



Perfil de Coyuntura Económica

ISSN: 1657-4214

wilman.gomez@udea.edu.co

Universidad de Antioquia

Colombia

Guataquí R., Juan Carlos; Cárdenas R., Jeisson; Montaña, Jaime
La problemática del análisis laboral de demanda en Colombia
Perfil de Coyuntura Económica, núm. 24, 2014, pp. 71-107
Universidad de Antioquia
Medellín, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=86141380004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

La problemática del análisis laboral de demanda en Colombia

Juan Carlos Guataquí R. – Universidad del Rosario*

Jeisson Cárdenas R. – Ministerio de Trabajo

Jaime Montaña – Ministerio de Trabajo

DOI: 10.17533/udea.pece.n24a04

-Introducción-I. Disponibilidad de Información sobre la Demanda de Trabajo en el Sistema Estadístico Nacional (SEN). II. Contexto Normativo de la Información Laboral de Demanda en Colombia.-III. Una nueva propuesta de análisis laboral de demanda.- Referencias Bibliográficas.-Anexos

Primera versión recibida el: 25 de Agosto de 2014; versión final aceptada el 28 octubre de 2014

Resumen: Este artículo constituye el punto de partida de una nueva metodología analítica del mercado laboral Colombiano. Dado que dicha metodología se concentra en el análisis de información de demanda, el artículo se enfoca en presentar detalladamente la problemática existente al respecto. Diferentes secciones se dedican a presentar la problemática jurídica, la disponibilidad de datos y los incentivos teóricos cruzados en torno a la disponibilidad o control

restringido de la información de vacantes. Finalmente se presentan los lineamientos básicos de la nueva metodología analítica, la cual no ha sido aplicada a este nivel de agregación y capacidad de procesamiento en ningún país latinoamericano. Su potencial de aporte al análisis del mercado de trabajo y formulación de políticas dirigidas a enfrentar el desempleo friccional y cíclico es notable.

* Agradecemos a Diana Cárdenas Gamboa por su permanente apoyo y estímulo al desarrollo de esta investigación. Las opiniones y comentarios incluidos en este documento son totalmente responsabilidad de los autores y no comprometen al Ministerio de Trabajo. Correo electrónico: juan.guataqui@urosario.edu.co

Palabras Clave: Vacantes, Intermediación, Colocación, Web-Scraping.

Abstract: This article is the departure point towards the implementation of a new methodological approach to the study of the Colombian labor market. As this approach is based on demand-side statistics, we present on detail and scope the different dimensions of the troublesome landscape to be found when approaching labor demand-side analysis in Colombia. The design of regulations, labor policy, data availability and economic incentives of agents are all inquired into. Conceptual foundations of a new methodological, that had not been applied for the Latin American context, approach are presented. We emphasize its remarkable potential on sustaining policies against frictional and cyclical unemployment.

Keywords: Vacancies, Intermediation, Web-Scraping, Demand-Side labor economics.

Résumé: Cet article constitue un point de départ important concernant le débat sur une nouvelle méthode d'analyse appliquée au marché du travail en Colombie, laquelle tourne autour d'une discussion sur la demande d'information. Nous consacrons des différentes sections à la présentation des questions juridiques, à la disponibilité des données et aux justifications théoriques concernant les contraintes sur l'information des postes de travail. Enfin, quelques lignes de l'article sont dédiées à la présentation de la nouvelle méthode d'analyse, laquelle n'a pas encore été mise en œuvre à ce niveau d'agrégation dans aucun pays d'Amérique latine. Nous concluons que

cette méthode a un potentiel analytique remarquable en ce qui concerne la formulation des politiques de lutte contre le chômage aussi bien frictionnel que cyclique.

Mots-clés: Offre d'emploi, intermédiation, emplacement de travail, Web-Scraping.

Clasificación JEL: J63,

Introducción

La información es uno de los elementos más importantes en el funcionamiento adecuado de un mercado de trabajo y de una economía en general. La evolución del análisis laboral teórico ha reconocido la importancia que tienen tanto la heterogeneidad de trabajadores y puestos de trabajo, como la información existente al respecto. Si reconocemos el desempleo como un desequilibrio económico entre la dinámica de la demanda de trabajo y la de la oferta de trabajo, debemos reconocer también que este desequilibrio es generado por diferentes factores tanto de índole microeconómica como macroeconómica. Ofreciendo un espacio para dicha diversidad, pero por otra parte tratando de establecer un marco estandarizado de trabajo, la economía laboral ha postulado tres modalidades de desempleo para caracterizar las circunstancias que explican dicho desequilibrio: el desempleo cíclico, el estructural y el friccional.

El desempleo cíclico suele asimilarse a la caída de la demanda agregada y a la dinámica de la actividad económica en general. El desempleo estructural hace referencia a los problemas de desencuentro entre el perfil de las vacantes y el perfil de los des-

empleados, sus profesiones, calificaciones o habilidades y competencias específicas. El desempleo friccional también aborda la heterogeneidad de los perfiles, en un contexto de corto plazo, donde el desempleo se reduce si existe información adecuada sobre la demanda de trabajo y la oferta. A modo de analogía, el desempleo friccional sería cero si existiera una especie de Subastador Walrasiano Laboral. La Intermediación (colocación) Laboral se dirige a enfrentar particularmente el desempleo friccional, y de la información obtenida en el desempeño de dicha actividad se retroalimentan las políticas dirigidas a enfrentar el desempleo estructural.

La heterogeneidad de los trabajadores en relación a sus habilidades, niveles educativos y distribución del ingreso tiene importante influencia sobre la colocación laboral. Es claro que el aporte productivo que un trabajador, basado en su dotación de capital humano, tiene un valor estratégico para la firma. Para los individuos de alta calificación, su potencial aporte productivo hace que existan tanto para ellos como para las firmas, incentivos a enviar señales mutuas de atracción, lo cual facilita la colocación laboral. Desafortunadamente, este no es el caso para individuos con bajos niveles de educación, pertenecientes por ende a grupos vulnerables. Si bien se entiende que ante los bajos niveles educativos la intermediación debería tener bajo costo e incluso, en circunstancias de auge económico, permitir un rápido *matching*, esto no resulta tan expedito en recesiones o fases de transición. En estas circunstancias, más que involucrarse en un contexto de competitividad económica, la intermediación apunta hacia un asunto de bienestar: poner

a disposición de la mayor cantidad posible de individuos, la información más óptima en relación a la disponibilidad de oportunidades laborales, y apoyar este proceso bajo la premisa de que dicha información sea manejada con las herramientas requeridas: asesoría ocupacional, redes sociales de apoyo y otras opciones que permitan mejorar las condiciones de empleabilidad de los grupos vulnerables.

La intermediación/colocación laboral comprende las entidades y procedimientos que facilitan la interacción de oferta y demanda entre trabajadores y empresas, con el objeto básico de proveer información que facilite el encuentro entre las vacantes disponibles y los trabajadores. El sistema de intermediación regula el emparejamiento entre trabajadores y empleos, garantizando una mayor productividad y un mejor nivel de ingresos para las firmas y los trabajadores respectivamente, optimizando el costo del reclutamiento. En resumen, la intermediación concilia y resuelve, en la medida de lo posible, las brechas y distorsiones entre oferta y demanda laboral ocasionadas por la presencia de asimetrías e información imperfecta en el mercado laboral.

Existen diversas formas de intermediación laboral, debido en parte a la dificultad de acceso a información privilegiada acerca de las habilidades de los candidatos que aplican a una vacante en particular. Algunas de las formas más comunes de colocación laboral son proveedores de información como los servidores de búsqueda de empleo en Internet, sindicatos, agencias temporales de empleo, y agencias públicas de asistencia a los desempleados.

El objetivo primordial de la intermediación laboral es la reducción en los costos de búsqueda de un empleo. La mediación genera un mecanismo que permite a las empresas conocer atributos de productividad de los trabajadores que no se reflejan explícitamente en su experiencia educativa, al mismo tiempo que informa a los trabajadores de las características de una empresa en particular que no se reflejan en los perfiles de las vacantes disponibles. Una gestión óptima incrementa la eficiencia del mercado al optimizar los apareamientos de atributos trabajador-empleo, que a su vez favorecen la productividad y los ingresos laborales, y asegura un acceso equitativo a las oportunidades de empleo.

La reducción en los costos de búsqueda puede reducir la presencia de selección adversa en el mercado laboral. La teoría económica y la evidencia empírica sugieren que los mercados laborales resultan ineficientes cuando trabajadores y firmas no están igualmente informados acerca de la heterogeneidad en puestos de trabajo y en habilidades (*información asimétrica*), o cuando una de las dos contrapartes del contrato laboral conoce más que la otra: las características del empleo o las habilidades de los trabajadores (*información imperfecta*) en este caso. En modelos de información perfecta la reducción en los costos de búsqueda, basada en la literatura clásica sobre costos de transacción, conlleva un incremento en el ingreso nacional, y en los salarios, al mismo tiempo que a una disminución en el desempleo. En los modelos de competencia imperfecta basados en distorsiones de información se origina un incentivo al “engaño” por parte de

trabajadores o firmas en los contratos de trabajo (*selección adversa*), que afecta la asignación eficiente del recurso humano en el mercado laboral.

Siguiendo este argumento teórico, la intermediación laboral es una respuesta natural a la provisión sub-óptima de información pública en el mercado de trabajo, que funciona transando información a un precio menor al costo que implica su búsqueda por parte de la fuerza laboral y los empresarios. Sin embargo, la consecución de información privilegiada genera una ventaja informativa cuya utilización genera una renta monopolística, o poder de mercado. Aunque tradicionalmente los servicios de intermediación laboral han sido públicos sin cargo a los usuarios, la aparición de nuevos esquemas de intermediación con sistemas de información ágiles y versátiles, como es el caso de los portales de empleo en Internet, han incrementado el valor relativo de la información de mercado. La explotación de la ventaja informativa es contrarrestada con la aparición de esquemas mixtos en donde las agencias públicas disminuyen los costos de operación para garantizar excelencia en los competidores privados. En ambos casos, la reducción de costos de búsqueda, y la aparición de esquemas mixtos de asistencia técnica a los desempleados, reducen los problemas de selección adversa, y mejoran la distribución de oportunidades de empleo.

Bien podría decirse que para analizar el estado efectivo del componente de demanda en el país, es necesario decantar el sistema de intermediación laboral desde lo teórico a lo sumamente práctico. En un modelo teórico neoclásico, con

trabajadores y puestos de trabajo homogéneos, la información no es necesaria, la colocación es inmediata y el desempleo involuntario es inexistente. En un modelo con heterogeneidad surgen problemas de información incompleta y asimétrica, y aparece el desempleo involuntario. En la práctica, en particular en el mercado laboral colombiano, y como veremos, no sólo surgen problemas de información, sino adicionalmente relacionados con la ineficiencia de las instituciones diseñadas / contempladas para enfrentarlos.

En este orden de ideas, la principal conclusión de este artículo es: *al momento actual, es imposible realizar un análisis integral del comportamiento del mercado de trabajo colombiano, que involucre diagnósticos simétricos y paralelos a partir de información proveniente ya sea del componente de oferta o de demanda*. La estructura de este artículo es la siguiente: a esta introducción, a modo de contexto teórico y conceptual, le siguen 3 grandes secciones. La sección 1 se concentra en el aspecto de disponibilidad de microdatos generados a partir de encuestas con orientación de demanda y módulos laborales. La sección 2 sustenta la eventual conclusión del artículo (la existencia de un deficiente stock de información impide analizar de forma integral el comportamiento del mercado laboral Colombiano), desde la perspectiva

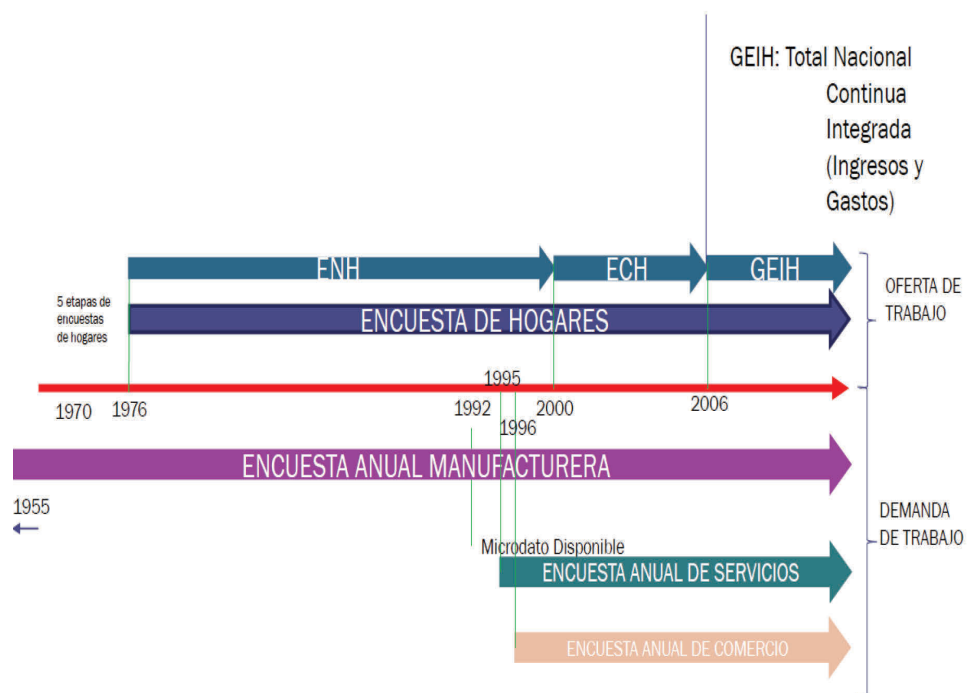
de las normas y arreglos institucionales que regulan el análisis de la información laboral desde la perspectiva de la demanda.

En la sección 3 se presentan, a modo de valor agregado, los principales lineamientos conceptuales de una nueva propuesta de análisis laboral basada en información de demanda, la cual supera las limitaciones identificadas en las secciones anteriores. En la cuarta sección se concluye.

I. Disponibilidad de Información sobre la Demanda de Trabajo en el Sistema Estadístico Nacional (SEN)

En esta sección presentamos los resultados de la revisión de fuentes de información sobre demanda de trabajo en el mercado laboral colombiano. El principal resultado de dicha revisión es que actualmente no se cuenta en Colombia con una fuente consolidada, geográfica y estadísticamente representativa, con adecuada periodicidad y calidad de información, sobre la demanda de trabajo. El gráfico 1 trata de sintetizar la disponibilidad, cronología, especialización sectorial y orientación de las principales encuestas con información laboral existentes en el Sistema Estadístico Nacional (DANE).

Gráfico 1
Información Laboral en el Sistema Estadístico Nacional.



Fuente: Elaboración de los autores.

a. Encuesta Anual de Servicios (EAS)

Esta encuesta se ha venido aplicando desde el año 1995, con incrementos progresivos de cobertura subsectorial (hotelería, publicidad, seguridad privada, etc). Su objetivo principal, como es el de otras encuestas sectoriales, es el identificar el comportamiento económico de las actividades sectoriales bajo su campo de análisis. En este orden de ideas, contiene información sobre el personal ocupado (temporal, permanente, agregado por sexo) y sus remuneraciones. El máximo nivel de desagregación a nivel de microdato está dado por la unidad de observación, la cual en este caso es

“Grandes Empresas”, que diligencian un formulario electrónico donde, entre otros tipos de información, en el cual presentan información agregada sobre su planta de personal (ver Anexo Ficha 1).

b. Encuesta Anual de Comercio (EAC)

Ofrece la posibilidad de disponer de información estructural del sector comercial, continua anual a partir de 1996. De forma semejante a la EAS, su unidad de observación es la empresa, en este caso comercial, y la información de tipo laboral es de índole semejante a la de dicha encuesta (Anexo Ficha 2).

c. Encuesta Anual Manufacturera (EAM)

Como veremos en la tercera sección de este documento, la EAM ha sido la fuente más recurrentemente utilizada para el análisis de la demanda de trabajo en Colombia. Posee una extensa cobertura cronológica, con continuidad anual desde el año 1955, en gran parte apoyada por cinco censos económicos (1944, 1954, 1971, 1990 y 2005). A nivel de microdatos, está disponible desde 1992. Se concentra en el sector industrial manufacturero, teniendo como agente representativo las firmas de dicho sector (establecimientos con diez o más personas ocupadas o un valor de producción estipulado anualmente, que funcionan en el país y que se definen como industriales, Anexo Ficha 3). Su módulo laboral es el más extenso y detallado que se puede encontrar en encuesta alguna de demanda a nivel nacional, pero la unidad de análisis es la planta de personal de la firma, no cada uno de los trabajadores. A partir de dicho lineamiento, la información laboral disponible, a grandes rasgos, permite analizar el nivel de calificación, posición ocupacional, tipo de vinculación, sexo y remuneración de trabajadores administrativos (*white-collar*) y operativos (*blue-collar*).

d. Encuestas Salariales

Las necesidades de reclutamiento de perfiles laborales específicos y de competitividad empresarial vía benchmarking salarial han llevado a numerosas firmas, de diversos sectores económicos, a utilizar encuestas salariales. En Colombia existen dos encuestas de este tipo, la Investigación Nacional de Salarios, producida por Acrip, y la Encuesta Salarial y Tendencias de

Gestión Humana, de Gestión Humana / Legis. La primera cuenta con mayor cobertura histórica y oportunidad cronológica (su último estudio fue aplicado en el año 2013), abarca 24 subsectores económicos, 731 firmas y 773 cargos estandarizados en las 9 principales ciudades del país. La Encuesta Salarial parece exhibir rezago cronológico (su última oleada aplicada fue en el año 2011), y su universo de aplicación hace referencia a 22 subsectores económicos, 530 cargos y 660 firmas. Si bien la información contenida en estas encuestas sería un insumo importante para el análisis integral del mercado de trabajo y la formulación de políticas públicas, la orientación estratégica de la información que contiene, tal y como lo mencionamos en las consideraciones teóricas presentadas en la Introducción a este documento, hacen que estas encuestas tengan un costo económico y un acceso restringido.

1.e Avisos Clasificados en Prensa

Este mecanismo de análisis sobre los requerimientos del segmento de demanda del mercado de trabajo fue el mecanismo más formal y estable de información, en particular para el segmento urbano de dicho mercado, hasta mediados/finales de los 1990s (Barnichon, 2010, p. 176). El Banco de la República ha realizado un aporte significativo al apoyar el trabajo de Álvarez y Hoffstetter (2013) quienes compilaron las vacantes anunciadas en avisos clasificados del diario El Tiempo, para la ciudad de Bogotá, en el periodo 1960-2010 y Arango (2013), quien realizó un ejercicio semejante para las siete principales ciudades de Colombia, desde 1976 al año 2012.

II. Contexto Normativo de la Información Laboral de Demanda en Colombia

Como hemos expuesto en nuestra revisión de fuentes de información, existen, a grandes rasgos, dos tipos de información de demanda de trabajo en Colombia. El primero de ellos hace referencia a puestos de trabajo, no necesariamente a vacantes, los cuales, si bien hacen parte de la demanda, no son el punto prioritario de análisis desde las políticas públicas. Información de este tipo es la utilizada en encuestas de salarios. Caso semejante ocurre con las encuestas de demanda del DANE, de carácter sectorial, aplicadas a firmas y establecimientos, en las cuales el análisis de la información de nómina constituye uno de los elementos analíticos de los determinantes de la evolución sectorial, pero no es particularmente detallado en las características de los puestos de trabajo.

El segundo está compuesto por información utilizada en actividades de intermediación / colocación laboral: las vacantes reportadas al Servicio Público de Empleo, las vacantes implícitas en los anuncios de empleo disponibles a través de Internet y las vacantes disponibles en bolsas de empleo y puestas a disposición de oficinas de egresados.

En aras del objetivo de política pública que se persigue, y a riesgo de ser extremadamente sintéticos, en el primer caso los puestos de trabajo / vacantes son un *medio* del objetivo de política, el cual es analizar los patrones y determinantes del comportamiento económico sectorial, análisis del cual el componente laboral es uno más. En el segundo tipo de informa-

ción, las vacantes hacen parte del *fin* de la política: el adecuado y exitoso encuentro entre un puesto de trabajo disponible y un potencial aspirante al mismo.

Hasta la promulgación de la ley 1636 de 2013, el país no contaba con un marco normativo adecuado para consolidar la información relacionada con la intermediación laboral, y en gran parte ello se debía a notorios vacíos jurídicos. Al revisar los términos del estudio de la Universidad Externado (2010) sobre la intermediación laboral, queda claro que gran parte de lo que dicho estudio considera “Agentes Naturales” de intermediación, no están regulados reactivamente por la ley, y tampoco cuentan con elementos legales que puedan permitir su uso proactivo.

A continuación enumeraremos los factores que, en nuestro criterio, explican esta situación. El primero de ellos es el confuso uso de la expresión “intermediación”. Surgió en el decreto 1421 de 1989, continuó en la Ley 50 (Arts. 95 a 97). Prosigue en el Decreto 3115 de 1997, que reglamenta dichos artículos. El uso de la expresión “intermediación” no es recurrente en la legislación que regula este servicio en varios países de América Latina. De hecho, las normas fundamentales (Convenios OIT 2 de 1919 y 88 de 1948) hacen referencia a “colocación”. El mismo Decreto 3115 sólo utilizaba la expresión “intermediación” al definirla (Art. 1) pero de ahí en adelante se refería a los agentes intermediadores como “Agencias de Colocación”.

Esto se combina con una marcada e identificada tendencia a satanizar la expresión, toda vez que como intermediación se entiende una forma de tercerización laboral.

Rojas (2004) resalta el notorio esfuerzo legislativo y de jurisprudencia, dirigido a impedir que bajo la figura de intermediación se permitan situaciones irregulares, pero la verdad es que los sucesivos gobiernos y sus legislaciones no han sido del todo exitosos. Ya el decreto 1433 de 1983 buscaba prohibir que las Empresas de Servicios Temporales realizaran actividades de intermediación, y el decreto 4588 de 2006, junto con la Ley 1233 de 2008 buscan que las Cooperativas de Trabajo Asociativo no realicen Intermediación Laboral. Esta voluntad puede llevar a niveles en que el remedio es peor que la enfermedad. Un ejemplo del máximo punto de confusión puede apreciarse en el Boletín de Prensa 040 de 2011 del Ministerio de Protección Social, cuyo encabezado pregona: “El Ministerio de la Protección Social Trabaja Contra la Intermediación Laboral”. Otro ejemplo es el estudio de Ermida y Colotuzzo para la OIT (2009) en el cual la “Intermediación” se considera dentro de la categoría conceptual “Subcontratación de Obras y Servicios” y la definición de Intermediación en la normatividad Colombiana se señala que el intermediario es quien contrata los servicios de trabajadores para que presten labores en beneficio de un empleador (sugerimos incluso una contradicción entre la definición de “Intermediador” del Art. 35 del Código Sustantivo de Trabajo y la definición de “Intermediación” del Decreto 3115 de 1997).

Desde la perspectiva de formulación de políticas la legislación Colombiana no definía claramente que es Intermediación Laboral, y por ende dicho vacío: i) no regula adecuadamente a los agentes de colocación ya contemplados en la Ley y, ii) sitúa en

una zona difusa a otros agentes “naturales”. Como ejemplo de esta última situación, existe evidencia anecdótica relacionada con las circunstancias de varias oficinas de egresados de universidades públicas y privadas, las cuales desarrollaban actividades de intermediación en el sentido del Art. 1 del Decreto 3115, pero negaban hacerlo a la luz de los artículos restantes de dicho decreto y más aún, del Art. 35 del CST.

La principal consecuencia de este vacío normativo era que no se contaba con el elemento coactivo que incentivara a las agencias de colocación privadas con ánimo de lucro, a proveer al gobierno la información relacionada con las vacantes que manejaban y el perfil de los aspirantes que recibían. Por ende no se podía contar con un panorama global de la intermediación laboral.

Como reacción a esta circunstancia, con la promulgación de la Ley 1636 y sus decretos y normas reglamentarias, el gobierno colombiano corrigió este vacío normativo, al redefinir la intermediación laboral como “servicios de gestión y colocación de empleo”, crear el Mecanismo de Protección al Cesante, y coaccionar a los diversos agentes que prestaban servicios de gestión y colocación de empleo, a proveer informes estadísticos mensuales sobre el movimiento de demandas y ofertas de trabajo, colocaciones y demás.

III. Una nueva propuesta de análisis laboral de demanda

A partir del año 2011 se establecen los lineamientos de la política para el fortalecimiento del Sistema de Capital Humano,

dentro de lo establecido en el Conpes 3674 de 2010. El objetivo principal de dicho sistema es que el mismo esté alineado con las necesidades productivas y sociales del país con el fin que la acumulación de capital humano sea acorde con los requerimientos del sector productivo y con los lineamientos de competitividad y política dictaminados por el gobierno central. Es de resaltar en este punto dos elementos que deben ser analizados con atención dentro de la estrategia propuesta: las necesidades del sector productivo y los lineamientos de competitividad y política del gobierno central.

Sobre las necesidades del sector productivo hay que mencionar dos hechos que son de suma importancia. El primero tiene que ver con la información disponible actualmente al respecto: No existe en Colombia una fuente de información sobre las necesidades de recurso humano por parte del sector productivo, en términos de habilidades, saberes y competencias. Existe, sin embargo una información parcial al respecto que es la obtenida por las mesas sectoriales. Sin embargo desde el punto de vista práctico, la información de las mesas sectoriales no responde a la necesidad de formación y a la demanda de trabajo, sino a la homogenización de los procesos, y esto en el mejor de los casos, dado que la efectividad de las mismas se ha puesto en discusión en diferentes escenarios, en cuanto no aporta a la pertinencia de la formación sino al contenido del trabajo.

El segundo hecho a resaltar es la información disponible que se tiene sobre las necesidades futuras del sector productivo. No existen en el país herramientas cuan-

titativas que aporten información sobre la estructura productiva y el cambio en la misma y como esta puede variar por la introducción de nuevas tecnologías de producción. Los modelos de pronóstico existentes en su mayoría se basan en proyecciones de la matriz insumo producto, con lo cual no se tiene información sobre fenómenos que deben tenerse en cuenta cuando se quiere modernizar e implementar sectores industriales con altos componentes de tecnologías de la información. Hoy en día es ampliamente aceptado el concepto de destrucción creativa dentro del mercado laboral. Ninguno de los sistemas existentes han incorporado esto, principalmente debido a que existe un vacío de información, pues no se conoce la estructura ocupacional de los sectores. Es por esto que ninguno de los sistemas de pronóstico es útil para la correcta planeación del recurso humano en cuanto no se conocen ni las necesidades actuales del sector productivo, ni los cambios remanentes de las políticas de competitividad que implemente el gobierno central.

El concepto principal sobre la cual se debe construir la estrategia de recurso humano, es el principio de *pertinencia*; sin embargo para que se pueda hablar de esta y poder evaluarla, se deben desarrollar sistemas de evaluación y monitoreo, así como un inventario de definiciones comunes para que se puedan identificar las carencias y se puedan identificar correctamente las necesidades del sector productivo. Hoy en día un inventario de competencias y cualificaciones no existe en Colombia, por el cual el objetivo último del Sistema de información de Capital Humano es que exista transparencia en estos conceptos,

y se tenga un lenguaje común para la medición, evaluación e identificación.

Desde el punto de vista de la academia conocer la estructura del mercado de trabajo, en términos ocupacionales, es de suma importancia. No obstante en términos de política, ¿urge ahondar en esto? ¿Es relevante esta información para la toma de decisiones y la evaluación de las políticas en el mercado de trabajo? La respuesta no es evidente, pues hay que conectar varias ideas para dar respuesta. Lo primero es tener en mente que cuando se identifica la estructura ocupacional de un sector, y las habilidades y competencias de cada una de las ocupaciones, se puede dar pronta respuesta y anticipar a crecimiento o ciclos de crecimiento del mismo. Con lo cual se tiene una respuesta para la mitigación del desempleo. De esta misma forma se articulan las políticas de formación, de innovación y competitividad sectorial, y por último se aumenta la empleabilidad por medio de políticas activas de empleo. Por lo cual la respuesta es afirmativa, a las preguntas arriba propuestas, en cuanto la información ocupacional en términos de habilidades y competencias permite por medio del ajuste en la oferta de formación aumentar la empleabilidad y disminuir el desempleo.

Con el ánimo de darle la importancia y responsabilizar institucionalmente al gobierno, el Decreto 4108 de 2011 por el cual se crea el Ministerio del Trabajo, asigna unas responsabilidades específicas a la Dirección de movilidad y formación para el trabajo. Dentro de estas están el monitoreo de las ocupaciones y la identificación de las tendencias ocupacionales. Esto im-

plica que las herramientas que se vienen utilizando sean obsoletas para tal fin y se deba desarrollar una metodología para el seguimiento y monitoreo ocupacional. A tal fin se han venido realizando diferentes estudios y se han aplicado diferentes metodologías cualitativas y cuantitativas con el fin de conocer las habilidades y competencias que el mercado requiere.

a. Contexto Conceptual: La necesidad de información laboral de demanda

Las discusiones sobre mercado laboral en Colombia, los modelos que se crean, los estudios contratados y las evaluaciones realizadas tienen un elemento en común desde hace más de dos décadas. En algún momento siempre la argumentación converge a un punto en el cual la discusión se y se optan por soluciones que no son óptimas, y que en muchos casos se alejan de lo técnico. Este punto, sin salida, es *que en Colombia no existe información adecuada sobre la demanda de trabajo*. Se ha intentado, durante el curso de las pasadas décadas, implementar una encuesta integral que ofrezca una contraparte conceptual adecuada a la ENH/ECH/GEIH, pero ha sido imposible, por una parte, coordinar a las instituciones, y por otra gestionar los recursos para tal fin. De este modo se ha venido abandonando la idea, y se ha perjudicado la discusión en temas de mercado laboral, tanto desde el punto de vista académico, como desde el punto de vista de la praxis política. Pero, ¿en términos prácticos porqué es importante analizar la demanda de trabajo?

Dentro del concepto de “creación destructiva” de Schumpeter (Reinert y Reinert, 2006), el crecimiento se da por

la sustitución de (trabajos en) sectores productivos por (trabajos en) sectores comparativamente productivos. Con esto hay un cambio en la mano de obra y en el tipo de empleos que se crean. Es un determinante fundamental, que el recurso humano pueda gestionar esta movilidad y adaptación de saberes para que no existan en la transicionalidad altas tasas de desempleo, y se pueda promover el cambio para estar bajo la frontera de posibilidades de producción rápidamente.

El mecanismo por el cual ocurre este cambio es la destrucción de empleos para el cual el retorno del sector está por debajo de los demás sectores y que para mantenerlo se debe incurrir en costos mayores o iguales a un sector en donde el ratio de capital humano y capital físico es menor. El estudio de la demanda de trabajo es fundamental para conocer los determinantes que incentivan o no la creación de más puestos de trabajo, útiles en los diferentes horizontes de política. Por lo cual contar información de la demanda de trabajo es relevante desde el punto de vista de la planificación y de la aplicación de las políticas de desarrollo y competitividad, además porque ayuda a aliviar a transicionalidad de la destrucción creativa de los puestos de trabajo. Hoy en día, esta información al no existir, se imputa de la información obtenida de las encuestas de hogares, es decir de la demanda efectiva, sin tener el detalle de habilidades y competencias de cada ocupación.

En Colombia, en la actualidad, no existe una fuente de información consensual y rigurosa sobre ocupaciones, en particular en términos de habilidades y cualificacio-

nes. Esto debe solucionarse para poder dar solución a asimetrías de información entre oferta y demanda de trabajo. Aún más grave no existe información sobre demanda de trabajo insatisfecha (vacantes) y menos aún, estimaciones de la demanda. Existe una hecho particular en los estudios de la economía colombiana donde la demanda de trabajo ha sido dejada a un lado; las estimaciones de funciones de demanda son aún más escasas (Vivas et Al., 1998; Arango y Rojas, 2003, Cárdenas y Bernal, 2001). La información de demanda de trabajo se ha evaluado únicamente desde tres fuentes:

- La demanda efectiva, que no tiene en cuenta las fuentes de demanda del mercado laboral, sino es tomada de la GEIH la cual parte de la oferta de trabajo, es decir desde las personas que son encuestadas en su muestra.
- Los modelos teóricos de equilibrio que se basan en cuentas nacionales, basadas en los supuestos teóricos, de los cuales se deriva la demanda de trabajo.
- Estudios que analizan el sector manufacturero como representativo del mercado de trabajo colombiano, a desmedro de la desindustrialización, el estancamiento del empleo industrial y su evidente tercerización (Parra, 2012)

Estudios de perfiles ocupacionales los cuales se recogen utilizando metodologías cualitativas y cuantitativas, caracterizando de demanda y oferta de trabajo para sectores específicos.

Sin embargo no existe información real que compare en los mismos términos la demanda y la oferta de trabajo. La proble-

mática que se presenta con los resultados, es que la información de la oferta y de la demanda “no hablan el mismo lenguaje”, por lo cual no son comparables. Es preciso resaltar que es necesario que la información de la demanda se encuentre pura, por lo que se tiene información respecto a la denominación de la ocupación. Por su parte se tiene la información categorizada de la encuesta de hogares, que está relacionada con la CNO de 1970. Esto hace que el análisis de las dos fuentes no sea compatible del todo.

Por otra parte, en los modelos de equilibrio general existentes en el país, la demanda de trabajo se infiere de la matriz insumo producto (de la estructura productiva) y su crecimiento pronosticado (Imputarle al crecimiento total la estructura identificada, por medio de factores fijos), y la oferta se infiere por medio de los perfiles que se recogen de las encuestas y ejercicios cualitativos. No obstante, desde el punto de vista teórico, estas inferencias no son iguales, en cuanto no representan la realidad del mercado pues no desglosan los niveles ni cualificaciones que debe contener la información de demanda de trabajo.

No existe en Colombia información cuya fuente sea la demanda de trabajo. Si bien existen iniciativas para hacerlo, estas resultan muy costosas en términos de recursos y de tiempo (es un proceso que requiere tiempo, con resultados lejanos, por lo cual no resulta de prioridad). No existe una comparación de las brechas de demanda y oferta en términos ocupacionales, que son relevantes porque son insumos para la diseño de políticas de corto y mediano plazo. Hasta ahora la información provista

por los métodos tradicionales solo hace cuenta del largo plazo. Se hace evidente la necesidad de herramientas que puedan proporcionar esta información.

b. La Solución Propuesta

El proyecto que estamos desarrollando tiene como objeto crear el soporte informático y metodológico para realizar, de manera continua y rutinaria, el estudio de brechas de demanda y oferta del mercado laboral colombiano, utilizando diferentes metodologías de análisis (cuantitativas tradicionales y no tradicionales como las relacionadas con los análisis de texto). El estudio de brechas se hará teniendo en cuenta que deberá estar contemplado el análisis en términos de ocupaciones, y estas deberán tener las habilidades y competencias requeridos por el mercado de trabajo.

Dado que en este momento no se cuenta con una clasificación nacional de habilidades y competencias, nos limitaremos a usar los diccionarios y taxonomías internacionales como referentes, lo cual permitirá hacer el análisis comparativo a nivel internacional. Esta aproximación hace que desde el punto de vista de política se puedan adaptar los programas y la oferta de formación a los estándares internacionales, con lo cual se puede promover el cambio tecnológico en las empresas pues el recurso humano contará con las habilidades para realizar dicha transformación.

Es importante además poder contabilizar de alguna manera la mano de obra que es requerida por el sector productivo, o contar con un proxy de esta información, dado que de este modo se puede ajustar la

oferta de trabajo creando mecanismos de actualización de saberes. Sobre este punto es importante resaltar el papel que tiene las competencias transversales, es decir la identificación de competencias que son comunes entre las ocupaciones, con el fin de proveerlas y facilitar la movilidad de recurso humano entre sectores.

Aproximación a metodologías de Big Data dentro del Estado y la academia

Desde más de una década, con la proliferación del uso de internet, se ha venido almacenando grandes cantidades de información sobre nuestro comportamiento. Transacciones de tarjetas de crédito, información de redes sociales, registros en plataformas, e información web se viene almacenando y está disponible para acceso público. La consultora internacional McKinsey fue la primera en acuñar el término *Big Data*, refiriéndose a volúmenes grandes de datos de diferentes fuentes y los mecanismos, algoritmos, administración y manejo de estos para obtener análisis y resultados del comportamiento de un modelo de negocio o mercado.

El almacenamiento de toda esta información abre fronteras para la investigación en diferentes áreas del conocimiento (Guojie y Xuequi, 2012). Edelman (2012) recoge las iniciativas que se han llevado hasta el momento bajo este nuevo esquema de datos, en los cuales se pasa de los conjuntos de datos estructurados, al levantamiento de información, estructuración y posterior cruce de la misma. El autor en su documento "Using Internet Data for Economic Research", recoge las experiencias dentro

de la economía utilizando esta perspectiva. Los ejemplos son variados y se encuentran desde aplicaciones macroeconómicas para levantamiento de precios y mirar la rigidez de los mismos (Carvallo, 2011), aplicaciones microeconómicas para identificar diferenciación de precios (Maier, 2005), hasta aplicaciones economía política (Hsie et al, 2011) y de economía de la salud (Ginsberg et al, 2009). De esta forma se deja a un lado la visión de los datos estructurados y se comienzan a aplicar metodologías para estructurar datos de diferentes fuentes para tener resultados y poder evaluar el comportamiento de diferentes variables.

Sin embargo esta información no es solamente una fuente de información, sino que representa una fuente de crecimiento y adquisición de conocimiento, que resultan en un aumento de la productividad y aumentan el bienestar del consumidor (Manyika, 2011). El modo para sacarle provecho es por medio de técnicas que se encasillan dentro de la minería de datos. La minería de datos tiene como fin buscar e identificar patrones inusuales e interesantes en los datos. Es así como según Fayyad et Al. (1996) estas aplicaciones de análisis de datos pueden definirse como la aplicación automatizada de algoritmos para detectar patrones en los datos.

De acuerdo a Bhandari et al. (1997), para que este procedimiento se lleve a cabo de forma correcta es necesario que la recolección de datos haya sido realizada utilizando un sistema especializado de recolección diseñado específicamente para la información que se va a manipular. Una vez descargada la totalidad de información se debe iniciar un pre-procesamiento de

los datos en donde se realizan pruebas de consistencia, las cuales muestran si es necesario o no realizar procesos de limpieza, transformaciones y enriquecimiento de la información, con el fin de asegurar que los datos sean tan exactos como sea posible antes de iniciar su análisis.

Estos análisis categorizan los valores y tendencias de los datos, extraen información estadística, y luego contrastan las categorías entre sí. De acuerdo a Gray et al. (1997), para estos análisis se requieren tres pasos:

1. Formular una consulta que extraiga los datos relevantes de una base de datos suficientemente amplia.
2. Extraer, en un archivo o una tabla, los datos agregados de la base de datos.
3. Analizar los resultados y formular una nueva consulta.

Sobre los resultados de este proceso hay que tener presente un cambio de paradigma en el modo como se enfrentan los datos comúnmente desde el punto de vista científico, porque se pasa de un paradigma en el cual no se persigue la exactitud y se acepta la incertidumbre y se deja a un lado el principio de causalidad por aceptar la correlación entre los mismos (Danyang, 2014). Este cambio de paradigma, implica un cambio en el modo como se conciben los datos desde el punto de vista científico, en el cual se construyen muestras basados en poblaciones y a partir de esto se hace inferencia estadística. Con la aparición de Big Data el problema de representatividad se deja a un lado porque se cuenta con una gran cantidad de información y el problema que se enfrenta es la estructuración y las

nuevas herramientas para analizar esta gran cantidad de información.

Si bien esta metodología se ha aplicado de manera exitosa en diferentes tipos de mercados, tal como se ha anotado anteriormente, en el mercado laboral no se ha hecho. Esto se deba posiblemente a que en los diferentes países existe un registro administrativo de las colocaciones de trabajos nuevos, dentro de los sistemas de información de servicio de empleo. No obstante siempre estos datos se toman con cautela porque usualmente no representan la totalidad de la información del mercado. Sin embargo en Colombia un registro de esta naturaleza no existe hasta el momento, pero urge de manera inmediata realizar análisis de la demanda de trabajo, y sobre todo de las habilidades y conocimientos requeridos para poder realizar alguna propuesta con base en información real para el mejoramiento de políticas y el ajuste de la oferta de formación.

Cabe anotar que los análisis de Big Data no se han realizado en Colombia hasta el momento, y solo se tiene análisis de los comportamientos de precios para la investigación de Carvallo de 2011, en donde se levantaron 3.9 millones de registros de precios. Este documento presenta la primera aproximación a esta metodología, y la primera dentro de análisis de la demanda en términos de competencias y habilidades.

Este uso de internet ha hecho que existan cambios en el mercado de trabajo. En si también en el modo en el que buscan trabajo las personas. Para países como EEUU, 1 de cada 10 ciudadanos americanos, ha buscado trabajo por medio de internet en el

último año¹. En enero del 2011, comScore, una firma dedicada a la contabilidad de visitas y sitios de internet, contabilizó un pico de 80.2 millones de visitas a sitios de empleo². De acuerdo con reportes de estadísticas más recientes estas cifras vienen aumentando y no solo las ofertas en los sitios en internet, sino las ofertas dentro de los portales en internet han hecho que más de un 45 por ciento de las ofertas de trabajo para 2013 se hayan hecho vía web (de manera pública o privada)³. Es por tanto que una de las formas para poder abarcar el mercado laboral, es por medio de este tipo de información que es pública a los buscadores de empleo. Desde nuestra aproximación lo que queremos lograr es identificar la mayor cantidad de fuentes de información y construir una bodega de datos que nos permita analizar la información que está alojada allí.

c. Open Data

Hoy en día la información es uno de los insumos más relevantes para muchas industrias que están basadas en el conocimiento. El mercado laboral y la información de mercado laboral no radica solamente para temas de política pública, sino también para el sector privado. Con esta información pueden hacer sus proyecciones de consumo, ahorro e inversión, variables que están correlacionadas.

No obstante hoy en día la mayoría de la información en Colombia es privada. Y en muchos casos la información que se recoge por instituciones públicas no está disponible al público. Esto hace que muchos de los análisis, y que innovaciones no se puedan implementar por el hecho que estos datos sean privados. Es por eso que es necesario cambiar de esquema, para un esquema en el que se permita la colaboración entre instituciones, y entre agentes del mercado. Esta visión propende por acercar el Estado a la ciudadanía, pero ante todo hacerla participe del análisis. De este modo la gobernabilidad se hace con transparencia y se pueden crear innovaciones desde el sector público y privado para poder llevar cambios e identificar problemáticas. Es por esto que dentro de lo que llevamos a cabo en este proyecto de investigación, queremos manifestar la necesidad de que existan este tipo de iniciativas, porque este trabajo si la información estuviese disponible se hubiese llevado a cabo seguramente hace mucho tiempo. Por temas de propiedad de la información estamos dejando de implementar y evaluar las políticas correctamente, y eso constituye un craso error. Por esto mismo se deben llevar a cabo proyectos privados e institucionales de datos abiertos, comúnmente conocido como Open Data.

Open data es el término que se refiere a la información que está disponible para todas

1 <http://www.bls.gov/news.release/ciaww.toc.htm>

2 <http://www.comscore.com/content/download/7577/130595/file/comScore%20MMX%20Ranks%20Top%2050%20U.S.%20Web%20Properties%20for%20January%202011.pdf>

3 Por web privada se entienden las ofertas generadas y alojadas en los sistemas internos de las compañías. Muchas de estas vacantes no salen al mercado público.

las personas. Esta información puede ser utilizada para cualquier propósito y sin ningún costo, previo el reconocimiento de la fuente, y de que no se pueda lucrar manipulando esa información.

Para que sea considerada como open data, la información debe contar con una licencia en la que se estipule que las personas que usen la información deben atribuir el crédito a quien sea que la esté publicando, y que las personas que mezclen esta información con otra deben entregar los resultados como open data.

Los principios de una adecuada open data son:

- Puede ser citada o referenciada, de modo que pueda ser fácilmente compartida y discutida.
- Está disponible en un formato estándar y estructurado, de modo que pueda ser fácilmente procesada.
- Garantiza la disponibilidad y consistencia en el tiempo, para que los demás puedan confiar en esta.
- Es rastreable, a través de cualquier proceso, de regreso al lugar donde se originó, para que otros puedan decidir si quieren confiar en ella y replicar los análisis y resultados.

Las iniciativas de Open data cada vez son más frecuentes en las democracias alrededor del mundo, buscando publicar los datos que actualmente son propiedad del gobierno y contribuir voluntariamente con los datos corporativos. Esto brinda transparencia y visibilidad de las instituciones gubernamentales, lo que implica que

aumente la confianza de los ciudadanos en las acciones del Estado.

Se traslada la información a una base de datos de un aplicativo de consulta de información. Esta información estará disponible para cualquier usuario en formato plano de texto para que sea manipulada, y también podrán generarse reportes en línea con las variables que se quieran. La idea es poder generar el flujo de información para que se dé un ajuste de mercado por retroalimentación. La información documental se organiza por categorías de contenido y la plataforma debe permitir realizar búsquedas por temas específicos, contenido, fecha, etc. De esta forma la información puede ser consultada por cualquier persona. Este tipo de disposición de la información debe estar basada en el principio de Open Data, por lo cual los usuarios de la información son todas las personas que tengan acceso a un computador conectado a internet.

Los posibles usuarios interesados son:

- Las instituciones de formación en cuanto pueden ajustar su oferta en base a las previsiones de mercado. Además se crea un sistema transparente de precios de matrícula que puede dinamizar el mercado y abrirlo regionalmente y así resolver los problemas de cobertura existentes en la actualidad.
- Las empresas en cuanto pueden tomar la información para realizar sus ejercicios de planeación de recurso humano y ajustar sus procesos de búsqueda y ajustar sus modelos operativos, ajustar sus planes de capacitación, así como su esquema de costos, de acuerdo a

la información suministrada por el Sistema de información.

- Los sindicatos en cuanto pueden informarse de la dinámica del mercado de trabajo, y proponer mejoras para el aumento de los retornos del trabajador basados en la productividad y el aprendizaje que puede llegar a proporcionarse mediante acuerdos colectivos en lo referente a capacitación.
- Investigadores para realizar sus propios cálculos, sus propias visualizaciones y alimentar el debate.
- Otras entidades del Estado para ajustar sus políticas basadas en el mundo del trabajo.
- Interesados en desarrollar APIs para utilizar la información que ha sido centralizada. En general cualquier persona o entidad que quiera generar conocimiento a través del uso de los datos.

Lo que resulta de estas implementaciones dentro de esta perspectiva es que los investigadores y usuarios pueden llegar a los mismos resultados y promover nuevas metodologías y análisis lo que enriquece el debate dentro y fuera de la academia. En específico, en esta propuesta, se dan luces para ampliar el debate de mercado laboral en Colombia.

d. Documento de memoria—Obtención de la información

Las ofertas online representan solo una pequeña parte de las transacciones que se realizan en el mercado de trabajo en Colombia. Incluso en países como Estados Unidos o Europa estos no son siquiera la mitad del total de transacciones. No

obstante como se verá posteriormente, los análisis de género, salarios y otras variables dentro de la base recogida concuerdan con la información que se puede tener del análisis de la Gran Encuesta de Hogares levantada por el DANE. La gran diferencia en la naturaleza de la información entre las dos, es en la fuente, dado que para los datos de nuestra base de datos levanta la información corresponde a datos de demanda de trabajo, y para lo recogido en la GEIH corresponde a información de oferta (que en realidad es la demanda efectiva, los puestos que están siendo ocupados). No obstante es la primera vez en Colombia que se realiza un análisis de vacantes con una cantidad de datos como la que tenemos y con fuentes tan diversas como la información de bolsas públicas y privadas de trabajo.

Desde que se comenzó el levantamiento de información se tenía claro la diferente procedencia de los datos que se iban a almacenar. Esto implicó desde el principio que se contaran con esquemas diferentes y una estrategia de homologación de los diferentes datos. Por este motivo, con el fin de realizar un análisis comparativo entre las variables del sistema se realizó la siguiente estructura de bodega de datos. En esta se almacenan toda la información de las diferentes bolsas de empleo y se homologa a una única estructura con el fin de poder analizar y hacer comparable.

e. Fuentes de información

En este aspecto se tuvieron en cuenta los siguientes agentes:

Bolsas de empleo públicas: Se realizaron convenios con el SENA, el SPE (en donde

se recogen los servicios de empleo de las universidades y de las cajas de compensación familiar que son prestadores de la Unidad Especial del servicio Público de Empleo) y algunas universidades. La información se recibió en Archivo plano de

texto y fue convalidada según la siguiente estructura de datos. Se realizó un proceso de homologación de las variables, en cuanto siendo provenientes de diferentes fuentes, estas tenían una estructura diferente.

Tabla 1:
Estructura de la base de datos.

Campo	Descripción del campo	Tipo de Campo
Num	Numero de oferta. Es el ID de la vacante dentro de la bodega de datos. Éste número es único y la función es poder identificar la vacante dentro de la bodega. Sirve también para hacer cruces en caso de ser necesario entre grupos y subgrupos.	Autoincrementable
Oferta_empleo	Es el "título" de la vacante, es decir la denominación que recibe la ocupación. Esto aporta información para realizar la categorización y dividir en clusters la base para la identificación de las habilidades de las ocupaciones.	Texto
Nombre_empresa	Nombre de la empresa	Alfanumérico
Sector	Sector de la empresa	Texto
Perfil_empresa	Campo de texto en donde se describe la	Area de texto (memo)
Cargo_equivalente	Categoría de empleo.com. Se analizaron otras páginas (zonajobs, entre otros) en los cuales se definen unas categorías ocupacionales.	Texto
Area_cargo	Area en donde se desempeña la persona	Texto
Subnivel_cargo	Nivel dentro de la empresa	Texto
Experiencia_anos_tot	Experiencia total	Numérico int.
Experiencia_anos_cargo	Experiencia realizando las funciones de la vacante	Numérico int.
Ciudades	Ciudad	Texto
Salario	Salario	Numérico int.
Profesion	Es el título de la persona que requiere la vacante. P.e. Economía.	Texto
Nivel_estudios	Nivel de estudios. (p.e. Técnico, universitario, Bachiller)	Texto
Idioma	Idioma que requiere la persona para desempeñar el cargo	Texto
Cantidad_vacantes	Cantidad de vacantes	Numérico int.
Tipo_contrato	Tipo de contrato	Texto
Tiempo_dedicado	Tiempo que debe dedicar para ejercer la ocupación.	Numérico int.
Genero	Género	Boleano
Fechap	Fecha de publicación de la vacante	Fecha
Fechar	fecha de expiración de la vacante.	Fecha
Codigo	número de identificación fuera de la bodega de datos	Numérico int.
Descripcion	Descripción de la ocupación	Area de texto (memo)
Requisitos_aplicar	Requisitos para aplicar	Area de texto (memo)

Web Scraping: Se trata de una técnica de consulta de información en la red, en la cual se realizan consultas a una página web, y se lee y almacena el contenido que se muestra en la página. Para realizar esto se construyó un código que permitía interpretar la información que se visualizaba en la página web, para posteriormente descargar de manera automática y ordenada la información.

Para la construcción de los algoritmos de interpretación de la información se siguió una metodología que combina dos propuestas: la propuesta por D.W. Embley et Al. (1999), en la cual, para poder pasar de datos sin estructura a datos estructurados – que es como nosotros los economistas estamos acostumbrados a trabajar- se modela conceptualmente el ordenamiento y estructura de las páginas web que contienen la información; y por otra parte la metodología que hace uso de consultas web y aprovechar la estructura de estos lenguajes para el despliegue de la información y así realizar la captura⁴.

Con el fin de que se entienda que la información obtenida, es la misma que fue introducida por los empresarios que ponen las ofertas en la web, es decir que la fuente de información es la demanda de trabajo, se explica de manera sucinta en que consiste el procedimiento. Para entenderlo de una manera sencilla, piénsese una página web como una forma de visualizar consultas de contenido, de una manera ordenada. El navegador interpreta esto por medio del lenguaje HTML, el

cual contiene dentro de sí una estructura en el lenguaje (el lenguaje HTML es un lenguaje de tags “<>”) y una estructura en el diseño y despliegue de la página web. Analizando el código HTML, que es interpretado para mostrar la información, se pueden aprovechar esos patrones para realizar consultas sobre esos elementos, de modo que se puede extraer el contenido éste. La programación puede hacer que esta tarea se haga repetidamente y de una manera homogénea por lo cual se asegura que en el levantamiento de la información se asegure confiabilidad y homogeneidad. Técnicamente, este proceso se denomina el análisis del esquema HTML, y consiste en mirar los elementos marcados y ordenarlos según prioridad para entender la funcionalidad y el contenido: de esta forma los tipos, token, elementos y atributos se analizan y se leen por medio de patrones.

Para poder aprovechar estos patrones se utilizó el lenguaje de búsqueda REGEX, en el cual a documentos de texto se le aplican búsquedas basadas en la aparición de una secuencia de caracteres, para obtener la posición dentro de un conjunto. Esto permite que se pueda extraer la información entre dos marcas (tags), simplemente encontrando la posición de las marcas y extrayendo el contenido que está en medio.

Además de realizar esto por medio del lenguaje, el análisis de los modelos, estructura y template de la visualización de la información (tablas, campos y tipo de contenido de los mismos), permite crear filtros realizar y asegurar la descarga de

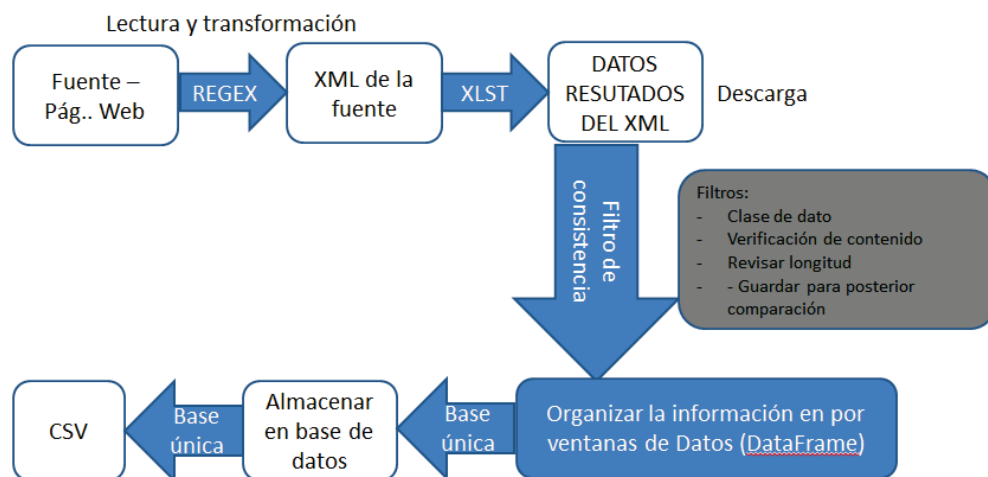
4 Ver J. Myllymaki (2002), Z. Lacroix (2003), Manuel Alvarez et al. (2007)

información y el contenido. Es decir conociendo la tabla y como se ordena la información en las tablas, márgenes y marcos de la página web, se creó un código de lectura y extracción de la información. No obstante respecto a este tipo de aproximación y metodología, aunque es eficiente y ayuda a la estructura del contenido, se asocia un problema: cualquier cambio de contenido en la estructura o visualización de la página o formato de las tablas, puede incurrir en un error del código y así no se obtenga información alguna. Esto implica esfuerzos adicionales en el mantenimiento

de los algoritmos de descarga y la continua supervisión de éstos.

Esta información se valida por tipo (por ejemplo si es un código que incrementa, si es una página web, si es un número o un texto) y posteriormente se copia en un formato XML. Esta conversión es necesaria para poder hacer la carga de información semiestructurada a campos estructurados dentro de la base de datos, y es una mejor práctica porque ante cambios en la información solo debemos cambiar la inserción en el documento XML y no en la base de datos directamente.

Gráfico 2
Mapa conceptual del proceso



Se automatiza una función para la cual toda la información que exista en la página web objetivo se descarga mediante un loop que revisa que la información sea descargada únicamente una sola vez tomando como referente la URL de destino. Esto debido

a que todas las bolsas de empleo muestran una oferta de empleo por cada URL en el dominio. Tras tener el resultado completo de la descarga se crea un XML para cada consulta hecha, tal y como se explica anteriormente.

El último paso es cargar el XML a una base de datos para su almacenamiento y consulta, para generar los archivos planos que nos permite utilizarlos en los diferentes paquetes estadísticos con los cuales estamos trabajando en el momento (Arcgis, R, STATA).

Frecuencia de la descarga:

Se analizó durante el periodo comprendido entre octubre de 2013 y enero de 2014 la frecuencia y tráfico de los portales web objetivos. Con esta información se automatizaron las descargas de manera bimensual, teniendo en consideración que los días en que se programan las descargas correspondan a un lunes o un viernes, de modo que correspondan a los días en los cuales hay un mayor número de ofertas activas. Se cuenta con información para las dos principales bolsas de empleo con esta metodología a partir de marzo de 2014 hasta la fecha.

f. Principales resultados

Esta sección tiene como objetivo presentar de manera general las descriptivas de todas las variables construidas a través de la base de vacantes consolidada por las diferentes fuentes de información.

De esta manera, sumando las diferentes fuentes de información homogenizadas y estandarizadas se cuenta desde febrero de 2014 hasta el mes de Junio de 2014 con 277,871 vacantes (tabla 2). La gran mayoría de las vacantes (92.29%) ofrecen solo un puesto de trabajo, pero también es de

resaltar que del total de avisos publicados 2.01% corresponden a contrataciones masivas donde se convocan más de 10 personas para realizar una misma labor.

Tabla 2

Número de puestos por vacante	Frecuencia	%
1	256,355	92.29
2	5,506	1.98
3	2,477	0.89
4	1,064	0.38
5	2,694	0.97
6	602	0.22
7	203	0.07
8	290	0.1
9	78	0.03
10	2,716	0.98
Más de 10	5796	2.01
Total	277,781	

Fuente: Ministerio del Trabajo. Cálculos propios

En la tabla 3 se puede observar la distribución departamental del total de puestos de trabajo, 65.47% de los puestos de trabajo se ofrecen en la ciudad de Bogotá, seguido de Antioquia (7.32%) y Valle (6.2%). De tal forma, se tienen 277,781 avisos de vacantes que en total ofrecen 614,832 puestos de trabajo en todo el país.

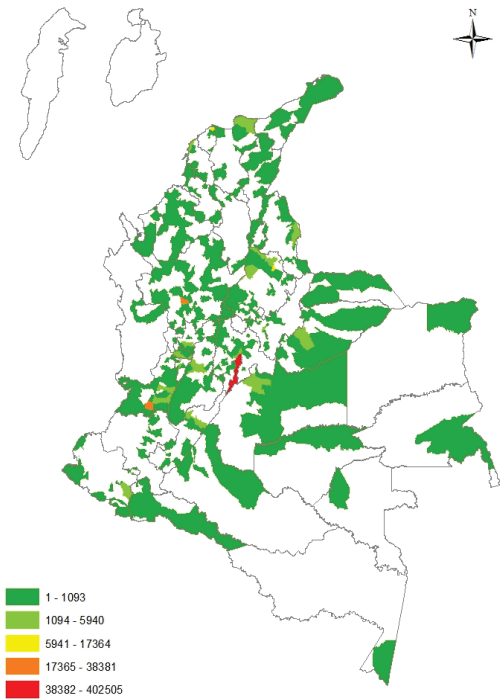
En la gráfica 3 se puede observar que las ofertas se concentran en las principales capitales del país, Bogotá (65.4%) (y Chía 0.39%), Medellín (6.2%), Cali (5.0%), Barranquilla (2.8%) y Bucaramanga (1.47%). En general, se tiene información de alrededor de 350 municipios.

Tabla 3

Departamento	Frecuencia	%
Bogotá	402,505	65.47
Antioquia	45,030	7.32
Valle	38,107	6.2
Atlántico	17,817	2.9
Santander	13,938	2.27
Cundinamarca	12,383	2.01
Meta	8,739	1.42
Bolívar	6,126	1
Risaralda	4,770	0.78
Tolima	3,310	0.54
Caldas	2,785	0.45
Huila	2,612	0.42
Nariño	2,536	0.41
Boyacá	2,417	0.39
Norte Santander	2,351	0.38
Casanare	1,991	0.32
Magdalena	1,982	0.32
Cesar	1,318	0.21
Quindío	1,317	0.21
Sucre	1,209	0.2
Córdoba	1,191	0.19
Cauca	1,092	0.18
Arauca	640	0.1
Putumayo	604	0.1
Caquetá	374	0.06
La Guajira	351	0.06
Choco	130	0.02
Amazonas	52	0.01
Guainía	33	0.01
Guaviare	18	0
Vichada	7	0
Vaupés	1	0
No reportado	37,096	6.03
Total	614,832	

Fuente: Ministerio del Trabajo. Cálculos propios

Gráfico 3:
Distribución municipal de las vacantes



Fuente: Ministerio del Trabajo. Cálculos propios

Por otra parte, la información de los salarios ofrecidos por puestos de trabajo se encuentra en rangos salariales. 50% de los puestos de trabajo ofrecen salarios entre \$550.001 y \$ 1.000.000, rango en el cual se incluye el salario mínimo. Seguido, 11.5% de los puestos de trabajo que ofrecen entre \$ 1.000.001 y \$ 1.500.000, mientras que solo 1.45% ofrece menos de \$ 550.000. Por otra parte, 29.5% de los puestos de trabajo no contiene información salarial o manifiesta estar sujeta a negociación, o se confirma en entrevista por lo cual no es posible identificar el salario ofrecido y quedan categorizadas como no reportan.

También es importante resaltar que por estos medios se ofrecen 457 trabajos que tienen un salario superior a los \$ 8.000.000.

Tabla 4

Rangos salariales	Frecuencia	%
Menos De \$550.000	8,919	1.45
\$ 550.001 - \$ 1.000.000	308,032	50.1
\$ 1.000.001 - \$ 1.500.000	71,139	11.57
\$ 1.500.001 - \$ 2.000.000	21,405	3.48
\$ 2.000.001 - \$ 2.500.000	9,059	1.47
\$ 2.500.001 - \$ 3.000.000	5,373	0.87
\$ 3.000.001 - \$ 3.500.000	3,495	0.57
\$ 3.500.001 - \$ 4.000.000	2,648	0.43
\$ 4.500.001 - \$ 5.500.000	1,357	0.22
\$ 5.500.001 - \$ 6.000.000	557	0.09
\$ 6.000.001 - \$ 8.000.000	840	0.14
\$ 8.000.001 y más	457	0.07
No reporta	181,551	29.53

Fuente: Ministerio del Trabajo. Cálculos propios

En la tabla 5 se puede observar que la experiencia laboral es un factor determinante para conseguir un empleo. Solo para 7.15% de los puestos de trabajo no es necesario tener experiencia alguna al aspirar al cargo, mientras que para 71.24% de los puestos de trabajo es necesario tener algo de experiencia laboral para al postularse. 21.6% de los puestos de trabajo no especifican en sus avisos si es necesario o no experiencia laboral para conseguir el trabajo.

Tabla 5

Requisito experiencia	Frecuencia	%
Sin experiencia	43,941	7.15
Experiencia	438,022	71.24
No reporta	132,869	21.61

Fuente: Ministerio del Trabajo. Cálculos propios

Al cruzar la información de salarios con los requerimientos de experiencia laboral se puede observar, en general, los cargos que exigen experiencia concentran más puestos de trabajo con mayores ingresos en comparación a las labores que no exigen experiencia laboral. Sin embargo, tanto los cargos que exigen o no experiencia, se concentran principalmente en salarios entre \$ 550.001 y \$ 1.000.000 mensuales (tabla 6).

Tabla 6

Rangos salariales	Sin experiencia	Experiencia
Menos De \$550.000	1.48	1.97
\$ 550.001 - \$ 1.000.000	81.94	70.59
\$ 1.000.001 - \$ 1.500.000	14.72	16.02
\$ 1.500.001 - \$ 2.000.000	1.22	5.57
\$ 2.000.001 - \$ 2.500.000	0.14	2.30
\$ 2.500.001 - \$ 3.000.000	0.04	1.40
\$ 3.000.001 - \$ 3.500.000	0.10	0.81
\$ 3.500.001 - \$ 4.000.000	0.00	0.70
Más de 4.500.001	0.35	0.64

Fuente: Ministerio del Trabajo. Cálculos propios

Ahora bien, del total de empleos que requieren experiencia y registran específicamente los años necesarios para aspirar al empleo, 48.3% requiere menos de un año de experiencia laboral mientras el 37.23% requiere al menos un año de experiencia (tabla 7).

Efectivamente, según se puede observar en la tabla 8, existe una correlación entre los años de experiencia requeridos y los salarios ofrecidos. 64.30% de los puestos de trabajo ofrecen un salario entre \$ 550.001 - \$ 1.000.000 y exigen un año de experiencia o menos. Mientras que los cargos que ofrecen más de \$ 5.500.000 requieren 5 años o más de experiencia laboral.

Tabla 7

Años de experiencia	Frecuencia	%
Menos de un año	118,415	48.34
1	91,203	37.23
2	19,442	7.94
3	6,092	2.49
4	1,471	0.60
5	3,301	1.35
6	4,058	1.66
7	249	0.10
8	213	0.09
9	21	0.01
10	404	0.16
Más de 10 años	85	0.03

Fuente: Ministerio del Trabajo. Cálculos propios

Con respecto al nivel educativo exigido para ingresar a determinado empleo, en la figura 4,⁵ se observa que 42.3% de los puestos de trabajo exigen como mínimo la terminación del bachillerato, 24.8% exigen técnico y 19.5% nivel tecnológico. Es importante resaltar que niveles inferiores al bachillerato no son muy demandados, en conjunto la demanda de estos niveles representa el 2.59% de los puestos de trabajo, mientras que los niveles de universitario y de posgrado suman 10.1%.

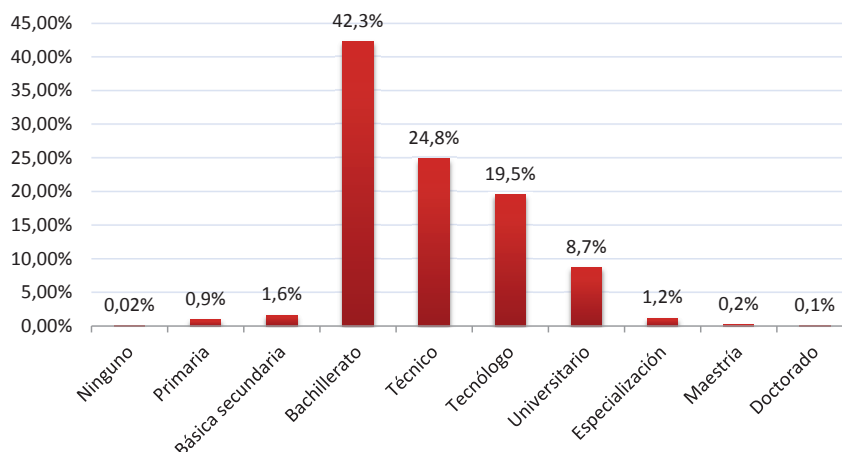
Tabla 8

Rangos salariales	Menos de un año	1	2	3	4	5 años o más
Menos De \$550.000	1.5%	0.5%	0.0%	0.0%	0.0%	0.0%
\$ 550.001 - \$ 1.000.000	38.6%	25.7%	2.4%	0.4%	0.1%	1.9%
\$ 1.000.001 - \$ 1.500.000	7.7%	7.4%	1.9%	0.5%	0.1%	0.3%
\$ 1.500.001 - \$ 2.000.000	1.3%	2.2%	1.1%	0.4%	0.1%	0.1%
\$ 2.000.001 - \$ 2.500.000	0.4%	0.7%	0.7%	0.4%	0.1%	0.1%
\$ 2.500.001 - \$ 3.000.000	0.2%	0.4%	0.4%	0.2%	0.0%	0.1%
\$ 3.000.001 - \$ 3.500.000	0.1%	0.5%	0.2%	0.1%	0.1%	0.1%
\$ 3.500.001 - \$ 4.000.000	0.0%	0.2%	0.1%	0.1%	0.1%	0.1%
\$ 4.500.001 - \$ 5.500.000	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%
\$ 5.500.001 - \$ 6.000.000	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
\$ 6.000.001 - \$ 8.000.000	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
\$ 8.000.001 y más	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%

Fuente: Ministerio del Trabajo. Cálculos propios

5 Cabe aclarar que la suma de los porcentajes no es exactamente el 100% ya que una misma vacante puede solicitar personas de diferentes niveles educativos para que realicen la misma labor.

Gráfico 4:
Distribución de los puestos de trabajo por nivel educativo



Fuente: Ministerio del Trabajo. Cálculos propios

También se evidencia en la tabla 8, que un mayor nivel educativo está asociado a un mayor nivel de salarios. Mientras que más de la mitad de los puestos de trabajo que exigen entre los niveles ninguno y tecnológico ofrecen salarios entre \$ 550.001 - \$ 1.000.000, aquellos puestos que exigen universitario o más se encuentran distribuidos en mayores niveles de salarios. Por ejemplo, 61.0% de los puestos de trabajo que piden nivel tecnológico ofrecen salarios entre \$ 550.001 - \$ 1.000.000 mientras que los cargos con nivel universitario es 32.0%.

En la tabla 9 se tiene que 85.58% de los puestos de trabajo no requieren específicamente un idioma para la postulación, mientras 11.53% requiere hablar español principalmente, ya que se necesitan personas para ser profesores de español o personas que hablen español fluidamente para atención al cliente. El segundo idioma más solicitado es inglés, para 2.5% (15,352) de los puestos de trabajo es necesario manejar este idioma mientras que en 1,029 empleos es fundamental saber francés.

Tabla 9

Rangos salariales	Ninguno	Primaria	Básica secundaria	Bachillerato	Técnico	Tecnólogo	Universitario	Especialización	Maestría	Doctorado
Menos De \$550.000	0.0%	0.9%	3.3%	2.5%	2.4%	1.9%	1.9%	0.4%	0.2%	0.0%
\$ 550.001 - \$ 1.000.000	89.9%	80.2%	84.0%	83.1%	69.6%	61.0%	32.0%	2.6%	9.4%	11.3%
\$ 1.000.001 - \$ 1.500.000	9.2%	16.0%	6.6%	12.0%	20.2%	26.8%	29.3%	4.3%	8.0%	4.2%
\$ 1.500.001 - \$ 2.000.000	0.0%	2.9%	2.7%	1.7%	4.5%	6.2%	15.0%	14.0%	9.9%	1.4%
\$ 2.000.001 - \$ 2.500.000	0.0%	0.0%	0.7%	0.2%	1.5%	2.0%	7.9%	21.8%	13.3%	2.8%
\$ 2.500.001 - \$ 3.000.000	0.0%	0.0%	0.0%	0.2%	1.0%	0.9%	4.4%	13.4%	20.3%	5.6%
\$ 3.000.001 - \$ 3.500.000	0.0%	0.0%	0.0%	0.2%	0.3%	0.4%	4.2%	8.7%	10.2%	2.8%
\$ 3.500.001 - \$ 4.000.000	0.0%	0.0%	0.0%	0.0%	0.3%	0.3%	2.9%	11.4%	8.5%	4.2%
\$ 4.500.001 - \$ 5.500.000	0.0%	0.0%	2.2%	0.0%	0.1%	0.2%	0.9%	9.4%	4.4%	2.8%
\$ 5.500.001 - \$ 6.000.000	0.0%	0.0%	0.5%	0.0%	0.0%	0.1%	0.4%	4.7%	4.6%	38.0%
\$ 6.000.001 - \$ 8.000.000	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.7%	5.2%	4.4%	4.2%
\$ 8.000.001 y más	0.9%	0.0%	0.0%	0.0%	0.1%	0.1%	0.4%	4.2%	6.8%	22.5%

Fuente: Ministerio del Trabajo. Cálculos propios

Tabla 10

Idioma	Frecuencia	%
Español	70,890	11.53
Inglés	15,352	2.5
Francés	1,029	0.17
Portugués	278	0.05
Mandarín	105	0.02
Holandés	66	0.01
Italiano	31	0.01
Alemán	18	0
Otro	866	0.14
No registra idioma	526,197	85.58

Fuente: Ministerio del Trabajo. Cálculos propios

Por otra parte, 81.18% de los puestos de trabajo no especifican si necesitan un hombre o una mujer para el cargo. Sin embargo, 3.46% (21,302) afirman que solo necesitan mujeres mientras que la demanda por hombres exclusivamente es mayor, 8.83% (54,283 puestos de trabajo). 6.53% afirman que son indiferentes entre hombres y mujeres.

Ahora bien, la mayoría de los anuncios de los puestos de trabajo (98.77%) no especifican preferencia por la edad de las personas pero cuando lo hacen, se requieren más

jóvenes (6,645 puestos de trabajo) que adultos (725 puestos de trabajo). Solo 162 empleos especifican que son indiferentes entre contratar jóvenes o adultos (tabla 12).

Tabla 11

Sexo	Frecuencia	%
Mujer	21,302	3.46
Ambos sexos	40,124	6.53
Hombre	54,283	8.83
No registra	499,123	81.18

Fuente: Ministerio del Trabajo. Cálculos propios

Tabla 12

Edad	Frecuencia	%
Joven	6,645	1.08
adulto	725	0.12
Ambos	162	0.03
No registra	607,300	98.77

Fuente: Ministerio del Trabajo. Cálculos propios

En la figura 5 se observa que principalmente se ofrecen contratos a término indefinido (100,932 puestos de trabajo) seguido de los contratos por obra o labor (59,796 puestos de trabajo) mientras que se ofrece el contrato de aprendizaje en 0.55% del total de empleos⁶.

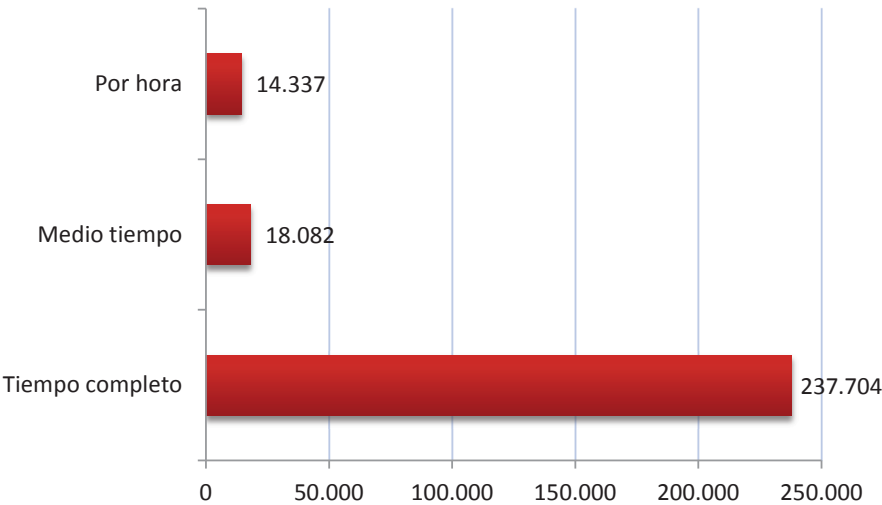
6 Es posible que un puesto de trabajo ofrezca diferentes tipos de contratación en ese caso se sumaron en cada una de las categorías correspondientes. Esto aplica también al aspecto de modalidad de trabajo.

Gráfico 5
Tipo de contrato



Fuente: Ministerio del Trabajo. Cálculos propios

Gráfico 6:
Modalidad de trabajo

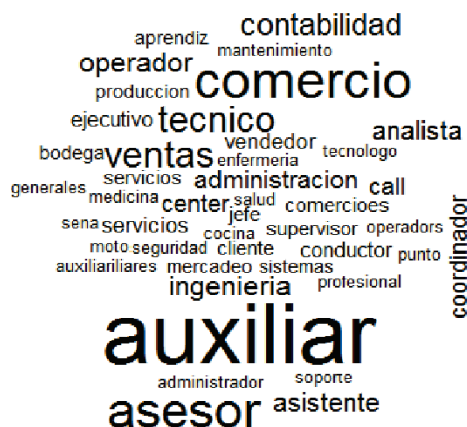


Fuente: Ministerio del Trabajo. Cálculos propios

Principalmente se requieren personas para trabajar tiempo completo (237,704 puestos de trabajo) mientras que los trabajos por hora solo representan el 2.33% del total de empleos ofrecidos.

A continuación se presenta el análisis de texto de los anuncios de las vacantes lo que permitirá reconocer cuales son las ocupaciones que más se requieren en el mercado laboral⁷. De esta forma se encontraron los cargos con mayor dinamismo actualmente en el mercado laboral, los cuales se puede observar en la figura 7. Los auxiliares y asesores son los cargos (grandes grupos ocupacionales) más demandados en el mercado laboral colombiano.

Figura 7
Cargos más demandados



Fuente: Ministerio del Trabajo. Cálculos propios

Más específicamente se puede ver en la figura 8 que el cargo más demandado entre los auxiliares son los auxiliares contables, seguido de auxiliares de bodega y servicios⁸. Mientras los asesores más demandados son los asesores comerciales, seguido de ventas y de call center.

Por último se presentan las estadísticas por sector⁹. En la figura 9 se tiene que el sector con mayor demanda personal es de consultorías/asesorías con el 75% del total de los puestos de trabajo que registran dicha información (248,794), seguido de Servicios (13.2%) y telecomunicaciones (4.1%).

De esta manera, se han presentado los principales indicadores que se pueden obtener de la base de vacantes que se compone de las diferentes fuentes de información ya una vez homogenizadas y estandarizada en un única bodega de datos. Como se ha mencionado previamente también hará parte de este sistema la base de datos de la Agencia del Empleo del SENA que permitirá estudiar más en profundidad la demanda por técnicos y tecnólogos.

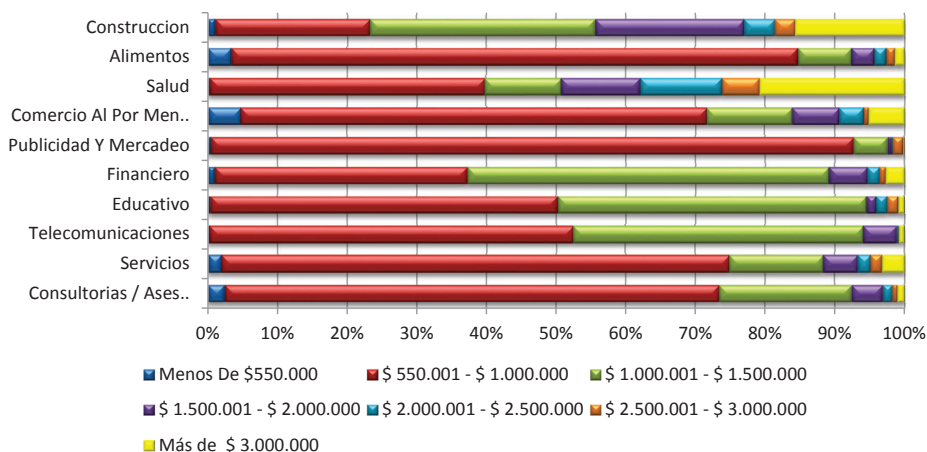
Además con la construcción de la bodega de datos en todo lo corrido del año se podrán realizar análisis dinámico de las ocupaciones, ya una vez clasificadas (O*NET y CIUO 08), que permitirá conocer cuales ocupaciones son las más demandadas a lo largo del tiempo, cuales han disminuido y cuales permanecen estáticas.

7 Esta información viene sin estructura, es decir, sin alguna clasificación de ocupaciones como la CIUO u O*NET, sino son directamente las palabras utilizadas por el empleador a la hora de colocar el anuncio. Próximamente ya se contará con esta información clasificada de acuerdo a la O*NET y la CIUO 08

8 Entre más gruesa sea la línea que conecta cada palabra con las demás más frecuente es su asociación

9 Cabe aclarar que esta variable sector corresponde solo a una de las fuentes de información y a sus categorías disponibles, ya que las demás tienen información dispersa sobre esta variable y no se ha estandarizado entre sí.

Figura 10
Distribución de salarios por sector



Fuente: Ministerio del Trabajo. Cálculos propios

Referencias Bibliográficas

- ALVAREZ, C., HOFFSTETTER, M. (2013), "Job Vacancies in Colombia: 1976-2012", *Borradores de Economía*, 797, Banco de la República.
- ARANGO, L. (2013), "Puestos de trabajo vacantes según anuncios de la prensa escrita de las siete principales ciudades de Colombia", *Borradores de Economía*, 793, Banco de la República.
- Arango, C. y 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.
- BARNICHON, R. (2009), "Building a Composite Help-Wanted Index", *Economics Letters*, 109.
- BHANDARI, I., COLET, E., PARKER, J., PINES, Z., PRATAP, R., RAMANUJAM, K. (1997), "Advanced Scout: Data Mining and Knowledge Discovery in NBA Data", *Data Mining and Knowledge Discovery*, 1(1).
- CÁRDENAS, M., BERNAL, R. (2001). *Determinants of labor demand in Colombia: 1976-1991*, Social Science Research Network Electronic Library.
- CARVALLO, O. (2011). Competencia, eficiencia y estabilidad financiera en la banca latinoamericana: Evidencia empírica 2001-2008. *Colección Economía y Finanzas*. Serie Documentos de Trabajo (121).
- DU, D., LI, A., ZHANG, L. (2014), "Survey on the Applications of Big Data in Chinese Real Estate Enterprise", *Procedia Computer Science*, 30.
- EDELMAN, B. (2012), "Using internet data for economic research", *Journal of Economic Perspectives*, 26(2).

- EMBLEY, D. W., CAMPBELL, D., JIANG, Y., LIDDLE, S., LONSDALE, D., NG, Y., SMITH, R. (1999), "Conceptual-model-based data extraction from multiple-record Web pages", *Data & Knowledge Engineering*, 31.
- ERMIDA, O., COLOTUZZO, N. (2009), *Descentralización, Tercerización, Subcontratación*, Organización Internacional del Trabajo.
- FAYYAD, U. (1996), Data Mining and Knowledge Discovery in Databases: Applications in Astronomy and Planetary Science, AAAI-96 Proceedings.
- 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.
- GINSBERG, J., MOHEBBI, M., PATEL, R., BRAMMER, L., SMOLINSKY, M., BRILLIANT, L. (2009), "Detecting influenza epidemics using search engine query data", *Nature*, 457 (19).
- GUATAQUÍ, J., FORERO, N., GARCÍA, A. (2009), "A quiénes afecta el desempleo? Análisis de la tasa de incidencia en Colombia", *Lecturas de Economía*, 70.
- GUOJIE L., XUEQI C., (2012) "Research Status and Scientific Thinking of Big Data", *Bulletin of Chinese Academy of Sciences*, 27(6).
- HSIEH, C., EDWARD, M., ORTEGA, D., RODRÍGUEZ, F. (2011). "The Price of Political Opposition: Evidence from Venezuela's Maisanta.", *American Economic Journal: Applied Economics* 3(2).
- Lacroix, Z. (2003), "Web data retrieval and extraction", *Data & Knowledge Engineering*, 44(3).
- MAIER-RIGAUD, F. P. (2005). "Switchings costs in retroactive rebates – What's time got to do with it?", *European Competition Law Review*, 26(5).
- MANYIKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C., HUNG, A. (2011), *Big Data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.
- MILLYMAKI, J. (2002), "Effective web data extraction with standard XML technologies", *Computer Networks*, 39(5).
- PARRA, M. (2012), *Trabajo temporal e indirecto: la pieza que faltaba para entender el mercado laboral colombiano*. Fedesarrollo.
- ROJAS, A. (2004), "La Intermediación Laboral", *Revista de Derecho*, 22. Universidad del Norte.
- REINERT, H., REINERT, E. (2006), "Creative Destruction in Economics: Nietzsche, Sombart, Schumpeter", en Backhaus, J. y Dreschsler, W. (Eds.), *Friedrich Nietzsche (1844-1900): Economy and Society*. Springer.
- UNIVERSIDAD EXTERNADO DE COLOMBIA (2010), *Situación de la Intermediación Laboral, Consideraciones y Expectativas en torno a un sistema nacional de intermediación laboral*, Banco Interamericano de Desarrollo, Banco Mundial.
- 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.

Anexo.

1. Ficha Resumen Encuesta Anual de Servicios

Contenido	La investigación obtiene su información a través de un formulario único que contiene dos módulos específicos para unas actividades económicas, y considera como unidad de estudio la empresa. Cuando se trata de empresas que desarrollan sus actividades en más de un sector económico (multi-actividad), se toma sólo la parte de los servicios investigados.
Casos	0
Variable(s)	303
Versión	2009
Productor	Dirección de Metodología y Producción Estadística - DIMPE.

Historia

La importancia creciente del sector de servicios en la economía colombiana plantea la necesidad de contar con información estadística sobre el tema para realizar el seguimiento y análisis de su comportamiento.

- En 1995 se dio inicio a la Encuesta Anual de Servicios (EAS) para las actividades de servicios de hoteles, restaurantes y agencias de viajes.
- En 1997 se incluyó la actividad de publicidad.
- En 1998, los servicios de obtención y suministro de personal y seguridad privada.
- En el 2000, las actividades de informática y correo.
- En 2003, los servicios de telecomunicaciones.
- En 2010 se rediseño el formulario de recolección, que se organizó en una estructura modular, con mayor detalle para el módulo de ingresos y gastos para un grupo específico de actividades económicas, y para el módulo de

personal ocupado de la actividad de educación superior.

Objetivo General

Conocer a través de la medición de valores absolutos de los principales agregados económicos (ingresos, producción bruta, consumo intermedio, valor agregado, personal ocupado, remuneraciones e inversión neta), la estructura y el comportamiento económico de las actividades de servicios investigadas a través de esta, determinadas con base en la clasificación CIU y el trabajo del Comité Interno de la EAS.

Tipo de estudio: Censo empresarial

Tipo de datos: Censo

Unidad de análisis

La unidad estadística es la empresa con NIT que, de manera exclusiva o predominante, se dedica a las actividades de servicios dentro del territorio nacional.

La unidad de observación y análisis la constituye la empresa que realiza actividades de servicios dentro del territorio nacional, de la cual existe y puede recopilarse información.

La Unidad de información es el contador, gerente, dueño o persona que maneja los balances de la empresa.

Universo de Estudio

Está conformado por empresas formalmente establecidas, residentes en el territorio nacional, cuya principal actividad es la prestación de servicios.

Población Objetivo

Está conformada por empresas formalmente establecidas, residentes en el territorio nacional, cuya principal actividad es la prestación de servicios conforme a la delimitación de la cobertura temática.

Productores y Patrocinadores

<i>Investigadores Primarios</i>	
Dirección de Metodología y Producción Estadística – DIMPE	Dependencia: Departamento Nacional de Estadística - DANE
<i>Otros Productores</i>	
Dirección de Metodología y Producción Estadística Encuesta Anual de Comercio	Dependencia: Departamento Nacional de Estadística - DANE
<i>Financiamiento</i>	
Departamento Nacional de Estadística – DANE	

Diseño de la encuesta

La EAS es un censo de grandes empresas y de acuerdo con los objetivos, por actividad de servicios se conformaron quince estratos. Los límites establecidos para cada uno de los estratos se encuentran en la Tabla 3. El universo se organizó en

estratos, caracterizados por ser homogéneos de acuerdo a los siguientes criterios: 1) de acuerdo con la principal actividad económica de la empresa, según el código CIIU Rev. 3 A.C.; 2) por ingresos anuales y cantidad de personal ocupado.

Tamaño de la encuesta

El tamaño de la encuesta se obtiene como la suma de los tamaños de todos los estratos de estudio. A su vez el tamaño de cada estrato es el resultado de sumar la cantidad de empresas de inclusión forzosa.

El tamaño es particular de cada año, dependiendo del mejoramiento del directorio y del comportamiento de la dinámica del sector. En 2010 el tamaño total de la EAS resultante fue de 6.748 empresas, de las cuales 5.819 fueron efectivas y 929 presentaron novedades (no pertenecen al alcance temático, están inactivas o sin localizar, entre otras). Un total de 5.343 empresas cumplieron con el alcance temático y con los parámetros de inclusión de la EAS en el año 2010, las cuales se distribuyeron en los quince estratos de estudio.

Recolección de datos

Fecha de recolección de datos

Inicio: 2012/05/15

Fin: 2012/10/31

Periodo: Anual

Modo de recolección de datos: Auto-diligenciamiento libre con formulario electrónico vía página Web.

Preguntas relacionadas con Mercado Laboral

NOMBRE	ETIQUETA	PREGUNTA
POMPSFR	Propietarios, socios y familiares sin remuneración mujeres	Propietarios, socios y familiares (sin remuneración fija) mujeres
POHPSFR	Propietarios, socios y familiares sin remuneración hombres	Propietarios, socios y familiares (sin remuneración fija) hombres
POTPSFR	Total propietarios socios y familiares sin remuneración	Total propietarios socios y familiares sin remuneración
POMPERM	Personal permanente mujeres	Personal permanente (contrato a término indefinido) mujeres
POHPERM	Personal permanente hombres	Personal permanente (contrato a término indefinido) hombres
POTPERM	Total personal permanente	Total personal permanente
POMTCDE	Personal temporal contratado directamente por la empresa mujeres	Personal temporal contratado directamente por la empresa mujeres
POHTCDE	Personal temporal contratado directamente por la empresa hombres	Personal temporal contratado directamente por la empresa hombres
POTTCDE	Total personal temporal directo	Total personal temporal directo
POMTMOE	Personal temporal en misión en otras empresas mujeres	Personal temporal en misión en otras empresas mujeres
POHTMOE	Personal temporal en misión en otras empresas hombres	Personal temporal en misión en otras empresas hombres
POT'TMOE	Total personal en misión	Total personal en misión
POMTCAG	Personal temporal suministrado por otras empresas mujeres	Personal temporal suministrado por otras empresas mujeres
POHTCAG	Personal temporal suministrado por otras empresas hombres	Personal temporal suministrado por otras empresas hombres
POTTCAG	Total personal temporal con otras empresas	Total personal temporal con agencias
POMPAU	Personal aprendiz o estudiantes por convenio mujeres	Personal aprendiz o estudiantes por convenio mujeres
POHPAU	Personal aprendiz o estudiantes por convenio hombres	Personal aprendiz o estudiantes por convenio hombres
POTPAU	Total personal aprendiz o estudiantes por convenio	Total personal aprendiz o estudiantes por convenio
POMTOT	Total de personal ocupado mujeres columna 1	Total de personal ocupado mujeres columna 1
POHTOT	Total de personal columna 2	Total de personal columna 2
POT'TOT	Total de personal ocupado columna 3	Total de personal ocupado columna 3
GPPSI	Salarios integral del personal permanente	Salario integral para el personal permanente
GPPER	Sueldos y salarios del personal permanente	Sueldos y salarios del personal permanente (en dinero y en especie, horas extras, dominicales, comisiones por ventas, viáticos permanentes.)

NOMBRE	ETIQUETA	PREGUNTA
GPPPSP	Prestaciones sociales del personal permanente comunes	Prestaciones sociales del personal permanente (vacaciones, primas legales extralegales, cesantías, intereses sobre cesantías)
GPPPST	Salarios y prestaciones del personal temporal contratado directamente por la empresa	Salarios y prestaciones del personal temporal contratado directamente por la empresa
GSPPT	Salarios y prestaciones del personal temporal en misión	Salarios y prestaciones del personal temporal en misión (solo para empresas especializadas en suministro de personal)
GPPCPO	Cotizaciones patronales obligatorias de salud, ARP y pensión del personal permanente y temporal directo, aprendices	Cotizaciones patronales obligatorias de salud, arp y pensión del personal permanente y temporal directo, temporal en misión y aprendices (excluye pensión para aprendices)
GPPANO	Aportes sobre nómina (Sena, cajas compensación fliar, ICBF)	Aportes sobre la nómina (Sena, cajas de compensación familiar, ICBF)
GPPAVO	Aportes voluntarios a compañías de seguro de vida o sistemas de salud prepagada	Aportes voluntarios a compañías de seguros de vida o de sistemas de salud prepagada
GPPGPA	Gastos causados por el personal aprendiz o estudiante por convenio (universitario, tecnólogo o técnico)	Gastos causados por el personal aprendiz o estudiante por convenio (universitario, tecnólogo o técnico)
GPPGPN	Otros gastos de personal	Otros gastos de personal no incluidos antes (incluya el auxilio de transporte) especifique en observaciones
GPPTPOT	Total gastos causados por el personal ocupado	Total gastos causados por el personal ocupado