicaciones  
y de la SI  
red.es

## 2 Análisis de programas formativos TICC

### 2.1 Propósito del estudio

En este capítulo mostraremos los resultados de la caracterización de la oferta formativa a partir de la información capturada en sitios web oficiales relativos a títulos universitarios y de formación profesional.

#### **FUENTE DE DATOS**

**TÍTULOS UNIVERSITARIOS** del Registro de Universidades, Centros y Títulos (RUCT) y **CUALIFICACIONES PROFESIONALES** del Instituto Nacional de las Cualificaciones (INCUAL).

### 2.2 Fuentes de datos.

#### 2.2.1 Fuentes para el análisis de títulos universitarios

Para el análisis de los títulos universitarios se hará uso de los planes de estudio disponibles en el Registro de Universidades, Centros y Títulos (RUCT)<sup>4</sup> donde se encuentran disponibles la mayoría<sup>5</sup> de los planes de estudios impartidos en las universidades españolas.

Para este análisis se han considerado tanto los títulos de Grado como los de Máster vinculados a las ramas de conocimiento: Artes y Humanidades, Ciencias, Ingeniería y Arquitectura y Ciencias Sociales y Jurídicas. Para limitar el estudio a los planes formativos dentro del sector TICC, dentro de cada rama, se han considerado como titulaciones TICC aquellas relacionadas con los siguientes ámbitos de estudio:

1. **Artes y Humanidades:** estudios en Bellas Artes, Diseño, Producción Artística, Edición, Dibujo, Radio, Industria Cinematográfica y Traducción Multimedia.
2. **Ciencias:** Matemáticas, Estadística, Física y Computación
3. **Ingeniería y Arquitectura:** estudios de Informática, software, web, Telecomunicación, Sistemas y Tecnologías de Información, Sonido e Imagen, Electrónica, Diseño Gráfico, Multimedia.
4. **Ciencias Sociales y Jurídicas:** estudios en Periodismo, Comunicación Audiovisual, Publicidad, Información y Documentación, Bibliotecas y Servicios de Información Digital, Consultoría y Gestión de la Información, Cine y Televisión, Gestión de la Información Digital, Tecnologías de la Información, Diseño, Creatividad para la Comunicación Publicitaria, Cine, Televisión y Medios Interactivos.

<sup>4</sup> RUCT, <https://www.educacion.gob.es/ruct/home>

<sup>5</sup> Se indica mayoría porque, tal y como se explicará más adelante, para algunos grados o másteres su plan de estudios no está disponible.

**FUENTE DE DATOS  
(UNIVERSIDAD)**

**585**

**TÍTULOS  
UNIVERSITARIOS  
TICC de Grado y  
Máster cuyo plan  
de estudios se  
encuentra  
disponible.**

Cabe destacar que hay una quinta rama, asociada a Ciencias de la Salud, que directamente no se ha considerado por no incluir ninguna titulación asociada al ámbito de las TICC.

Algunos de los planes formativos objeto de estudio, a pesar de estar incluidos en el registro oficial de titulaciones, no tienen su plan de estudio disponible, lo que ha obligado a reducir el presente estudio a aquellas titulaciones que han permitido el acceso a su plan de estudios a través del portal de RUCT.

Tras aplicar el proceso de crawling o rastreo el 11 de julio de 2019, se ha podido acceder al plan de estudios de **976** títulos TICC, entre grados y másteres de las diferentes ramas de conocimiento. En la Tabla 7 se detalla la distribución de estos planes de estudio entre las diferentes ramas, indicando el número total de títulos por rama y nivel (grado (G) o máster (M)), cuántos de ellos corresponden a TICC y en cuántos de ellos la información del plan de estudios se encuentra disponible (585 en total).

**TABLA 7: DISTRIBUCIÓN DEL NÚMERO DE TÍTULOS UNIVERSITARIOS (GRADOS (G) O MÁSTERES (M)) OFERTADOS EN LAS DIFERENTES RAMAS DEL CONOCIMIENTO. SE INCLUYE EL NÚMERO DE TÍTULOS PERTENECIENTES A TICC Y EL NÚMERO DE TÍTULOS CUYO A CUYO PLAN DE ESTUDIOS SE PUEDE ACCEDER.**

Rama	Nivel	Títulos TICC	Plan de estudios
Ingeniería y Arquitectura	G	291	263
	M	319	233
Ciencias	G	26	23
	M	38	23
Arte y Humanidades	G	60	49
	M	65	42
Ciencias Sociales y Jurídicas	G	184	153
	M	257	190

Para facilitar el posterior análisis de cada plan de estudios, su descarga se ha acompañado de información sobre el nombre del título, universidad, rama de conocimiento, nivel y enlace al plan de estudios.

## 2.2.2 Fuentes para el análisis de cualificaciones profesionales

### FUENTE DE DATOS (FORMACIÓN PROFESIONAL)

# 72

### CUALIFICACIONES PROFESIONALES.

El portal web del Instituto Nacional de las Cualificaciones<sup>6</sup> (INCUAL) incluye un total de 26 familias profesionales de las que se han seleccionado (siguiendo, principalmente, las recomendaciones del informe del ONTSI<sup>7</sup>) las siguientes tres como relacionadas con el sector de contenidos digitales:

1. Artes Gráficas
2. Imagen y Sonido
3. Informática y Comunicaciones

Dentro de cada una de estas familias, el portal web de INCUAL proporciona un listado de las cualificaciones profesionales de cada familia junto con un documento pdf que incluye la información completa de cada cualificación. En particular, contiene:

- La competencia general.
- Las unidades de competencia que la forman. Además, por cada unidad, detalla sus realizaciones y el contexto profesional al que se adecuan.
- El entorno profesional en el que se enmarca la cualificación.
- La formación asociada, indicando los módulos formativos que la componen. Además, por cada módulo formativo se detallan las capacidades y criterios de evaluación, contenidos y las capacidades a adquirir por los alumnos para ser desarrolladas (o completadas) en un entorno profesional.

Dado que esta información se encuentra muy bien estructurada para cada cualificación profesional, y disponible para todas ellas, la exploración o crawling de esta web se ha centrado en descargar el contenido completo de estos pdf. Aunque cabe destacar, que para facilitar el estudio posterior, cada uno de estos documentos se ha acompañado del nombre de la cualificación, familia profesional en la que se incluye y nivel de cualificación que tiene asignado.

Tras lanzar el proceso de búsqueda el 11 de julio de 2019, se han descargado **72** cualificaciones profesionales, distribuidas en:

- 31 cualificaciones de la rama de Artes Gráficas,
- 18 cualificaciones de Imagen y Sonido y
- 23 cualificaciones de Informática y Comunicaciones.

<sup>6</sup> INCUAL, <http://incual.mecd.es/>

<sup>7</sup> "Estudio sobre los Perfiles Profesionales y las Titulaciones Relacionadas con el Sector TICC" realizado por el ONTSI en febrero de 2014.

## 2.3 Metodología

La metodología de trabajo sigue los mismos pasos que la descrita en el Apartado 1.2 para el procesado de las ofertas de empleo, por lo que se van a aplicar con los siguientes cuatro pasos:

1. Descarga de los planes formativos, extrayendo en cada caso la información del plan de estudios más los metadatos descriptores del mismo (título, rama, ...).
2. Procesado de los planes formativos para transformarlos en un conjunto de términos que caractericen cada plan de estudios. Del mismo modo que para las ofertas de empleo, se aplicarán los siguientes pasos:
  1. Identificación de términos.
  2. Eliminación de palabras comunes (stopwords).
  3. Identificación de n-gramas.
  4. Identificación de plurales.
  5. Edición manual de stopwords. Inclusión de stopwords específicas para cada corpus.
  6. Generación de la bolsa de palabras de cada plan formativo; es decir, por cada plan formativo se contará cuántas veces aparecen cada término.
3. Identificación de perfiles. Utilizando la bolsa de palabras que caracteriza cada plan formativo, se buscarán los términos que co-ocurren con alta frecuencia en un número significativo de documentos. Este procesado proporciona un listado de perfiles, el tamaño de cada perfil en el corpus y la representación de cada documento o plan formativo sobre estos perfiles. Véase el Subapartado 1.2.3 para una explicación más detallada de estas salidas.
4. Visualización de resultados. Se utilizará la misma herramienta de visualización que en el análisis de ofertas de empleo, ya que permite el análisis detallado de las salidas de los diferentes modelos de perfilado.

## 2.4 Resultados del análisis

En esta sección se incluye una descripción de los perfiles obtenidos para los planes formativos universitarios (grados y másteres), cualificaciones profesionales y cursos de formación. Para ello, se realiza un análisis con 15 perfiles, ya que para todos los casos proporciona un conjunto de perfiles con suficiente valor semántico y evita la aparición de perfiles irrelevantes y ruidosos.

Para realizar este análisis, para cada perfil, se incluye:

- Un identificador del perfil.

- Su importancia, medida como el porcentaje de documentos o planes formativos que dicho perfil representa en el corpus.
- Las palabras o términos característicos.

A lo largo de las siguientes subsecciones, se mostrarán estos resultados para cada tipo de nivel educativo.

### 2.4.1 Análisis de títulos universitarios

La Tabla 8 resume el contenido de los perfiles extraídos sobre la colección de planes de estudios universitarios. De los 15 perfiles originales, se han eliminado, utilizando la herramienta de perfiles incluida en el proyecto, 3 perfiles considerados ruidosos por considerar que incluían únicamente términos de escaso valor para el corpus analizado.

**TABLA 8: CARACTERIZACIÓN DE LOS PERFILES DE TÍTULOS UNIVERSITARIOS DETECTADOS (12 PERFILES).**

Nº perfil	Importancia	Términos
0	27,7%	Becas, Horarios, Convocatorias plaza, beca, MES, Informática, mención, optatividad, convocatoria, cuatrimestre, semestre
1	8,2%	Informática, BBDD, Sistemas Operativos Informática, software, computador, redes, seguridad, ingeniería, computación, sistemas_operativos, informático, arquitectura
2	7,7%	Audiovisual y Periodismo audiovisual, periodístico, género, artístico, periodismo, cine, contemporáneo, producto, televisión, comunicativo
3	7,6%	Electrónica y Comunicaciones electrónica, comunicaciones, redes, industrial, ingeniería, circuito, telecomunicación, señal, electrónico, instrumentación
4	6,7%	Modalidad y tipo de docencia (transversal) presencial, clase, tutoría, autónomo, hora, exposición, teórico, individual, discusión, debate
5	6,6%	Publicidad y Marketing marketing, publicitario, publicidad, estrategia, campaña, producto, sociedad, estratégico, negocio, marca

<b>6</b>	5%	Intercambio, Movilidad (transversal) destino, enviar, plaza, beca, firmado, videojuego, modificación, contrato, trámite, documento
<b>7</b>	4,8%	Modalidad de docencia, Herramientas docentes (transversal) virtual, máster, solicitud, presencial, TIC, alegación, PS, educación, aula, foro
<b>8</b>	2,9%	Modalidad de docencia, Evaluación (transversal) presencial, semipresencial, distancia, modalidad, tutoría, videoconferencia, convocatoria, presencialidad, examen, extraordinario
<b>9</b>	2,4%	Matemáticas ecuación, matemático, reducido, ordenador, laboratorio, diferencial, geometría, álgebra, lineal, variable, tutoría
<b>10</b>	1,5%	Telecomunicaciones, Redes, Señales ordenación, comunicaciones, modificar, electrónica, señal, circuito, telecomunicaciones, redes, vicerrectorado, participativo
<b>11</b>	1%	Herramientas docentes (transversal) virtual, didáctico, unidad, telefónico, foro, buzón, tutorial, supuesto, tutoría, BBDD

Analizando en detalle los términos que forman cada perfil, nos encontramos con 2 tipos de perfiles que podríamos asociar a las siguientes categorías o temáticas:

- Perfiles temáticos: asociados a las diferentes ramas incluidas en el conjunto de datos: Informática, Audiovisual y Periodismo, Electrónica y Comunicaciones, Publicidad y Marketing, Matemáticas y Telecomunicaciones.
- Perfiles transversales: que incluyen otros conceptos habituales en las descripciones de los planes de estudio, asociados asimismo a diferentes aspectos concretos, como pueden ser la movilidad, evaluación, herramientas docentes, etc.

#### 2.4.2 Análisis de cualificaciones profesionales

Los resultados del modelo de perfiles sobre la colección de cualificaciones profesionales se encuentran en la Tabla 9. En este caso se han considerado adecuados todos los perfiles obtenidos, por

lo que no ha sido necesario utilizar la herramienta de edición de perfiles.

**TABLA 9: CARACTERIZACIÓN DE LOS PERFILES DETECTADOS PARA LAS CUALIFICACIONES PROFESIONALES (15 PERFILES)**

Nº perfil	Importancia	Términos
0	11,4%	Impresión, Flexografía impresión, primo, tinta, cilindro, medición, impresor, huecograbado, flujo, entintado, preimpresión
1	10%	Técnico de redes redes, red, comunicaciones, local, interconexión, fijo, alarma, móvil, monitorización, inalámbrico
2	9,9%	Programación, diseño web lenguaje, BBDD, usuario, programación, web, desarrollado, consulta, gestor, relacional, orientado
3	8,9%	Producción audiovisual, cine y televisión audiovisual, cinematográfico, postproducción, televisión, cámara, guión, narrativo, captación, televisivo, vídeo
4	7,8%	Administración de servidores usuario, servidor, explotación, instalar, servidores, mensajería, administrar, hardware, copia, monitorización
5	6,9%	Encuadernación encuadernación, tapa, manipulado, plegado, hilo, lomo, apilado, cosido, corte, duro
6	6,3%	Edición y maquetación editorial, multimedia, texto, interactivo, animación, obra, maquetación, autor, libro, prototipo
7	6%	Impresión, encuadernación preimpresión, transformado, industrial, impresión, encuadernación, departamento, ensayo, flujo, tinta, proveedor
8	5,8%	Fotografía fotográfico, procesado, digitalización, revelar, vectorial, película, químico, fotolito, luz, encargo

<b>9</b>	5,6%	Ingeniero de sonido sonido, musical, sonoro, sesión, señal, animación, sala, mezcla, vivo, micrófono
<b>10</b>	4,7%	Impresión, estampación matriz, estampación, estampa, boceto, grabado, pantalla, litográfico, tinta, serigráfico, fotolito
<b>11</b>	4,6%	Microinformática microinformático, periférico, conexión, consumible, red, antivirus, sustitución, instalar, texto, avería
<b>12</b>	4,3%	Empaquetado y envasado envase, embalaje, cartón, artículo, ondulado, papelería, complejo, papel, fabricación, estructural
<b>13</b>	4,2%	Técnico audiovisual, espectáculos en directo espectáculo, vivo, evento, artístico, ensayo, representación, iluminación, escénico, colectivo, escena
<b>14</b>	3,5%	Diseñador gráfico ilustración, artístico, libro, estudio, formal, estético, económico, conceptual, obra, plástico

En este caso, los perfiles se pueden desglosar en tres grupos, cada uno de ellos asociado a cada una de las ramas de las cualificaciones profesionales:

1. **Artes Gráficas**, a la pertenecen los perfiles 0, 5, 6, 7, 8, 10, 12 y 14, que si los agrupásemos, representarían el 49% del corpus.
2. **Imagen y Sonido**, donde nos encontraríamos a los perfiles 3, 8, 9 y 13, representando un 24,5% del corpus
3. **Informática y Comunicaciones**, formado por los perfiles 1, 2, 4 y 11 y conteniendo el 22,3% del corpus.

A diferencia de los planes de estudios universitarios, en este caso no nos hemos encontrado perfiles transversales, sino que los planes formativos son muy específicos de las diferentes ramas a las que pertenecen. Por otro lado, cabe destacar que el hecho de que los perfiles de artes gráficas representen aproximadamente la mitad del corpus se debe a que por un lado el número de cualificaciones en esta rama es 31, frente a las 18 y 23 de las otras ramas, y que la longitud media de estos planes formativos es mayor, lo que hace que domine su presencia frente a los perfiles de otros ámbitos.

Por otro lado, podemos concluir que los planes de estudio entre ramas son bastante independientes, ya que no se encuentran perfiles comunes a varias ramas.

Es interesante resaltar que los perfiles obtenidos en este caso están mucho más centrados en contenidos que en el caso de las titulaciones universitarias: se ha podido extraer un número mayor de perfiles temáticos a pesar de disponer de menos de la décima parte de titulaciones. El principal motivo para ello es que, a diferencia de lo que ocurre con las memorias de verificación de las titulaciones universitarias, los documentos pdf disponibles en el caso de cualificaciones profesionales están centrados en los contenidos de los programas. Para solventar esta dificultad, se propone como línea de extensión la obtención de textos de mayor calidad para la extracción de los perfiles de las titulaciones universitarias. Para ello, se podría trabajar a partir de la memoria verificada extrayendo únicamente las secciones relevantes. Alternativamente, se propone implementar un nuevo crawler de datos que explore la estructura del portal RUCT para extraer, en los casos en que esta información se encuentra desglosada y estructurada en el propio portal, los programas de estudio y contenidos de las asignaturas que componen los planes de estudio.