

## Metodología para el análisis de demanda laboral mediante datos de Internet: el caso colombiano

Recibido: 14 de julio de 2015- Aceptado: 19 de octubre de 2015

Doi:

Jeisson Arley Cárdenas Rubio<sup>†</sup>  
Juan Carlos Guataquí Roa<sup>‡</sup>  
Jaime Mauricio Montaña Doncel<sup>§</sup>

---

### Resumen

La coyuntura laboral reciente del desempleo en Colombia muestra que su componente cíclico se ha reducido a su mínima expresión y que, por ende, la mayor parte del desempleo actual hace referencia a las tipologías friccional y estructural del desempleo, y puede ser explicado en gran parte por el “*mismatch*” existente entre la demanda y la oferta de trabajo. La carencia de información sobre las necesidades y requerimientos del sector productivo ahondan y mantiene, aún más, las brechas existentes entre ambos componentes del mercado laboral. Dada esta situación, este trabajo propone una solución a la falta de información sobre los requerimientos cuantitativos y cualitativos de la demanda de trabajo, desde una metodología de *big data*, es decir, la aglomeración y sistematización de grandes cantidades de información, teniendo como fuente de información la utilizada por las bolsas de empleo colombianas en sus actividades de colocación. Procesada esta información, y luego comparada con la información de oferta laboral disponible en encuestas de hogares, puede concluirse que una

---

<sup>†</sup> Facultad de Economía, Universidad del Rosario. Correo electrónico: cardenasrjeisson@urosario.edu.co.

<sup>‡</sup> Facultad de Economía, Universidad del Rosario. Correo electrónico: juan.guataqui@urosario.edu.co

<sup>§</sup> Departamento de Economía, Universidad de Turín. Correo electrónico: jaimem.montana@gmail.com

---

Para citar este artículo: Cárdenas, R. J. A., Guataquí, R. J. C., & Montaña, D. J. M. (2015). Metodología para el análisis de demanda laboral mediante datos de internet: el caso colombiano. *Revista de Economía del Rosario*, 18(1), 93-126. doi:

base de vacantes construida a partir de *big data* en Colombia, es una excelente herramienta para conocer la demanda que, por distintas habilidades, tiene el sector productivo y, por ende, la formulación de políticas públicas integrales de educación y formación, sumamente requeridas en el país. Este artículo documenta el esfuerzo pionero desarrollado al respecto.

*Palabras clave:* vacantes, demanda de trabajo, competencias, big data.

*Clasificación JEL:* J63

## A methodological approach for labor demand analysis based on internet data: The Colombian case.

### Abstract

The recent performance of unemployment in Colombia suggests an almost complete reduction of cyclical unemployment, and the likely prevalence of its frictional and structural components. This is probably due to a qualitative mismatch between labor supply and demand. From this last side of the labor market, there has been a clear and persistent lack of accurate information about its requirements in terms of capabilities and skills. We propose a big data methodological approach based on help-wanted advertising. Our data base, once compiled and compared with employment data from the Colombian household survey proves to be a more accurate and detailed source of information about occupations, skills and vacancies than any source available to date in the Colombian Statistical System. On this paper we detail the context and methodological guidelines of this pioneer case study.

*Keywords:* vacancies, labor demand, Skills, big data.

*JEL Classification:* J63

## Metodologia para a análise de procura laboral utilizando dados da internet: O caso Colombiano

### Resumo

A conjuntura laboral recente do desemprego na Colômbia mostra que sua componente cíclica se tem reduzido a sua mínima expressão e que, portanto, a maior parte do desemprego atual faz referência às tipologias friccional e estrutural do desemprego, e pode ser explicado em grande parte pelo “mismatch” existente entre a procura e a oferta de trabalho. A carência de informação sobre as necessidades e requerimentos do setor produtivo aprofundam e mantêm ainda mais, as brechas existentes entre ambos os componentes do mercado laboral. Dada esta situação, este trabalho propõe uma solução à falta de informação sobre a tipologia do desemprego, desde uma metodologia de big data, é dizer, a aglomeração e sistematização de grandes quantidades de informação, tendo como fonte de informação a utilizada pelas bolsas de emprego colombiana em suas atividades de colocação. Processada esta informação, e logo comparada com a informação de oferta laboral disponível em enquetes de lares, pode se concluir que uma base de vacantes construída a partir de big data na Colômbia, é uma excelente ferramenta para conhecer a procura que por distintas habilidades, tem o setor produtivo, e, portanto, a formulação de políticas públicas integrais

de educação e formação, sumamente requeridas no país. Este artigo documenta o esforço pioneiro desenvolvido ao respeito.

*Palavras-chave:*

*Classificação JEL:* J63

## 1. Una visión desagregada del mercado de trabajo

Anteriormente, se concebía al mercado de trabajo como un único ente en el cual no había distinciones entre sectores, ocupaciones o habilidades de los trabajadores. No existía una visión *completa* del mercado laboral que contemplara la complejidad de sus dimensiones.

El modelo de referencia, con el que se explican los niveles de ocupación y desempleo, se puede encontrar teóricamente en las contribuciones de Mortensen-Pissarides (1994, 2001). Basados en este tipo de concepción, se ha venido recolectando información cuantitativa y cualitativa, con el ánimo de dar respuesta a los diferentes problemas de política y realizar ejercicios que permitan entender la realidad del mercado laboral de Colombia.

Entre los esfuerzos aunados se encuentran numerosos levantamientos de perfiles ocupacionales sectoriales, entre los cuales cabe resaltar los realizados por entidades gubernamentales como el Ministerio del Trabajo en conjunto con la red de observatorios del mercado laboral - ORMET<sup>1</sup> -, las cámaras de comercio, y PNUD<sup>2</sup> para diferentes zonas del país. Además de estos esfuerzos el Departamento Administrativo Nacional de Estadística (DANE), y en particular con la información de la Gran Encuesta Integrada de Hogares (GEIH) se han realizado contribuciones para entender la oferta laboral del país.

Con el fin de poder abordar el análisis, utilizando lo que hemos denominado como visión *completa* del mercado de trabajo, en el país se hizo uso de dos enfoques para caracterizar la demanda de trabajo: el primero, puede encontrarse por ejemplo en el levantamiento de perfiles ocupacionales, se utilizó la concepción de que cada mercado era independiente entre sí y estaba completamente caracterizado (industria, localización geográfica, sector), y que existían ocupaciones específicas para cada mercado, así como los saberes, competencias y habilidades eran específicos al mismo.

---

<sup>1</sup> Para ver un ejemplo consultar [http://www.mintrabajo.gov.co/component/docman/doc\\_download/1451-perfiles-ocupacionales-sector-construccion.html](http://www.mintrabajo.gov.co/component/docman/doc_download/1451-perfiles-ocupacionales-sector-construccion.html)

<sup>2</sup> Ídem.

Sin embargo de acuerdo a esta práctica de análisis, se han hecho supuestos que son limitantes, tales como que cada mercado analizado es independiente y autónomo<sup>3</sup>. Por otra parte, se agregaban y omitían estas características, utilizando la información estadística proporcionada por el DANE. En este enfoque simplemente se omitía la existencia de estos mercados o se controlaba económicamente.

Utilizando este tipo de datos en Colombia, que si bien no eran adecuados y suficientes (Guataquí et al., 2014), se intentó dar una caracterización del mercado laboral. Hay que reconocer, sin embargo, que si bien los esfuerzos han traído consigo muchos frutos para entender las dinámicas y desafíos propios del mercado colombiano, la falta de datos de caracterización de la demanda ha hecho que la investigación se vea estancada desde hace ya más de una década, como se resaltaba desde Vivas et al. (1998).

La transición teórica de mercados atómicos, totalmente separados, o mercados en donde existe una integración completa y sin diferencia, ha venido perdiendo campo. Un ejemplo de esto se puede encontrar en modelos teóricos y empíricos en los cuales los sectores u ocupaciones son interdependientes entre sí (Stops y Mazzoni, 2010; Stops 2014), o cuando se introducen heterogeneidades en la composición de la oferta y demanda de trabajo (Barnichon y Figura, 2010).

Dentro de la praxis e implementación de política pública se ha dado por hecho desde hace tiempo que esta inter-relación es existente. Un ejemplo de esto es la política de balance de competencias, en donde se consideran habilidades y competencias transversales a las ocupaciones, y por medio del rentrenamiento se puede aumentar la reinserción laboral disminuir la duración del desempleo.

Es entonces que la generación de nuevas fuentes de recolección y análisis de información de demanda y oferta de trabajo se hacen necesarias. Fuentes que contengan el detalle requerido para la planeación, análisis y evaluación de políticas públicas, así como insumo del debate académico y la generación de conocimiento.

Este documento tiene como objeto la justificación y presentación de una metodología para la recolección y análisis de datos de demanda de trabajo, con la finalidad de contribuir a los aspectos mencionados anteriormente, reabrir las discusiones y estudios de la demanda de trabajo en Colombia y anticipar la proliferación del uso de las tecnologías, como método para la recolección de información.

---

<sup>3</sup> Por lo que no hay relación entre los sectores o las ocupaciones, así como la procedencia de la oferta de trabajo.

## 2. Necesidad de información para realizar un análisis desagregado de la demanda de trabajo

El mercado de trabajo en Colombia se enfrenta a muchos desafíos que se han también identificado internacionalmente en otros mercados. Un ejemplo de esto es que se ha dado por hecho que hay una brecha entre demanda y oferta de trabajo, sin verificación empírica, dado que los datos disponibles no pueden dar explicación de tales hechos. De tal forma, en varias dimensiones del mercado de trabajo se han tomado *de facto* aseveraciones de las que no se puede comprobar la veracidad en Colombia, debido a la falta de datos.

Para ilustrar esta dinámica considérese el siguiente caso: En Colombia varios sectores del sector productivo y en general los empresarios, han venido manifestado de no poder encontrar el capital humano para crecer y llenar sus vacantes. Por otro lado existe un problema de empleabilidad en jóvenes recién egresados de la educación técnica, tecnológica y universitaria. Ambas verdades se han tomado como ciertas incluso sin el respaldo de datos que lo soporten y basado en la experiencia de otros países.

Si bien existen análisis internacionales (Almeida et al., 2012), que ilustran una situación en la cual cerca de la mitad de los empleadores en el mundo no pueden llenar trabajos para recién egresados<sup>4</sup>, y una proporción igual de jóvenes afirman que no se usan las habilidades y competencias en el ejercicio de sus trabajos, la información no se ha verificado en Colombia.

Sin posibilidad de verificar la información, en cuanto los estudios de demanda no tienen el grado de detalle requerido debido principalmente a que la información de demanda de trabajo no cuenta con ésta. Entre la información que se necesita se pueden encontrar: no existe una caracterización de las vacantes por los requerimientos en términos ocupacionales identificados por región, salario, experiencia, entre otras variables que resultan fundamentales para analizar una situación así. Tampoco existe una caracterización en términos de habilidades y competencias para indicar los requerimientos del sector productivo.

En resumen, en Colombia la información de demanda disponible es insuficiente, y no cuenta con las características necesarias para realizar este tipo de análisis (Guataquí et al., 2014). Las razones más relevantes para sostener lo anterior son:

---

<sup>4</sup> Denominados en inglés “*entry-level jobs*”.

- La información de demanda y oferta de trabajo no contiene información relativa a las habilidades y competencias. En el caso de la GEIH (que es una encuesta de oferta) se considera la demanda efectiva, pero no se cuenta con información de cargos no ocupados o vacantes persistentes.
- No se puede analizar los aspectos cíclicos de la demanda de trabajo, pues la información de la GEIH es un *stock*. En cambio, la información de flujo es relevante en cuanto puede dar información sobre la creación y destrucción de empleo por motivos temporales, así como la destrucción de ocupaciones por obsolescencia tecnológica o su creación. En estos últimos casos puede servir de insumo de una clasificación ocupacional adaptada a la demanda de Colombia.
- No existe en Colombia en la actualidad información de demanda que contenga, o en su defecto con la cual se pueda estimar, el número de vacantes por sector y ocupación a nivel desagregado<sup>5</sup>.
- En caso de existir información parcial, la información no está en el mismo lenguaje. Un ejemplo de esto son las diferentes clasificaciones utilizadas para la clasificación ocupacional o la falta de actualización de estas clasificaciones.

De acuerdo a lo anterior se hace necesaria actualizar y crear nuevos métodos para la identificación de la demanda. Sin esta información es imposible realizar ejercicios de pronóstico y prospectiva para la planificación de políticas de capital humano y así dar cumplimiento a los lineamientos o recomendaciones nacionales, tal como lo es el CONPES 3674. La recolección y análisis de información con este grado de detalle traería beneficios para realizar ejercicios académicos que requieran incorporar el número de vacantes o las habilidades requeridas, y comparar los resultados que se dan como dados para saber si aplican o no en el mercado colombiano.

El objetivo del presente documento es aportar el soporte metodológico y en términos de infraestructura, para proveer de manera continua y automatizada información de demanda con el grado de detalle necesario para efectuar análisis y así promover la construcción de metodologías para el correcto diseño e implementación de políticas, así como para promover la investigación en lo que refiere al aspecto de demanda en el mercado laboral colombiano.

---

<sup>5</sup> Existe una nueva serie de tiempo de vacantes en medios impresos (Álvarez y Hofstetter, 2014), sin embargo al analizarla en comparación a la Encuesta de Capital Humano 2014, la representatividad de dicha serie no es suficiente. Así mismo la serie no tiene el detalle a nivel geográfico, sector, ocupación o habilidad y competencia requerida.

### **3. ¿Sobre qué teorías y supuestos se forja el proyecto de levantamiento de información de demanda?**

La falta de información de demanda ha sido un elemento que se ha reconocido desde hace ya más de una década. Sin embargo los esfuerzos para construir una encuesta de demanda no han dado fruto en cuanto realizar este tipo de ejercicios y mantenerlos, es muy costoso. Por otra parte, esta encuesta ha sido planteada como un esfuerzo conjunto de diferentes entidades, y no se ha llevado a cabo debido a un problema de coordinación entre las entidades, hecho entendible pues esto afecta en su presupuesto.

Dados los altos costos y la imposibilidad de un cercano acuerdo entre las entidades competentes para llevar la encuesta de demanda de trabajo a cabo, se optó por investigar y plantear una metodología que fueran viables en términos de costos y esfuerzos. En esta sección se describen los supuestos y teorías sobre los que nos basamos para elegir la metodología de levantamiento de información, y los aspectos que potencialmente queremos que incluya la base, aportando una sucinta explicación para ello.

#### *Disponibilidad y acceso a la información*

La información sobre demanda de trabajo puede ser levantada por dos fuentes: la primera es directa, en la cual por medio de cuestionarios, se aplican con una frecuencia determinada preguntas a las empresas sobre la necesidad y cantidad de recurso humano requerido por el sector productivo, así como del recurso empleado. En Colombia, esta opción se ha descartado, debido al monto de recursos requeridos para solventarla y mantenerla. La segunda es indirecta, en la cual se toman la información que es expuesta al mercado para atraer al recurso humano. Esto se hace por medio del análisis de vacantes en los diferentes medios en donde se publican. Esto se hace con el supuesto de que una vacante funciona como una forma de dar una señal sobre qué tipo de recurso se necesita; así la vacante de trabajo constituye una fuente de información en sí misma.

Las vacantes se pueden encontrar en fuentes privadas, siendo estas sistemas de información a los que solo se puede tener acceso siendo parte de una organización o empresa (las bases de datos de las empresas y universidades) o en fuentes públicas (medios de comunicación, radio, televisión, medios impresos e internet). Existe una serie de vacantes en Colombia que utiliza los medios impresos, limitándose al número de vacantes y no con el grado de detaldeseado (Álvarez y Hofstetter, 2014). Dicha serie viene desagregada a nivel de ciudades pero no presenta ninguna información de ocupaciones, ha-

bilidades, sectores, edad requerida, experiencia requerida u otra información objeto del presente. Además de esto, los medios impresos representan una porción relativamente pequeña y en proceso de obsolescencia, comparada con la información de medios como internet.

Por este motivo se optó por la búsqueda almacenamiento y análisis de las vacantes ofrecidas por medio de internet, en cuanto la morfología de un anuncio de trabajo en internet contaba con la información que se buscaba.

Así mismo esta solución presentaba la oportunidad de en un futuro capturar la mayoría de las ofertas en medios, dada la tendencia en la penetración de internet y de telefonía móvil y uso de datos móviles en Colombia (MINTIC, 2015; SIC, 2014). En este sentido la iniciativa representa una oportunidad para una futura captura de las dinámicas y estado del mercado laboral en Colombia.

A nivel internacional existen experiencias de compilación y análisis de vacantes. Australia, Canadá, Estados Unidos<sup>6</sup> y Holanda tienen series de vacantes sobre las cuales se han analizado diferentes aspectos del mercado laboral<sup>7</sup>. Sin embargo las vacantes en línea, su compilación, reclasificación y uso de todo el texto del anuncio es algo que no se ha hecho anteriormente y que se propone para entender ver la dinámica de las necesidades en términos ocupacionales y como de las habilidades y competencias requeridas en el mercado.

Si bien es una innovación, estas ideas están presentes desde hace tiempo atrás, y se dan hoy gracias a los avances en términos de almacenamiento y versatilidad en los lenguajes de programación que han permitido que la automatización sea más accesible. Un ejemplo de la importancia que este tipo de información tiene se puede encontrar en Dunlop (1996), en el cual se resalta la importancia de la construcción de estructuras ocupacionales detalladas con el fin de entender las dinámicas propias del mercado laboral. En específico se centra en la necesidad de tener información que contenga:

- Información ocupacional que tenga detalles como la compensación, disponibilidad de formación para el trabajo, educación, características respecto a edad, sexo, raza, entre otras.
- Información sobre requerimientos a trabajos para recién egresados.
- Factores que afectan la demanda de las ocupaciones.

---

<sup>6</sup> Para una revisión de la historia de este tipo de levantamiento de información para Estados Unidos ver Clark & Phillips (2002).

<sup>7</sup> Se han hecho esfuerzos desde mediados del siglo XX. En especial La creación del Help Wanted Index y JOLTS. (Abraham and Wachter, 87; Amoah, 2000; Joseph, 1966; Dunlop, 1966)



Esta estructura coincide en gran parte con el nivel de detalle que tiene como objeto el presente documento para recolectar información de mercado laboral. Este proceso dado que hace uso de soporte y herramientas ligadas a la tecnología, y dado el número de información potencial a almacenar, se basó en gran medida en los avances que se han realizado en la última década, especialmente a los avances en Big Data y su metodología de análisis. Este tipo de aproximaciones se han hecho populares en los últimos años y son comunes para el sector productivo, pero su utilización en el gobierno aún resulta limitado (Kambatla et al., 2014).

Otras fuentes que son de carácter público (Servicio de Empleo y SENA), que cuentan con la misma morfología en general, complementaron la información que se obtuvo de internet, con el fin de asegurar la mayor cobertura.

### *Big data en este proyecto*

Como se ha mencionado anteriormente, la información de demanda laboral es escasa en Colombia donde la tasa de desempleo entre enero y septiembre de 2015 es de 9.2% y la tasa de informalidad para el mismo periodo es de 48.3%. Estos resultados pueden ser explicados por los desajustes que se presentan entre lo que se está demandando por parte de las empresas y lo que se está ofreciendo por parte de la fuerza laboral (UKCES, 2012). Por tanto, es de gran importancia conocer las características de la demanda para reducir las asimetrías de información entre la oferta y la demanda laboral.

Sin embargo, hasta ahora no ha existido información sobre demanda insatisfecha (vacantes) y han sido escasas las estimaciones de demanda en país. Una revisión de estudios en la economía colombiana muestra que existe una escasez de análisis de demanda laboral dada la escasez de información (Vivas et al., 1998; Arango y Rojas, 2003; Cárdenas y Bernal, 2001; Guataquí et al., 2014). Consecuentemente, brillan por su ausencia los modelos que expliquen las altas tasas de desempleo, informalidad, subempleo etc. debido a los desequilibrios entre la oferta y demanda laboral. Como resultado, las necesidades de recurso humano en el país y sus implicaciones sobre los resultados del mercado laboral son desconocidas.

En este contexto, se busca dar solución a la ausencia de información con una propuesta innovadora, pues recolectar información de demanda laboral a través de encuestas a las empresas resulta muy costoso en términos de recursos y tiempo. Sin embargo, gran parte de la información de demanda laboral ya existe en internet (Edelman, 2012; Kurekova, 2014) y lo que hace falta es organizarla rigurosa y sistemáticamente para su posterior análisis.

En este documento se propone como alternativa para analizar la demanda laboral del país, utilizar la información de vacantes que está disponible en las principales páginas web. Estas fuentes son útiles para entender los requerimientos de los empleadores como la experiencia laboral, el nivel educativo, el salario y zona geográfica. Por otra parte, contiene una descripción de los conocimientos técnicos y cognoscitivos que son necesarios para desempeñar ciertos cargos, lo cual puede ser utilizado para determinar las necesidades del sector productivo.

Carnevale et al. (2014) señalan las ventajas que tiene utilizar este tipo de información:

- Recolectar información por estos medios es costo-efectivo.
- Produce estimaciones suplementarias de la demanda con el grado de detalle de ocupaciones, industria y geografía.
- Muestra la demanda relativa para diferentes tipos de habilidades a niveles educativos.
- La naturaleza en tiempo real de los avisos clasificados de vacantes permite la detección de tendencias en la demanda laboral. Lo que brinda herramientas para la prospectiva económica.
- La información en tiempo real es útil para alinear los currículos educativos con las necesidades de la fuerza laboral, en un contexto donde existen rápidos cambios industriales.

Por otra parte, Reimsbach-Kounatze (2015) señalan que internet se ha convertido en una fuente potencial de información que complementa los datos recolectados por las oficinas de estadística y además abre la posibilidad de nuevos enfoques de análisis.

De esta manera, se propone el uso de Big Data y de técnicas como el Web Scraping (raspado web) y Text Mining (minería de texto) para compilar y analizar la información sobre vacantes en Colombia. A continuación se hace un recuento de lo que es el Big Data, sus características y sus implicaciones para el análisis del mercado laboral.

De acuerdo a Joyanes (2013), no existe un consenso sobre la decisión del término Big Data, sin embargo, la mayoría de definiciones concuerdan en que el Big Data se caracteriza principalmente por los grandes volúmenes de información, las diferentes fuentes de donde pueden venir los datos y la velocidad con la que se necesita capturar determinada información.

El volumen se refiere a un conjunto de datos que excede las capacidades de las herramientas informáticas comunes en cada sector para compilar y analizar determinada información (Manyika et al., 2011). En este sentido, no se define

un tamaño fijo para clasificar los datos como Big Data<sup>8</sup>, pues esto depende de las tecnologías comunes que se encuentren disponibles tanto en el sector como en determinado momento.

La velocidad con la que fluyen los datos también es un factor relevante para definir la información en el marco del Big Data, por tanto la creación, el flujo constante en tiempo real y su utilización determina la naturaleza de la información (Gartner, 2011; Reimsbach-Kounatze, 2015). Claros ejemplos de lo anterior es la información de redes sociales y las miles de transacciones financieras que se pueden realizar en cuestión de minutos.

La última característica principal de Big Data es la variedad de sus fuentes, las cuales pueden ser de tipo estructurado, semiestructurado y no estructurado, que al ser analizadas en conjunto conducen al reto de utilizar nuevas técnicas. (Joyanes 2013).

Por fuentes estructuradas se entienden, los datos que contienen campos fijos, definidos y ordenados. Este es el caso de la información derivada de encuestas como: la Gran Encuesta Integrada de Hogares-GEIH (Encuesta de oferta laboral) y encuestas sectorizadas de demanda laboral en donde se encuentra la Encuesta Anual de Servicios -EAS, Encuesta Anual Comercio-EAC y la Encuesta Anual Manufacturera-EAM, realizadas por el Departamento Administrativo Nacional de Estadística-DANE. (Guataquí et al., 2014) Los anteriores son datos estructurados, con campos fijos definidos y que contiene información sobre las características de las personas y empresas, con los cuales tradicionalmente se han hecho los estudios de mercado laboral.

Por otra parte, los datos semiestructurados no tienen campos fijos pero tienen etiquetas, encabezados, nodos, entre otros, que permiten identificar los elementos dentro de los datos. Un ejemplo de lo anterior son los lenguajes de programación en XML, JAVA o HTML, en los que se basa la creación de gran parte de las páginas web. En este sentido, los datos utilizados en el caso de la demanda laboral, provienen de fuentes semiestructuradas ya que la fuente primaria de información son los portales web cuya información está en formato HTML.

Por último, se encuentran los datos no estructurados los cuales no tienen ningún tipo de campo definido ni etiquetas que permitan identificar los elementos dentro de los datos. Por ejemplo, los formatos de videos, audio, libros, mensajes de texto, correos electrónicos y textos de anuncios clasificados. En consecuencia, para realizar el análisis de información en este trabajo también

---

<sup>8</sup> Con esta definición se evita el hecho que lo que parecía grandes volúmenes de información en el pasado, hoy en día ya no se considera como lo suficientemente grande.

es necesario el uso de técnicas para manejo de datos no estructurados ya que la información sobre descripción de las vacantes, se compone principalmente de textos (párrafos) donde el empresario describe sus requerimientos para llenar determinada vacante.

Para realizar lo anterior, se hace uso de text mining (minería de texto) la cual consiste en reducir la información proveniente de un texto a través de la identificación de patrones comunes (Joyanes, 2013; Franks, 2012), esta técnica se utiliza frecuentemente para analizar redes sociales, discursos, opiniones en foros y para el caso particular de mercado laboral, los textos de los anuncios de las vacantes. Como se explicará más adelante, con la técnica text mining se leen los datos no estructurados como es el texto del anuncio de la vacante, se procesa y a partir de este texto se crean datos estructurados para su análisis.

#### 4. Recolección de la información

Para la recolección de la información se utilizó la metodología de *web-scraping*. Esta consiste en analizar las propiedades del código bajo el que se muestran las páginas web en los navegadores y recolectar la información necesaria de acuerdo a los parámetros de identificación que se le da a la programación. Los análisis económicos y ejercicios econométricos con este tipo de datos se hacen cada vez más frecuentes. Un ejemplo de esto es el análisis de precios y precios de consumidores realizado por Cavallo (2012).

Con la aplicación de éste tipo de metodologías, se puede aprovechar la estructura estática de la programación de las páginas web y las características del lenguaje HTML, aprovechando su estructura basada en *tags*, para obtener la información<sup>9</sup>. Los *tags* son elementos fijos en la programación de cualquier sitio web con la que se le dan las características de forma, color y contenido.

Con el fin de demarcar las acciones para interpretar por el navegador, cada elemento tiene dentro del lenguaje un comienzo y un fin, y entre ellos el contenido que caracteriza el atributo. Así dentro de cada pareja de *tags* se encuentra la información de ese atributo, siendo este la información de interés. Analizando la morfología de las vacantes online, se establecieron elementos comunes que podían obtenerse de diferentes fuentes, tomando éstos para automatizar una serie de tareas en el sistema operativo que lanzan la búsqueda

---

<sup>9</sup> En Guataquí et Al. (2014) se detalla en manera más precisa el procedimiento de recolección.

de contenido de los elementos identificados. Estas tareas evalúan cada uno de los *tags* y recogen los que se han identificado que tienen información relevante.

Al realizar la descarga de la información se procede a organizarla dentro de una base de datos; estos procesos se programan con una frecuencia regular para asegurar dos descargas mensuales de información. La frecuencia con la que se realizará la descarga a partir del 2016 es diaria, con el fin de tomar posibles vacantes que estén activas en periodos muy cortos, dado que actualmente se viene realizando de manera quincenal.

El algoritmo de recolección contiene además un componente con el cual se valida la captura de información, haciendo énfasis en el tipo de dato recogido, en la longitud y otros parámetros con los que se pueden identificar errores en los datos recolectados. Este proceso se divide en dos partes: en la primera parte se recoge la información en HTML y se extrae el contenido pasándolo a formato XML. Esto permite posteriormente insertarlo en una base de datos.

En la base de datos cada observación corresponde a una única URL, con lo cual en principio se eliminan los posibles duplicados en la base de datos. No obstante existía la preocupación que los datos recolectados tuvieran duplicados por fuentes de información, es decir, que dado el bajo costo de poner una vacante en línea, una misma empresa lo hiciera en varios sitios web, y así nosotros la contáramos repetidas veces. Con la finalidad de eliminar esta posibilidad, se realiza una prueba rutinaria en la cual se eliminan las observaciones duplicadas (o sospechosas de serlo). En este proceso se tomaron las características como la fecha de publicación el título de la vacante, el texto de la descripción, el salario ofrecido, para armar una medida de similitud. Con esta medida se eliminaron las observaciones que sobre pasaban un límite definido.

Dada las características y el volumen de la información, se requirió adaptarse a nuevas técnicas de manejo y almacenamiento de este gran volumen de datos. Con el fin de comparar se realizó un ejercicio en el cual se comparó la capacidad de recursos que se necesitaba instalar utilizando *web-scraping* o si se realizaba manualmente. La planta necesaria al realizar la recolección utilizando este método es el 10% del caso en que se realiza manualmente<sup>10</sup>. El costo en software es también bajo, dado el uso de software de distribución libre, tanto para la recolección, el almacenamiento y el análisis<sup>11</sup>.

---

<sup>10</sup> Tomando como potencial de recolección mensual 108 600 vacantes disponibles, si el promedio de recolección es de 7 a 10 minutos por vacante, se necesitan 806 400 minutos al mes para realizar la tarea. Esto equivale al trabajo de 38 personas.

<sup>11</sup> Entre los software utilizados hay que resaltar la versatilidad que han aportado herramientas como R, Mongo DB, el lenguaje Phyton y Hadoop.

## 5. Homogenización de la información

En el contexto de esta propuesta, se entenderá por vacante al anuncio que coloca el empleador, sin embargo, este anuncio puede tener uno o más puestos de trabajo disponibles, por ejemplo, si un anuncio dice “Se necesitan diez auxiliares contables para trabajar de manera inmediata, salario mínimo”, esta es una vacante con diez puestos de trabajo.

A pesar de que la información, en su mayoría, esta desorganizada no quiere decir que no se pueda unificar y reorganizar, de tal manera que se realicen análisis de la información. Por el contrario, como se tiene tanta información que describe los requerimientos de los empresarios para acceder a un puesto de trabajo, esta se constituye en una fuente de información importante para entender la demanda de trabajo en el país y orientar la oferta laboral para que puedan acceder a un puesto de trabajo

Posteriormente, los datos se descargan en un formato plano para cada una de las fuentes de información utilizadas, en este estudio se utilizaron cuatro fuentes de información, por lo tanto se tienen cuatro archivos planos diferentes con su propia estructura de variables, por ejemplo, una de ellas tiene un campo especialmente destinado a la información del sector de la empresa que solicita la vacante, otra fuente tiene información especialmente para el tipo de horario laboral, otra tiene información para el tipo de contrato que se ofrece.

También existen campos en común en las diferentes fuentes, por un lado, se tiene que todas las fuentes tienen una variable llamada “título de la vacante”, la cual contiene información específica sobre el oficio o el cargo ofrecido por el empleador; otro campo en común se denomina “descripción”, el cual es un campo de texto extenso donde los empresario pueden describir los detalles de vacante, salarios, tipo de empresa, nivel educativo mínimo requerido, horarios, incluso preferencias sobre la edad y el sexo de los aspirantes. Estas dos variables son de gran importancia, ya que con ellas, como más adelante se explicará, se puede identificar la ocupación y su respectivo código, de acuerdo con la clasificación O\*NET y CIUO, lo que permite tener un análisis estadístico estructurado con comparabilidad internacional.

Otros campos disponibles en común son la fecha de publicación de la vacante y su fecha de retiro<sup>12</sup>, con ellas se puede establecer en qué momento exacto estaba disponible la vacante, lo cual es útil para la identificación de estacionalidades en el requerimiento de ciertas vacantes.

---

<sup>12</sup> Solo una de las fuentes no reporta información alguna sobre la fecha de retiro de la vacante.

Con estas similitudes y diferencias se comienza a construir una base homogénea, que permita hacer análisis de información de mercado laboral de las principales bolsas de empleo que se encuentran en el país. El primer paso para hacer esto es unir todas las bases en un solo archivo, cuyas únicas variables en común son el título, la descripción, la fecha de publicación y la fecha de cierre de la vacante, los demás son campos específicos de cada una de las fuentes que contienen información de forma dispersa de variables relevantes como el salario, el nivel educativo y la experiencia laboral.

A partir de este punto, se utilizan paquetes estadísticos con el fin de identificar y extraer expresiones regulares<sup>13</sup>, es decir, con la búsqueda de patrones de palabras o palabras clave en todos los campos disponibles, se comienza a construir cada una de las variables<sup>14</sup>. De esta manera, con la búsqueda de patrones se construyen las variables de idiomas; por ejemplo, si un anuncio de una vacante registró en cualquiera de sus campos las expresiones “inglés”, “francés”, “alemán”, “mandarín”, el buscador de palabras crea una variable denominada “inglés” que toma valor de 1 si el anuncio en algunos de los campos registró que se necesitaba nivel de inglés para acceder a la vacante, y 0 si no se especifica en ninguno de los campos el idioma inglés como requisito para aspirar a una vacante; y así se realiza sucesivamente para cada idioma.

De manera semejante al caso anterior, se conforman las variables de nivel educativo. Si la vacante registró en sus campos un nivel educativo de primaria, se crea una variable *dummy* denominada “primaria” que indica si la vacante requiere personal con nivel de primaria, si registró bachillerato (con todos sus posibles sinónimos), entonces se crea una *dummy* que indica si la vacante requiere trabajadores con bachillerato. De tal forma, se construyen las siguientes variables por cada nivel educativo: “sin educación”, “primaria”, “secundaria”, “bachillerato”, “técnico”, “tecnólogo”, “universitario” “especialización”, “maestría” y “doctorado”. La construcción de estas *dummies*, al no ser excluyente la una de las otras, plantea una ventaja para identificar las vacantes que son indiferentes para los empresarios entre un nivel educativo y otro, por ejemplo, se puede identificar en cuántas y en el cuáles vacantes un nivel técnico es indiferente a tener un nivel de tecnólogo.

Otra variable de gran importancia es la experiencia laboral, con la búsqueda de patrones se determina si una vacante requiere alguna experiencia laboral para poder aspirar al cargo, se construyen la *dummy* “experiencia”

---

<sup>13</sup> Para la búsqueda de patrones regulares en este trabajo se utilizó el programa REGEX

<sup>14</sup> Para el éxito del establecimiento de los patrones de búsqueda adecuados, es necesario realizar una extensa revisión manual para identificar los patrones más pertinentes para identificar lo que se requiere.



donde 1 significa que la vacante requiere experiencia laboral y 0 si no específica experiencia laboral y “sin experiencia” que toma valor de 1 si la vacante no requiere experiencia laboral alguna y 0 si no específica información sobre experiencia. Ahora bien, como el caso de las variables de educación, al construir una *dummy* por cada categoría se pueden identificar las vacantes donde tener experiencia laboral o no es indiferente. De esta manera, se identificaron *dummies* para las variables sobre tipo de contrato, horario laboral, preferencias en la edad y género.

Por otra parte, es importante identificar en qué municipio del país se encuentra disponible la vacante. Dado que esta variable puede tener alrededor de 1123 categorías (en Colombia existen alrededor de 1123 municipios diferentes), crear esta cantidad de *dummies* para cada una sería ineficiente, ya que un anuncio de vacante, como se mostrará más adelante, se ofrece en un solo lugar, por ello se aplica la misma búsqueda de palabras clave o patrones de los nombres de los municipios y se construye una variable única que permite identificar el municipio donde la vacante es ofrecida y se le asigna su respectivo código, de acuerdo con la división político administrativo (divipola).

El salario ofrecido por las empresas es construido a partir de la combinación tanto de patrones textuales como de patrones numéricos, por ejemplo, para identificar si una vacante en el 2014 ofrecía un salario mínimo se tuvo que buscar por SMLV, SMMLV, SMLMV, SLMV, MLV, SLML, SMLE, SMMLG, S.M.L.V., S.M.M.L.V., salario mínimo, 61600, entre otros, para poder identificar la vacante ofrecía un salario mínimo con similares patrones se extrajeron los demás valores para los demás ingresos ofrecidos.

Sin embargo, dado que una de las fuentes ya venía con la variable de ingresos ofrecidos, en categorías fijas, y para tener todos salarios homogenizados para análisis se clasificaron todos los ingresos de acuerdo a estas categorías<sup>15</sup>: menos de \$550000; \$550001 - \$1000000; \$1000001 - \$1500000; \$1500001 - \$2000000; \$2000001 - \$2500000; \$2500001 - \$3000000; \$3000001 - \$3500000; \$3500001 - \$4000000; \$4000001 - \$4500000; \$4500001 - \$5500000; \$5500001 - \$6000000; \$6000001 - \$8000000; y \$8000001 en adelante.

Por último, queda la construcción de la variable ocupación, esta es de vital importancia, pues es la que indica la clase de recurso humano que necesita el país y, en trabajos posteriores, las habilidades y competencias más relevantes en el mercado laboral, información de importancia para la gestión y orientación del recurso humano. De cierta manera, clasificar esta información, ya sea

---

<sup>15</sup> Dadas las limitaciones que puede presentar la información categorizada, se están haciendo modelos de imputación para asignarle, tanto a las variables que venían categorizadas como a las que no reportaban, ingresos un valor continuo.



en O\*NET o CIUO 08, digitada por los empresarios, presenta retos importantes, ya que se encuentra sin categorización alguna y aunque el análisis de frecuencias de palabras simples da algunas luces sobre las ocupaciones que requiere el país, es limitada porque no está estandarizada, ni permite la agregación o desagregación de datos ni comparabilidad internacional.

Organizar dicha información manualmente es una tarea muy costosa en términos de tiempo y recursos (habría que clasificar alrededor de 654000 registros de cada una de las ocupaciones correspondientes), por ello se ha recurrido al uso de *Web Services*<sup>16</sup> y a la búsqueda de patrones o palabras clave para identificar a qué categoría ocupacional corresponden los anuncios de las vacantes. En primer lugar, es necesario hacer una depuración del título de la vacante variable que tienen en común todas las fuentes de información por medio de la minería de texto se hace la eliminación de palabras que no contengan información sobre el perfil ocupacional, por ejemplo, “urgente”, “se solicita”, “necesario”; las preposiciones de la lengua española, nombres de empresas y los municipios y departamentos del país, entre otras.

Se cambiaron expresiones plurales por singulares para algunas palabras, por ejemplo, se cambió “vendedores” por “vendedor” o “asesores por asesor”, modificando las terminaciones de las palabras, en este caso “ores” por “or”; también se hizo la eliminación de tildes y otros signos de puntuación.

Por otra parte, se utilizaron motores de búsqueda, disponibles en la página web de la O\*NET, que al introducir ciertas palabras clave sobre ocupación, arrojan como resultado la denominación ocupacional más cercana (semánticamente) de acuerdo con esas palabras claves y su respectivo código de acuerdo con la clasificación O\*NET-SOC.

Ahora bien, aún sigue siendo muy costoso en términos de tiempo copiar uno a uno el título de la vacante en el buscador de la O\*NET, y copiar ese resultado en la base de datos. Para evitar este problema, una vez tenida la información del título de la vacante lo más depurada posible, se utiliza el *web service* de la página de Internet de la O\*NET, que permite hacer consultas con las palabras clave del título de la vacante, de manera automatizada, y descargar dicho resultado<sup>17</sup>.

Una vez obtenido el resultado de clasificación, por medio del *web service*, se realizó un proceso de revisión manual para la identificación de patrones

---

<sup>16</sup> El uso de *web service* facilita y agiliza el trabajo de codificación, ya que no es necesario programar cada uno de los patrones sino que este servicio arroja el resultado más similar a las palabras clave del título de la vacante, lo que deja solo para programar con aquellos casos donde el resultado del *web service* no fue correcto.

<sup>17</sup> Este servicio permite hacer 25.000 consultas de manera continua en un día.

que definitivamente dicho servicio no clasificó correctamente, dada la propia naturaleza coloquial del lenguaje colombiano, como por ejemplo el caso del “todero”, “islero”, “oficiales de construcción”, entre otros, a los cuales se les aplicó el análisis de patrones de palabras, tanto del título de la vacante como la descripción, para poder asignar una clasificación acorde al requerimiento del empresario. Con esta última identificación de patrones, queda lista la variable ocupación con la clasificación, la cual permitirá otorgar a cada categoría los conocimientos, aptitudes y habilidades que se requieren en cada ocupación, además, con el uso de correlativas se puede pasar de la clasificación O\*NET a la clasificación CIVO 08 de la OIT para ejercicios de comparación con las estadísticas nacionales e internacionales.

De esta manera, se partió de una información existente pero desorganizada y se construyó una base de datos con 654310 vacantes (anuncios) que corresponden a 2201479 puestos de trabajo para 2014 con las variables más relevantes (ver anexo 1), para su análisis estadístico.

Como lo han señalado Kureková et al. (2014), en países en desarrollo, la información disponible en las páginas web es una fuente valiosa para analizar el mercado laboral, más aún cuando las encuestas de demanda laboral son escasas o nulas. Sin embargo, por la naturaleza de esta información, pueden existir problemas de sesgos, coberturas, representatividad y no respuesta, problemas que toda encuesta con diseño probabilístico es susceptible de padecer. Con el fin de poder identificar los alcances y limitaciones que tiene esta información web, a continuación se presentan algunas de las posibles estadísticas de la base de vacantes, con su respectiva comparación con las estadísticas de empleo oficiales, para validación de resultados.

Por medio de la Encuesta de Productividad y Formación del Capital Humano, aplicada por el DANE a tres sectores de la economía (industria, comercio y servicios) en 2012 se puede establecer que, independientemente del sector económico, las vacantes potencialmente capturadas mediante de las técnicas mencionadas anteriormente, pueden equivaler a más del 50 % del total de vacantes que se buscan en el país. Ahora bien, esta proporción podría incrementarse con el paso del tiempo, dado el mayor uso de páginas web y servicios de empleo para la consecución de trabajadores, por lo tanto, el uso de información de este tipo y su respectiva consolidación dan bastantes luces sobre lo que ocurre en el mercado laboral pues su proporción no es despreciable.

En Colombia no existe aún una encuesta de demanda que permita comparar las características de la demanda con el tipo de información que se encuentra disponible en la web. Por tal motivo, para analizar en qué medida la distribución de las ofertas de empleo reflejan la realidad del comportamiento del mercado laboral colombiano, se recopiló información sobre población municipal y

**Tabla 1.** Medios de búsqueda de personal

Medios para la búsqueda de personal	Industria	Comercio	Servicios
Servicio Nacional de Aprendizaje (SENA)- Servicio público de empleo	12,3%	13,7%	10,4%
Anuncios en medios de comunicación <sup>18</sup>	12,2%	10,8%	10,4%
Bolsas de trabajo de Universidades y otras organizaciones.	8,4%	6,9%	10,8%
Contacto con otras instituciones educativas	2,1%	2,7%	4,0%
Headhunters / Bolsas de empleo	6,7%	6,5%	6,7%
Bases de datos / registros propios	17,4%	18,3%	18,7%
Sitios web laborales	16,7%	13,7%	20,2%
Redes informales (amigos/conocidos/colegas, etc.)	23,8%	26,9%	18,0%
Ferías de empleo	0,5%	0,5%	0,8%
Proporción de vacantes potencialmente capturadas	56,2%	51,6%	58,4%

Fuente: elaboración propia con base en la Encuesta de productividad y formación del capital humano del DANE.

departamental, tomadas de las proyecciones de población realizadas a partir del censo de 2005, así como la población ocupada (demanda efectiva) calculada con base en la Gran Encuesta Integrada de Hogares (GEIH) de 2014.

Con ello se busca identificar si existe cierta similitud con los datos oficiales, aun así, si existen diferencias entre la información web y lo que se encuentra en encuestas oficiales puede ser porque existen brechas en lo que se demanda y en lo que se ocupan las personas.

Se comparó tanto la población como el número de personas ocupadas por departamento, con las cifras recopiladas en la base de vacantes. En la tabla 2 se observa que entre los cinco principales departamentos (Bogotá D.C., Antioquia, Valle, Cundinamarca, Santander y Atlántico), el ordenamiento entre el número de ofertas de empleo es correspondiente con el número de ocupados y con el total de población a cada uno de ellos. Asimismo, guardan similar orden de proporción los departamentos con menos vacantes y menos población general y ocupada (Caquetá, Chocó, La Guajira y Sucre). Vale la pena resaltar que la correlación entre número de ocupados y vacantes por departamento fue de 0,9115.

Por lo tanto, existe una correspondencia entre el tamaño del mercado laboral, medido tanto por el número de personas como ocupados, y la información

<sup>18</sup> Muchas de las vacantes publicadas en medios de comunicación como prensa nacional y local también hacen uso de las páginas web buscadoras de empleo, por ello esta categoría hace parte de las vacantes potenciales recogidas.

de vacantes. La información de vacantes está más concentrada en Bogotá en primer lugar por el tamaño de su economía y en segundo por el mayor uso de los medios informáticos por parte de los empleadores para la búsqueda de personal.

**Tabla 2.** Distribución de puestos de trabajo, población general y ocupada por departamento

Departamento	Población total	Ocupados	% Puestos de trabajo
Bogotá D.C.	16,30%	19,40%	59,0%
Antioquia	13,40%	13,10%	8,50%
Valle	9,60%	9,50%	6,00%
Santander	4,30%	5,30%	4,20%
Atlántico	5,10%	5,10%	3,10%
Cundinamarca	5,50%	5,20%	2,10%
Meta	2,00%	1,80%	1,90%
Risaralda	2,00%	2,20%	1,80%
Bolívar	4,30%	4,20%	1,30%
Tolima	2,90%	3,10%	1,20%
Caldas	2,10%	1,90%	1,20%
Nariño	3,60%	3,90%	1,00%
Boyacá	2,70%	2,80%	1,00%
Huila	2,40%	2,0%	1,00%
Otros Departamentos	1,0%	1,0%	1,0%
Magdalena	2,60%	2,20%	0,90%
Quindío	1,20%	1,60%	0,80%
Córdoba	3,50%	3,00%	0,70%
Cesar	2,10%	2,00%	0,70%
Norte Santander	2,80%	3,10%	0,40%
La Guajira	2,00%	1,80%	0,40%
Cauca	2,90%	2,70%	0,30%
Sucre	1,80%	1,70%	0,30%
Choco	1,00%	1,10%	0,30%
Caquetá	1,00%	1,00%	0,30%

Fuente: elaboración propia con base en la Encuesta de productividad y formación del capital humano del DANE.

También se realizó una comparación del salario que reportan las personas ocupadas en la GEIH con respecto al que ofrecen los empleadores por medio de la información de vacantes. En la figura 1 se puede observar la proporción por rangos ingresos recibidos de los trabajadores y los ingresos ofrecidos por las vacantes. Al incluir la totalidad de los ocupados de la GEIH, el rango salarial con mayor participación es el de ingresos menores de \$550000, mientras que las ofertas de empleo ofrecen principalmente salarios entre \$550001 y \$1000000. Ahora bien, puede que la información de vacantes se concentre más en las cabeceras del país dado que en el resto este tipo de sistemas para la consecución de personal puede no ser tan efectivo, por ello se analizan los ingresos de los trabajadores, teniendo en cuenta solo las cabeceras, sin embargo, se observó similar comportamiento al nacional.

Ahora, puede ser que la información de vacantes vaya más acorde con la información de los ocupados formales<sup>19</sup>, ya que los medios de búsqueda de empleo que se capturan a través de este sistema son los más formales. Al comparar, se observa mayor similitud en el comportamiento de los ingresos, donde existe mayor concentración tanto en vacantes con ingresos ofrecidos como en ingresos recibidos de los trabajadores, entre 550001 y 1000000 de pesos, también se observa similar distribución de ingresos a lo largo toda la serie<sup>20</sup>. De esta forma, parece ser que la información de vacantes puede describir en mejor forma el comportamiento del mercado laboral formal.

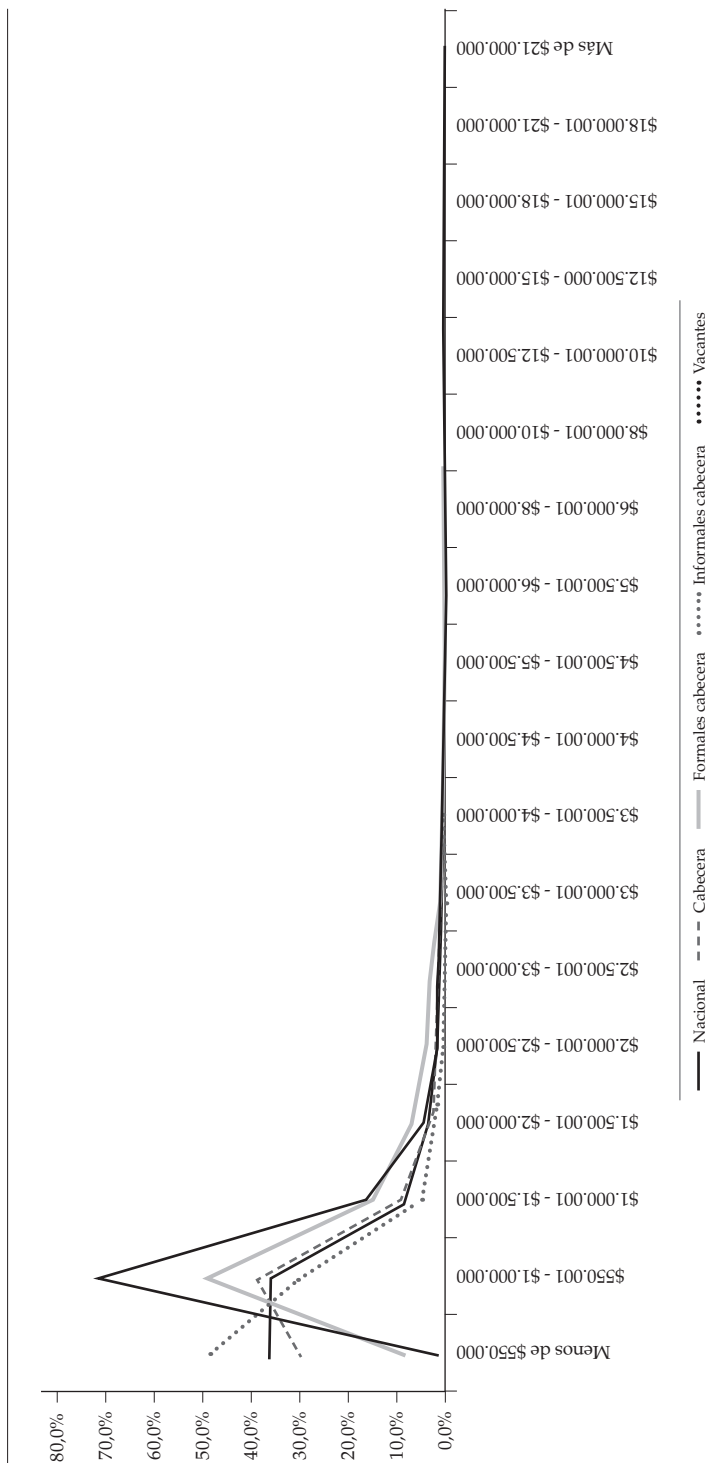
También se realizó un ejercicio de correlación para medir la relación que existe, a nivel de ocupación (a 3 dígitos), entre los reportes de la GEIH y los de la base de vacantes. El coeficiente obtenido fue de 0,7548. En este mismo sentido, en la tabla 3, se muestra las 15 ocupaciones más demandadas, según la base de vacantes (que recoge poco más del 56 % del total de vacantes en el país), y su correspondencia con el *ranking* de la GEIH, 9 de las 15 ocupaciones más demandas también hacen parte de las 15 primeras ocupaciones por parte de la demanda efectiva (GEIH).

De esta manera, existe cierta correspondencia entre lo que se encuentra por parte de la demanda y la oferta, sin embargo, también en esta tabla se puede interpretar como una forma de no correspondencia entre lo que se está

---

<sup>19</sup> Para medir formalidad se tomó la definición empleada por el DANE, en el cual se contabiliza como sector informal a todas aquellas empresas con menos de cinco empleados.

<sup>20</sup> Aunque la información de vacantes concentra en mayor proporción en el rango de 550.001 a 1 millón de pesos en comparación a los ingresos recibidos de los trabajadores, esto puede ser porque existe una brecha entre lo que se está ofreciendo y lo que se está demandando, lo que provoca una ausencia de personal en esos rangos salariales.



**Figura 1.** Distribución por niveles salariales  
Fuente: elaboración propia a partir de la base de vacantes y la GEIH (2014).

**Tabla 3.** Comparación de las ocupaciones encontradas en la base de vacantes y en la GEIH de 2014

Código Onet 3D	Grupo y Subgrupo	% Ocup GEIH	% Ofertas empleo	Ranking vacantes	Ranking GEIH
412000	Vendedores de puntos de venta al por menor	4,00%	9,10%	1	6
433000	Empleados Financieros	3,20%	5,20%	2	12
519000	Otras ocupaciones relacionadas con la producción	4,30%	5,00%	3	5
151000	Ocupaciones informáticas	2,10%	5,00%	4	17
435000	Distribución, despacho, agenda y registro	5,50%	4,00%	5	2
434000	Empleados de archivo e información	2,50%	3,40%	6	14
131000	Especialistas en administración de negocios	1,60%	3,30%	7	22
291000	Atención sanitaria y tratamientos	2,20%	3,10%	8	15
419000	Otras ocupaciones relacionadas con ventas	1,10%	2,90%	9	28
132000	Especialistas financieros	2,10%	2,80%	10	18
172000	Ingenieros	1,10%	2,80%	11	27
514000	Trabajadores del metal y el plástico	4,90%	2,70%	12	4
113000	Gerentes con especialidad operativa	1,80%	2,50%	13	19
436000	Asistentes administrativos y secretarios	3,50%	2,30%	14	10
537000	Transporte de materiales	3,60%	2,00%	15	8

Fuente: elaboración propia a partir de la base de vacantes y la GEIH (2014).

demandando y lo que se está ofreciendo en el mercado laboral colombiano. Aunque para determinar en concreto lo anterior, es necesario profundizar más sobre este aspecto, lo cual va más allá de los límites de este trabajo, pero que para futuras investigaciones es tema de vital interés sobre todo para la planificación y orientación del recurso humano.

Por otra parte, entre más alto el salario ofrecido mayor es la experiencia laboral exigida. Aunque la mayoría de salarios, independientemente de la experiencia exigida se encuentran entre \$ 550.001 y \$1.500.000 (Tabla 4).

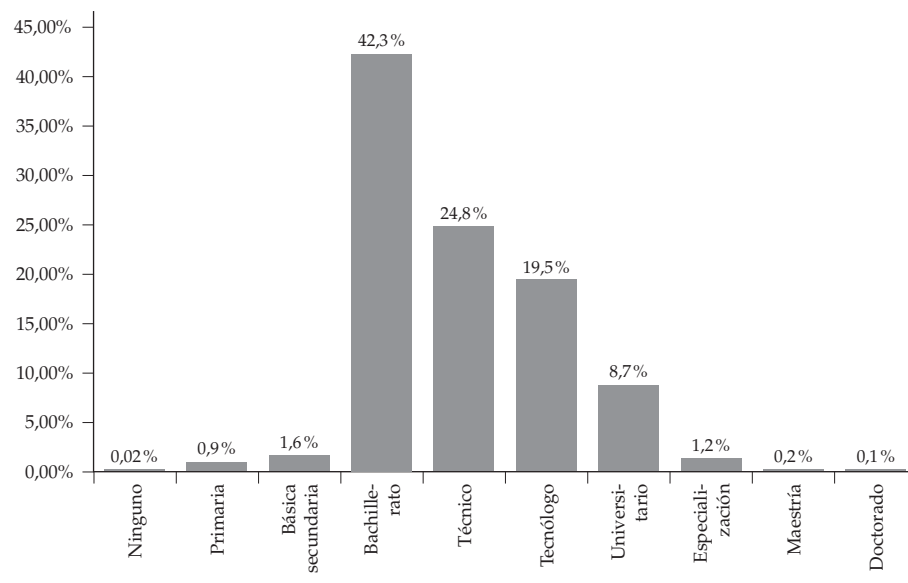
Aunque el nivel educativo más exigido por los empresarios es bachillerato (figura 2<sup>21</sup>), entre mayor nivel educativo mayor la remuneración. En la tabla 5, se observa que el 64,7% de las vacantes que exigen personas con doctorado ofrecen salarios mayores a \$5500000.

<sup>21</sup> Los porcentajes suman exactamente el 100%, ya que para un mismo anuncio de vacante se puede solicitar diferentes niveles educativos.

**Tabla 4.** Rangos Salariales y Experiencia

Rangos salariales	Menos de un año	1 año	2 años	3 años	4 años	5 años o más
Menos de \$550.000	1,50 %	0,50 %	0,00 %	0,00 %	0,00 %	0,00 %
\$550.001 - \$1.000.000	38,60 %	25,70 %	2,40 %	0,40 %	0,10 %	1,90 %
\$1.000.001 - \$1.500.000	7,70 %	7,40 %	1,90 %	0,50 %	0,10 %	0,30 %
\$1.500.001 - \$2.000.000	1,30 %	2,20 %	1,10 %	0,40 %	0,10 %	0,10 %
\$2.000.001 - \$2.500.000	0,40 %	0,70 %	0,70 %	0,40 %	0,10 %	0,10 %
\$2.500.001 - \$3.000.000	0,20 %	0,40 %	0,40 %	0,20 %	0,00 %	0,10 %
\$3.000.001 - \$3.500.000	0,10 %	0,50 %	0,20 %	0,10 %	0,10 %	0,10 %
\$3.500.001 - \$4.000.000	0,00 %	0,20 %	0,10 %	0,10 %	0,10 %	0,10 %
\$4.500.001 - \$5.500.000	0,00 %	0,10 %	0,00 %	0,00 %	0,00 %	0,10 %
\$5.500.001 - \$6.000.000	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %
\$6.000.001 - \$8.000.000	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %	0,10 %
\$8.000.001 y más	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %	0,10 %

Tomado de Guataquí et Al. (2014)

**Figura 2.** Distribución de los puestos de trabajo por nivel educativo

Tomado de Guataquí et Al. (2014)



**Tabla 5.** Rangos Salariales y Nivel Educativo

Rangos salariales	Ninguno	Primaria	Básica secundaria	Bachillerato	Técnico	Tecnólogo	Universitario	Especialización	Maestría	Doctorado
Menos de \$550000	0,00%	0,90%	3,30%	2,50%	2,40%	1,90%	1,90%	0,40%	0,20%	0,00%
\$ 550001 - \$1000000	89,90%	80,20%	84,00%	83,10%	69,60%	61,00%	32,00%	2,60%	9,40%	11,30%
\$1000001 - \$1500000	9,20%	16,00%	6,60%	12,00%	20,20%	26,80%	29,30%	4,30%	8,00%	4,20%
\$1500001 - \$2000000	0,00%	2,90%	2,70%	1,70%	4,50%	6,20%	15,00%	14,00%	9,90%	1,40%
\$2000001 - \$2500000	0,00%	0,00%	0,70%	0,20%	1,50%	2,00%	7,90%	21,80%	13,30%	2,80%
\$2500001 - \$3000000	0,00%	0,00%	0,00%	0,20%	1,00%	0,90%	4,40%	13,40%	20,30%	5,60%
\$3000001 - \$3500000	0,00%	0,00%	0,00%	0,20%	0,30%	0,40%	4,20%	8,70%	10,20%	2,80%
\$3500001 - \$4000000	0,00%	0,00%	0,00%	0,00%	0,30%	0,30%	2,90%	11,40%	8,50%	4,20%
\$4500001 - \$5500000	0,00%	0,00%	2,20%	0,00%	0,10%	0,20%	0,90%	9,40%	4,40%	2,80%
\$5500001 - \$6000000	0,00%	0,00%	0,50%	0,00%	0,00%	0,10%	0,40%	4,70%	4,60%	38,00%
\$6000001 - \$8000000	0,00%	0,00%	0,00%	0,00%	0,10%	0,10%	0,70%	5,20%	4,40%	4,20%
\$8000001 y más	0,90%	0,00%	0,00%	0,00%	0,10%	0,10%	0,40%	4,20%	6,80%	22,50%

Tomado de Guataquí et Al. (2014)

Por último en la tabla 6, se muestran las habilidades y competencias más demandadas, identificadas a través de los patrones en el texto del anuncio de la vacante. Para el 65,5% de las vacantes se requiere buena presentación personal, el 32,30% requiere buen servicio al cliente (Servicio) y para el 21% se necesitan conocimientos en Excel. Por su parte, existe demanda por conocimientos de programación en softwares como SAP, SIIGO Y HELISA en los cuales se diseñan productos informáticos de gestión empresarial como software contable y administrativo.

**Tabla 6.** Habilidades y competencias más demandadas

Habilidades	Porcentaje de vacantes	Habilidades	Porcentaje de vacantes
Presentación Personal	65,50%	Honestidad	1,50%
Servicio	32,30%	Ágil	1,40%
Excel	21,00%	Puntual	1,00%
Comunicación	13,90%	Emprendedor	0,90%
Seguridad	8,70%	Word	0,70%
Actitud	8,00%	Tolerancia	0,50%
Responsabilidad	6,90%	Sap	0,40%
Mercadeo	6,00%	Ortografía	0,30%
Liderazgo	4,90%	Matemática	0,30%
Proactivo	4,00%	Redacción	0,30%
Fluidez	3,30%	Siigo	0,30%
Access	3,30%	Helisa	0,20%
Compromiso	2,60%	Linux	0,10%
Office	2,10%	Regencia	0,10%

Fuente: elaboración propia a partir de la base de vacantes

## 11. Conclusiones

En este documento se han presentado, detalladamente, las justificaciones, lineamientos y resultados de la aplicación de un nuevo enfoque metodológico al análisis de la demanda de trabajo en Colombia. Se han sustentado las causas y consecuencias relacionadas con el diagnóstico de que el análisis de la economía laboral en Colombia ha estado claramente sesgado a desarrollarse a partir de información provista por el componente de oferta de trabajo del mercado laboral, en particular, las encuestas de hogares. A más de veinte años de que se iniciara el proceso de inserción del país en la economía internacional, por medio de la apertura económica, el país no ha podido desarrollar a plenitud los elementos de política pública del recurso humano requeridos para generar en este elemento estratégico las condiciones apropiadas para generar la competitividad que el factor trabajo debe proveer al sistema económico nacional y a la inversión extranjera directa.

Ante la evidencia planteada por este diagnóstico, los elementos fundamentales de esta propuesta parten de la premisa de aceptar que un alto y creciente porcentaje de la colocación laboral se está y se seguirá llevando a cabo mediante

elementos provistos en Internet. Una vez se sustentaron los aspectos históricos que giraron en torno al surgimiento de la corriente analítica del *big data*, se han ofrecido los principales principios de rigor metodológico relacionados con la herramienta, en particular, enfatizando el hecho de que no requiere de un diseño muestral estadístico. Se ofreció adicionalmente una breve reseña de la literatura pertinente, enfatizando experiencias semejantes en las que el enfoque de *big data* ha permitido superar cuellos de botella de la política pública, generando una plataforma analítica que ha permitido conocer, en gran detalle y a un costo comparativa y significativamente menor que aquel en que se habría incurrido al implementar una encuesta específicamente necesaria para el problema de política pública analizado.

A continuación se presentó la metodología en cada uno de sus componentes, describiendo el proceso cumulativo de obtención de información mediante convenios institucionales, la adición de información no institucionalizada mediante web crawling, la consolidación de la bodega de datos (datawarehouse) y los procesos de depuración de información valiosa que no se incluía bajo criterio genérico y la imputación de información relacionada con salarios y remuneraciones. Igualmente se describió en detalle la georreferenciación de las vacantes consolidadas a nivel de municipio.

Como resultado de la aplicación del enfoque metodológico que hemos desarrollado, se cuenta con una bodega de datos compuesta de 2.201.479 puestos de trabajo, para las cuales, por primera vez, y a diferencia del enfoque ocupacional de rutinas laborales que contiene el desarrollo de una labor (CNO 1970) y en el cual han sido compiladas históricamente las vacantes reportadas al Servicio Público de Empleo del SENA, las cuales, junto con la información de la Encuesta Anual Manufacturera y la Muestra Mensual Manufacturera, constituyen la referencia histórica más recurrente en el análisis de información laboral de demanda en Colombia. Se considera que, con el aporte que esta metodología puede ofrecer al análisis del mercado laboral colombiano y a la formulación de políticas públicas relacionadas con habilidades, competencias y requerimientos de la demanda, se puede generar un sólido, riguroso y representativo cuerpo de literatura que subsane la debilidad histórica que caracteriza el análisis de la economía laboral en Colombia.

## Referencias

Abraham, K. G., & Wachter, M. (1987). Help-wanted advertising, job vacancies, and unemployment. *Brookings Papers on Economic Activity*, 207-248.

- Acemoglu, D. & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4,1043-1171.
- Almeida, R., Behrman, J., & Robalino, D. (Eds.). (2012). The right skills for the job?: Rethinking training policies for workers. World Bank Publications.
- Álvarez, C. & Hofstetter, M. (2014). Job vacancies in Colombia: 1976–2012. *IZA Journal of Labor & Development*, 3 (15).
- Amoah, B. (2000). Help-wanted index. *Perspectives*, 14-18.
- Arango, C. & Rojas, Á. M. (2003). Demanda laboral en el sector manufacturero colombiano: 1977-1999. *Borradores de Economía*, Bogotá, D.C., Banco de la República, no.247.
- Autor, D. & Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, 31(2 Part 2), S59-S96.
- Barnichon, R., & Figura, A. (2010). Labor market heterogeneity and the aggregate matching function. *American Economic Journal*, en imprenta.
- Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2014). Microeconomic evidence of creative destruction in industrial and developing countries. *IZA Discussion Paper*, 1374.
- Bureau of Labor Statistics [en línea]: *Computer and internet use at work in 2003*. 2005 [fecha de consulta: 4 Abril 2014]. Disponible en: <http://www.bls.gov/news.release/pdf/ciuaw.pdf> /.
- Cárdenas, M., & Bernal, R. (2001). *Determinants of labor demand in Colombia: 1976-1991*, Social Science Research Network Electronic Library. [fecha de consulta: 28 enero 2014]. Disponible en: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=467545](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=467545).
- Carnevale, A., Jayasundera, T., & Repnikov, D. (2014). *Understanding Online Job Ads Data: A technical report*. Georgetown University. Center on Education and the Workforce, 1- 20.
- Cavallo, A. (2013). Online and official price indexes: measuring Argentina's inflation. *Journal of Monetary Economics*, 60(2), 152-165.
- Clark, K. A., & Phillips, M. A. (2002). *A comparison of job openings surveys: Concepts and survey design*. Bureau of Labor Statistics, Office of Survey Methods Research.
- ComScore [en línea]: *Travel and Careers in Focus at Start of New Year*. 2005 [fecha de consulta: 28 marzo 2014]. Disponible en: <http://comscore.com/content/download/7577/130595/file/comScore%20Media%20Metrix%20Ranks%20Top%2050%20U.S.%20Web%20Properties%20for%20January%202011.pdf> />.
- David, H. (2013). The “task approach” to labor markets: an overview. *Journal for Labour Market Research*, Institut für Arbeitsmarkt- und Berufsforschung

- (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany], vol. 46(3), pages 185-199.
- Davies, B., Gore, K., Shury, J., Vivian, D., Winterbotham, M. (2012). *UK Commission's Employer Skills Survey 2011: UK Results*. UK Commission for Employment and Skills. UKCES Evidence report, 38.
- Devins, D., Bickerstaffe, T., Nunn, A., Mitchell, B., McQuaid, R., Egdeell, V., & Lindsay, C. (2011). *The role of skills from worklessness to sustainable employment with progression*. UK Commission for Employment and Skills UKCES.
- Di Gropello, E., Tan, H. W., & Tandon, P. (2010). *Skills for the Labor Market in the Philippines*. World Bank. <https://openknowledge.worldbank.org/handle/10986/2514> License: CC BY 3.0 IGO.
- Dickerson, A., Wilson, R., Kik, G., & Dhillon, D. (2012). *Developing occupational skills profiles for the U.K.: a feasibility study*. UK Commission for Employment and Skills. Evidence Report 44.
- Du, D., Li, A., & Zhang, L. (2014). Survey on the applications of big data in Chinese real estate enterprise. *Procedia Computer Science*, 30, 24-33.
- Dunlop, J. T. (1966). *Job vacancy measures and economic analysis*. In *The measurement and interpretation of job vacancies*. NBER.
- Edelman, B. (2012). Using Internet data for economic research. *The Journal of Economic Perspectives*, 26(2): 189-206.
- Elias, P., & Purcell, K. (2013). *Classifying graduate occupations for the knowledge society*. Institute for employment research. University of Warwick. *Futuretrack Working Paper*, 5.
- Embley, D.W., Campbell, D.M., Jiang, Y.S., Liddle, S.W., Lonsdale, D.W., Ng, Y-K. & Smith, R.D. (1999a). Conceptual-model-based data extraction from multiple record web pages. *Data and Knowledge Engineering*, 31 (3), pp.227-251.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37.
- Franks, B. (2012). *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, New Jersey: Wiley.
- Froy, F., Gigúere, S., & Meghnagi, M. (2012). *Skills for competitiveness: a synthesis report*. OECD.
- Prentice, A. (2011). *CEO Advisory: 'Big Data' Equals Big Opportunity*. Gartner Analysis.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
- Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M. (1997), "Data Cube: A Relational Aggregation Operator Generalizing

- Group-By, Cross-Tab and Sub-Totals", *Data Mining and Knowledge Discovery*, 1(1): 29-53 (1997).
- Guataquí, J., Cárdenas, J., & Montaña, J. (2014). La problemática del análisis laboral de demanda en Colombia. *Perfil de Coyuntura Económica*, 24, pp. 71-107.
- Guojie L., & Xueqi C. Research Status and Scientific Thinking of Big data [J]. *Bulletin of Chinese Academy of Sciences*, 2012, 27(6).
- Healy, J., Mavromaras, K., & Zhu, R. (2011). Consultant report securing australia's future stem: Country comparisons. Australian Council of Learned Academies.. Recuperado de <http://www.acola.org.au/>, 6 sept. 2015.
- Holt, R., Sawicki, S., & Sloan, J. (2010). *A theoretical review of skill shortages and skill needs*. UK Commission for Employment and Skills. Evidence Report 44.
- Joseph, M. L. (1966). Job Vacancy Measurement. *Journal of Human Resources*, 59-80.
- Joyanes, L. (2013). *Big Data. Análisis de grandes volúmenes de datos en organizaciones*. México: Alfaomega.
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573.
- Kurekova, L, Beblavy, M & Thum, E. (2014). *Using internet data to analyze the labour market: a methodological enquiry*. IZA DP 8555, IZA, Bonn.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Technical report, McKinsey Global Institute.
- MINTIC [En línea]: Boletín Trimestral de las TIC. [Fecha de consulta 13 de noviembre 2015]. Disponible en: [http://colombiatic.mintic.gov.co/602/articles-11128\\_archivo\\_pdf.pdf](http://colombiatic.mintic.gov.co/602/articles-11128_archivo_pdf.pdf)
- Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, 61(3), 397-415.
- Mortensen, D. T., & Pissarides, C. A. (2001). *Taxes, subsidies and equilibrium labour market outcomes*. CEP Discussion Paper 519. London School of Economics and Political Science.
- Myllymaki, J. (2002). Effective web data extraction with standard xml technologies. *Computer Networks*, 39(5), 635-644.
- Reimsbach-Kounatze, C. (2015). *The proliferation of Big Data and implications for Official Statistics and Statistical Agencies: A preliminary analysis*. OECD Digital Economy Papers, No. 245, OECD Publishing. <http://dx.doi.org/10.1787/5js7t9wqzv8-en>
- Rothwell, J. (2014). *Using Internet data to analyze the labor market: A methodological enquiry*. Brookings Institution.

SIC [En línea]: *Estudio del sector telecomunicaciones en Colombia*. [Fecha de consulta 13 de noviembre 2015]. Disponible en [http://www.sic.gov.co/recursos\\_user/documentos/promocion\\_competencia/Estudios\\_Economicos/Estudios\\_Economicos/Estudios\\_Mercado/Estudiosectorial-Telecomunicaciones.pdf](http://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/Estudiosectorial-Telecomunicaciones.pdf)

Stops, M., & Mazzoni, T. (2010). *Matchingprozesse auf beruflichen Teilarbeitsmärkten/Job Matching on Occupational Labour Markets*. *Jahrbücher für Nationalökonomie und Statistik*, 287-312.

Stops, M. (2014). Job matching across occupational labour markets. *Oxford Economic Papers*, 66(4), 940-958.

Vivas, A., Farné, s. y Urbano, d. (1998). “Estimaciones de Funciones de Demanda de Trabajo Dinámicas para la Economía Colombiana”, *Archivos de Economía*, Bogotá, Departamento Nacional de Planeación, (julio), no.92

UKCES (2012). *UK Commission for Employment and Skills. Employer Perspectives Survey*.

## Anexo

**Tabla 7.** Variables creadas

Nombre de la variable	Leyenda de ayuda	Valores permitidos
Código	Campo único de identificación de la vacante	Campo numérico
Mes	Mes de descarga de la información	Campo numérico
Fuente	Identifica de dónde proviene la información	Campo texto
Título de la vacante	Cargo u oficio de la vacante	Campo texto
Descripción	Descripción de la vacante, actividades a realizar y requisitos para aplicar	Campo texto
Municipio	Municipio donde está disponible la vacante	Campo texto
Código_municipio	Código divipola del municipio donde está disponible la vacante	Numérico
Departamento	Departamento donde está disponible la vacante	Campo texto
Código_Departamento	Código divipola del municipio donde está disponible la vacante	Numérico
Empresa	Nombre de la empresa colocadora de la vacante	Campo texto
Cantidad_vacantes	Número de puestos de trabajo por anuncio	Numérico
Salario	Salario ofrecido	Menos de \$550000
		\$550001 - \$1000000
		\$1000001 - \$1500000

*Continúa*

Nombre de la variable	Leyenda de ayuda	Valores permitidos
		\$1500001 - \$2000000
		\$2000001 - \$2500000
		\$2500001 - \$3000000
		\$3000001 - \$3500000
Salario	Salario ofrecido	\$3500001 - \$4000000
		\$4500001 - \$5500000
		\$5500001 - \$6000000
		\$6000001 - \$8000000
		\$8000001 y más
Años_experiencia	Experiencia realizando las funciones de la vacante o relacionadas	Numérico
Fecha_publicación	Fecha de publicación de la vacante	Fecha
Fecha_de_vencimiento	Fecha de expiración de la vacante	Fecha
Inglés	Conocimientos en inglés	1=Sí 0=No
Francés	Conocimientos en francés	1=Sí 0=No
Español	Conocimientos en español	1=Sí 0=No
Mandarín	Conocimientos en mandarín	1=Sí 0=No
Portugués	Conocimientos en portugués	1=Sí 0=No
Alemán	Conocimientos en alemán	1=Sí 0=No
Hombre	Si la vacante requiere hombres, específicamente	1=Sí 0=No
Mujer	Si la vacante requiere mujeres, específicamente	1=Sí 0=No
Joven	Si la vacante requiere jóvenes, específicamente	1=Sí 0=No
Experiencia	Si la vacante requiere alguna experiencia laboral previa	1=Sí 0=No



Nombre de la variable	Leyenda de ayuda	Valores permitidos
Sin experiencia	Si la vacante no requiere experiencia laboral previa	1=Sí 0=No
Contrato_fijo	El contrato ofrecido es a término fijo	1=Sí 0=No
Contrato_indefin	El contrato ofrecido es a término indefinido	1=Sí 0=No
Contrato_aprendi	El contrato ofrecido es un contrato de aprendizaje	1=Sí 0=No
Contrato_servi	El contrato ofrecido es un contrato de prestación de servicios	1=Sí 0=No
Contrato_obra	El contrato ofrecido es un contrato por obra o labor	1=Sí 0=No
t_completo	Si se requiere trabajar la jornada completa	1=Sí 0=No
t_medio	Si se requiere trabajar media jornada	1=Sí 0=No
por_hora	Si se requiere trabajar por horas	1=Sí 0=No
ninguno	No se requiere ningún nivel educativo	1=Sí 0=No
primaria	Se requiere al menos nivel educativo de primaria	1=Sí 0=No
Básica secundaria	Se requiere al menos nivel educativo de básica secundaria	1=Sí 0=No
Bachillerato	Se requiere al menos nivel educativo de bachiller	1=Sí 0=No
Técnico	Se requiere al menos nivel educativo técnico	1=Sí 0=No
Tecnólogo	Se requiere al menos nivel educativo tecnológico	1=Sí 0=No
universitario	Se requiere al menos nivel educativo universitario	1=Sí 0=No

Continúa

Nombre de la variable	Leyenda de ayuda	Valores permitidos
Especialización	Se requiere al menos nivel educativo especialización	1=Sí 0=No
Maestría	Se requiere al menos nivel educativo maestría	1=Sí 0=No
Doctorado	Se requiere al menos nivel educativo doctorado	1=Sí 0=No
Código ocupación soc-o*net	Clasificación soc-onet	Numérico
Denominación ocupación soc-o*net	Clasificación soc-onet	Campo texto
Código Ocupación ciuo 08	Clasificación ciuo 08	Numérico
Denominación Ocupación ciuo 08	Clasificación ciuo 08	Campo texto
Conocimientos	Tipo de conocimientos de acuerdo a la O*NET	Campo texto
Aptitudes	Tipo de aptitudes de acuerdo a la O*NET	Campo texto
Habilidades	Tipo de habilidades de acuerdo a la O*NET	Campo texto