



INSTITUTO POLITÉCNICO NACIONAL

**ESCUELA SUPERIOR DE INGENIERÍA MECÁNICA Y ELÉCTRICA
UNIDAD PROFESIONAL "ADOLFO LÓPEZ MATEOS"**

**"METODOLOGÍA PARA LA PREDICCIÓN DE PROBLEMÁTICAS
SOCIALES CON HERRAMIENTAS COMPUTACIONALES"**

TESIS

**QUE PARA OBTENER EL TÍTULO DE:
INGENIERO EN COMUNICACIONES Y ELECTRÓNICA**

PRESENTAN:

ESPINOSA ARREDONDO EVARISTO DE JESUS

ASESORES:

**Dra. María Elena Acevedo Mosqueda
Ing. Catalina Patiño Gallegos
M. en C. Roberto Galicia Galicia**



MÉXICO, D.F. DICIEMBRE 2015

Contenido

Capítulo 1	5
1.1 Resumen.....	5
1.2 Objetivos	6
1.2.1 Objetivo General	6
1.2.2 Objetivos Específicos	6
1.3 Problemática a Resolver.....	7
1.4 Justificación	8
1.4.1 Panorama de la niñez en el mundo y México	8
1.4.2 Pobreza y Trabajo Social	10
1.4.2 Representar problemáticas sociales con polinomios matemáticos.....	11
1.4.3 Elección de necesidades sociales	11
1.4.4 El beneficio de reducir la desigualdad	12
1.4.5 Estudio de comportamientos sociales con herramientas computacionales	13
1.4.6 Desarrollo tecnológico, social y emprendimientos sociales.	14
Capítulo 2	16
2.1 Antecedentes	16
2.2 Estado del Arte	19
2.2.1 OECD.....	19
2.2.3 Declara.....	19
2.2.4 Con Vos en la Web	19
2.2.5 UNICEF.....	20
2.2.6 Agencia Espacial Mexicana.....	20
2.2.7 Singularity University	20
Capítulo 3	22
3.1 Marco Teórico	22
3.1 Cualidades de la Información.....	22

3.2 Concepto de Sistema de Información	23
3.3 ¿Qué es una base de Datos?	24
3.4 Modelo de Base de Datos	25
3.5 Modelo Entidad Relación	27
3.6 Conjuntos de entidades	27
3.6 Conjuntos de relaciones de bases de datos	28
3.7 Correspondencia de cardinalidades	29
3.8 Reducción de un esquema E-R a tablas.....	30
3.9 Representación tabular de los conjuntos de entidades fuertes	31
3.10 Representación tabular de los conjuntos de entidades débiles	31
3.11 Representación tabular de los conjuntos de relaciones	32
3.12 Combinación de tablas	32
3.2 SQL.....	33
3.2.1 Creación y borrado de una base de datos relacional	34
3.2.2 Creación de tablas	35
3.2.3 Sentencias de manipulación.....	36
3.2.4 Subconsultas	38
3.3 Matlab	39
3.4 C#.....	40
Capítulo 4	41
4.1 Marco Metodológico.....	41
4.1.1 Introducción a la Metodología	41
4.1.2 Bases de Datos	42
4.2 Extracción de bases de datos	47
4.3 Base de Datos Final	48
4.4 Modelado Matemático.....	49
4.5 Modo de predicción	52
4.6 Interfaz	53
Capítulo 5	55
5.1 Caso de estudio: México	55
5.2 Resultados	56
5.3 Conclusiones.....	59
Bibliografía	61

Capítulo 1

1.1 Resumen

En esta tesis, se presenta una metodología para predecir problemáticas sociales, a partir de herramientas computacionales y bases de datos oficiales de la OCDE (1) y de la UNICEF (2).

Primeramente, se extraen las bases de datos de los exploradores de los organismos internacionales mencionados previamente para, posteriormente, modelar, con lenguaje MySQL y en función del modelo relacional, bases de datos de necesidades sociales de niños y adolescentes en México.

Después, estas bases son representadas en polinomios matemáticos y en gráficas tridimensionales, en donde cada eje representa un índice social extraído de las mismas.

Finalmente, se presenta un interfaz gráfica, desarrollada en lenguaje C#, sobre la plataforma Visual Studio, la cual, desde su fundamento, es simple en su semántica, diseño y programación, para que pueda ser comprendida y utilizada por cualquier trabajador social ajeno a las tecnologías de la información.

Así, se tiene, gracias a la aplicación de herramientas computacionales, tanto un panorama general de las problemáticas actuales como una predicción de problemáticas futuras en base a evidencia, desde la métrica de los índices estandarizados, de los derechos básicos de niños y adolescentes, para prevenirlas y mejorar las estrategias que las atienden.

De esta forma, se presentan elementos estadísticos e informáticos que profesionalicen la toma de decisiones basadas en evidencia dentro de los emprendimientos sociales y el trabajo social.

1.2 Objetivos

1.2.1 Objetivo General

Desarrollar una metodología que permita la predicción de problemática sociales en función de indicadores sociales con la ayuda de modelos de bases de datos e interfaces gráficas.

1.2.2 Objetivos Específicos

- Definir la metodología de forma general obtener el objetivo general buscado.
- Seleccionar un caso de uso que permita explotar la metodología.
- Detallar los pasos de la metodología en base a las experiencias obtenidas en la aplicación de la misma en el caso de uso seleccionado.

1.3 Problemática a Resolver

A partir del año 2006, cuando Muhammad Yunus ganó el Premio Nobel de la Paz, surgió dentro del ambiente empresarial la tendencia de los negocios sociales. Un negocio social, se enfoca en resolver una problemática social, preferentemente, desde el desarrollo de ciencia y tecnología, dando prioridad a la economía local. Al mismo tiempo, genera ganancias de forma responsable con la comunidad, gracias al trabajo de científicos, trabajadores sociales y emprendedores. Sin embargo, en muchas ocasiones, y desde mi experiencia como emprendedor social en República Checa, Argentina y Brasil, pude comprobar que una gran parte de este tipo de trabajo se realiza con pocas herramientas tecnológicas, y es difícil encontrar evidencias y datos que sustenten el trabajo. Además, cuando se buscan inversiones, al ser negocios de tipo social, es complicado mostrar hechos y cifras que fortalezcan las negociaciones con potenciales inversionistas.

Nos encontramos en plena época de Big Data. Las bases de datos en la red crecen de manera exponencial, siguiendo la Ley de Moore; no solo crece la cantidad de información, sino la complejidad para manejarla, interpretarla y usarla en la vida real. No obstante, la mayoría de las bases de datos nunca serán vistas por los seres humanos y, consecuentemente, no se transformarán en conocimiento.

Organismos internacionales como la OCDE y UNICEF, tienen enormes bases de datos de indicadores sociales, pero que pueden ser interpretadas y tratadas, en su mayoría, por personas que conozcan algún lenguaje para tratar bases de datos, como SQL, que existe en la mayoría de los manejadores incluyendo, MySQL. Por tanto, los emprendedores y trabajadores sociales, que son quienes realmente necesitan esa información y podrían transformarla en proyectos con valor agregado, se ven en dificultades para acceder al conocimiento estancado en interminables archivos.

Según cifras del Banco Mundial, en 2014, los emprendimientos sociales y ONG's experimentaron, alrededor del mundo, un flujo de inversión de más de 500 billones de Dólares. La OCDE, en 2014, tuvo un presupuesto de 396 millones de dólares. La UNICEF, en 2014, tuvo un presupuesto de 1.2 billones de dólares. Si los

emprendedores y trabajadores sociales tuviesen una plataforma digital fácil y simple de manejar, que les mostrara un panorama estadístico de temas sociales preciso y fácil de interpretar, podrían hacer su trabajo de forma más eficiente, y destinar los flujos de inversión en emprendimientos sociales, hacia los temas que realmente se tienen que resolver. Pero, prácticamente no existen plataformas simples y enfocadas a necesidades sociales específicas, la mayoría requieren conocimientos de matemáticas y bases de datos, que no son dominados por gran parte de los emprendedores sociales, haciendo que las bases de datos existentes en la red pierdan valor agregado, pues no pueden ser aprovechadas por los profesionales que realmente necesitan esa información.

1.4 Justificación

Para este trabajo, la metodología será aplicada en las problemáticas sociales de niños y adolescentes en México. Por tanto, es necesario, primero, comprobar por qué se quiere trabajar con este grupo demográfico.

1.4.1 Panorama de la niñez en el mundo y México

De acuerdo a la UNICEF, diariamente mueren 17,000 niños a causa de enfermedades que pudieron haber sido prevenidas. Cada 10 minutos, en algún lugar del mundo, una mujer adolescente muere por actos de violencia. La mitad de los niños menores de 5 años sufren de desnutrición.

Según las Naciones Unidas, de mantener los modelos antiguos de desarrollo, solamente 49% de los niños asistirán a la escuela. De acuerdo a la UNESCO, si todos los niños tuvieran el mismo acceso a la enseñanza, el ingreso per cápita aumentaría un 23% en los próximos 40 años, y si todas las mujeres cursaran la educación primaria, los matrimonios precoces y la mortalidad infantil podrían disminuir a una sexta parte y la mortalidad materna a dos tercios.

La UNICEF establece que *“la educación es un derecho humano fundamental y una herramienta decisiva para el desarrollo de las personas y las sociedades. Por ello, lograr la enseñanza primaria universal es el segundo de los Objetivos de Desarrollo del Milenio (ODM) de Naciones Unidas”*

Sin embargo, 57 millones de niños y niñas en todo el mundo no van a la escuela y otros tantos reciben una educación discontinua o deficiente (15 millones de niñas y

10 millones de niños) probablemente nunca pisarán un aula de mantenerse las tendencias actuales.

El 42% de los niños sin escolarizar vive en países pobres que padecen conflictos. En todo el mundo en desarrollo, solo 87 de cada 100 niños completa la educación primaria. Las tasas más bajas de escolarización primaria corresponden a África subsahariana, donde sólo el 65% de los niños y niñas en edad de estudiar acuden a la escuela. Hay al menos 250 millones de niños que no saben leer, escribir ni contar bien. 1 de cada 5 adolescentes no va a la escuela. Un 54% de los jóvenes que están fuera de la escuela son niñas. Las tasas más bajas de asistencia a Secundaria se registran en África oriental y meridional, donde la matriculación escolar de los varones es del 24% y la de las niñas del 22%. La educación secundaria es decisiva para la formación, el desarrollo y la protección de los adolescentes.

En México, 21.2 millones de niños, niñas y adolescentes (53.8 %) se encuentran en condición de pobreza, de los cuales 4.7 millones de ellos (11.9 %) viven en pobreza extrema. Además, 3 Millones de niños son explotados laboralmente.

Entender las causas de las problemáticas sociales desata airadas discusiones y polariza perspectivas, derivando en intereses económicos y políticos que condicionan las estrategias de trabajo para dar soluciones tangibles y reales. Los gobiernos no han sido capaces de transformar al mundo, pero la tecnología sí.

Es incuestionable e ineludible la necesidad de desarrollar ciencia y tecnología dentro del ecosistema empresarial actual; sin embargo, la narrativa convencional de estos avances obedece, generalmente, a necesidades mercantiles. La explosión en las últimas décadas de las tecnologías de la información ha transformado radicalmente las tendencias de desarrollo de negocios y ha impactado irremediablemente la vida de quienes tenemos acceso a sus beneficios.

Tanto el trabajo y emprendimiento social, como las políticas públicas y el desarrollo de responsabilidad social corporativa, han seguido una línea de trabajo basada en perspectivas, no en hechos.

1.4.2 Pobreza y Trabajo Social

De acuerdo a Marta Cubero, quien previamente había desarrollado una metodología para interpretar problemáticas sociales desde una metodología con técnicas de lógica difusa, para poder traducir dichas problemáticas en evidencia matemática que sirviera a las organizaciones sociales, que, para este trabajo, es el segmento del mercado laboral a quien le sirve esta metodología, “Las organizaciones sociales desarrollan acciones y servicios de interés general, construyendo cohesión e inclusión social, luchando contra la pobreza y discriminación, evitando que los colectivos más sensibles y más desfavorecidos de nuestra sociedad queden al margen de unos niveles mínimos de bienestar social. Resulta imposible pensar en el crecimiento sin inclusión y cohesión social.”

Dichas organizaciones,- continúa García Cubero- han ido creciendo a lo largo de los años de forma muy significativa, dando cada vez más calidad y profesionalidad a los servicios que prestan, generando innovación, creatividad y empleo de forma muy significativa. Precisamente, para poder seguir ofreciendo calidad y profesionalidad, es necesario cuidar la salud de los profesionales que diariamente interactúan o entran en contacto con la población más desfavorecida.

No hay que olvidar que los trabajadores de las organizaciones sociales desempeñan una labor enormemente útil y necesaria para la sociedad en la que vivimos, ya que suponen un gran caudal de cooperación humana en pro del desarrollo de ciertos grupos sociales menos favorecidos de nuestra sociedad. Dentro de estas organizaciones, existen situaciones concretas que pueden ser generadoras de riesgos laborales, que, para el enfoque de este trabajo, son:

- Existe poca evidencia matemática para sustentar el trabajo social que, a su vez, sea fácil de interpretar y entender por los trabajadores sociales.
- A pesar de los grandes flujos de inversión hacia proyectos sociales, es complicado conseguir patrocinios, pues dicha inversión no está, en muchas ocasiones, bien destinada hacia lo que, con evidencias, debe ser lo más importante.

1.4.2 Representar problemáticas sociales con polinomios matemáticos

En todos los trabajos científicos y técnicos, ya sean de investigación o de aplicación práctica, las medidas sin incertidumbre no tienen ningún valor, pero esto aún no se ha considerado en el ámbito de las ciencias sociales. Esta forma de pensar afecta a la credibilidad de los resultados, ya que la incertidumbre afecta a todo el proceso. Como mucho se calcula la incertidumbre derivada del tratamiento estadístico de las respuestas.

A través de polinomios se puede representar el conocimiento común en un lenguaje matemático, por lo que es posible trabajar a la vez con datos numéricos y términos lingüísticos.

1.4.3 Elección de necesidades sociales

Es necesario, para esta metodología, decidir y comprobar qué necesidades sociales se van a investigar, desde herramientas computacionales. Para comprobar por qué se trabajará en los factores sociales que se verán a lo largo de este trabajo, retomemos el trabajo de García Cubero, quien establece que “en medio de la década de mayor crecimiento, los elementos centrales del Desarrollo Social son: Pobreza, Desempleo, Integración Social. En la Declaración del Milenio se recogen ocho objetivos referentes a la erradicación de la pobreza, la educación primaria universal, la igualdad entre los géneros, la mortalidad infantil, materna, el avance VIH/SIDA, el sustento del medio ambiente y el fomento de una asociación mundial para el desarrollo.”

Desde la concepción del desarrollo social, la pobreza quedará entendida como el proceso antitético al desarrollo, y tiene lugar cuando las necesidades humanas de las personas se ven insatisfechas de forma involuntaria y prolongada. Aparece no solamente cuando la renta es insuficiente sino también cuando los individuos se ven privados de ciertos servicios sociales como pueden ser la educación, la salud, la participación social, y el mantenimiento de un nivel de vida igual al de sus semejantes.

Según la Declaración de Copenhague 95, sigue Cubero, se estableció como referencia que *“la base del desarrollo social es la ampliación de las capacidades y*

la mejora de las condiciones de vida de todas las personas, situando la satisfacción de sus necesidades básicas y de sus necesidades sociales, como ser social, en el centro de las políticas y actividades nacionales e internacionales". Menor pobreza, mayor participación, menos desempleo, mayor cohesión e integración social, y mayor equidad son algunos de los referentes básicos del modelo de desarrollo social propugnado en la Declaración de Copenhague.

Si consideramos la pobreza *"como un bajo nivel no sólo de ingresos y de consumo, sino también de instrucción, salud y nutrición, y de otras áreas del desarrollo humano"* (World Bank, 2000). En consecuencia, *"están en situación de pobreza las personas, familias y grupos, cuyos recursos económicos, sociales y culturales son tan limitados, que les excluyen del modo de vida que se considera aceptable en el país en que viven"*.

A pesar de la complejidad y el carácter dinámico de este fenómeno parece existir cierto consenso en que la exclusión social incluye dificultades o barreras en al menos tres grandes dimensiones; dimensión económica, dimensión política y dimensión social (relacional).

Esta nueva visión del concepto de exclusión social permite dar cuenta de tres aspectos claves: su origen estructural, su carácter multidimensional y su naturaleza procesual.

El estilo de desarrollo ha dejado al margen a un gran número de personas que, ante la competitividad y readaptación constante que impone el modelo social, se ven imposibilitadas de acceder a las oportunidades de que dispone el resto.

Analizar la exclusión social exige analizar una serie de procesos sociales que tienen la doble vertiente de procesos socio-económicos configuradores de la estructura social, y de procesos de erosión de las condiciones de integración social y, por ello, de generación de una sociedad excluyente.

1.4.4 El beneficio de reducir la desigualdad

La desigualdad se encuentra a la alza en los países de la OCDE, y sigue siendo muy alta en los países no miembros. Recientemente, el crecimiento económico era

considerado como el objetivo más importante en la mayoría de los países. Esto era cierto, pero no satisfacía todas las necesidades de su población. Una parte de la motivación para el énfasis dado al crecimiento económico, se apoyaba en la asunción de que el crecimiento económico llevaría, automáticamente, bienestar a la gente. De acuerdo a un estudio publicado por la OCDE, esto no es necesariamente cierto. Primero, el crecimiento económico hace crecer el bienestar económico solamente para un número limitado de personas en la población total. Segundo, el crecimiento económico no se traduce necesariamente en mejores resultados no económicos, como la salud o la educación, que son necesarios para que la gente pueda participar de manera plena en la economía y la sociedad. Por tanto, el cómo los países crecen, y la disyuntiva de traducir ese crecimiento en beneficios para sus ciudadanos, y lograr un crecimiento económico inclusivo, es un punto fundamental en las agendas de desarrollo social y científico de los países miembros de la OCDE. (OCDE, 2014)

En contraste, un crecimiento fuerte en los salarios de los habitantes de países en vías de desarrollo, ha llevado a incrementar la clase media convencional, definida para estos países como individuos que viven en casas con las necesidades básicas cubiertas. Sin embargo, la población en dichos países es muy vulnerable, debido a la proliferación del trabajo informal, el desempleo y el escaso servicio a servicios de salud y beneficios sociales. Todos estos factores, juntos, chocan en problemáticas sociales catastróficas, como las mostradas en los índices de crimen.

1.4.5 Estudio de comportamientos sociales con herramientas computacionales

En el MIT, el equipo dirigido por el Dr. Nathan Eagle había predecido comportamientos sociales esperando las variables en función de lo que arrojará su estudio.

En este trabajo, se consigue diferenciar cuatro necesidades sociales: educación, problemas sociales, acceso a servicios y ciencia y desarrollo, para contrastarlas y facilitar el trabajo del algoritmo. Además, se toma un fractal del universo poblacional: los niños y adolescentes.

Considerando esas necesidades sociales como categorías marginales, si se crea una metodología capaz de analizarlas con herramientas computacionales, además de emitir una recomendación y una predicción, se consigue crear una propuesta para incluir esas necesidades al ecosistema digital, además de crear una herramienta que cambie la forma de trabajar bajo percepciones, y el desarrollo social sea en base a evidencias.

De seguir bajo la tendencia actual donde el desarrollo social sin una predicción a futuro, estaríamos caminando a ciegas y creando una brecha entre el imparable desarrollo tecnológico y el necesario desarrollo humano. Si se crea una convergencia, se puede utilizar, en este caso, el desarrollo de software para comprobar cuáles serán las problemáticas futuras en función de los factores actuales, y así prevenirlas. Por ello el enfoque del estudio en el sector infantil y juvenil, para mejorar su ambiente de desarrollo.

En muchas instancias, la ciencia se está quedando detrás del mundo comercial en la habilidad de inferir significancia de una base de datos y tomar decisiones basadas en esa significancia. Sin embargo, el mundo comercial es comparativamente simple: las cosas que pueden ser descritas por algunos números o por un nombre, son manufacturadas para ser compradas y vendidas.

El tratamiento de base de datos, depende de tres actividades básicas: captura, curación y análisis.

1.4.6 Desarrollo tecnológico, social y emprendimientos sociales.

Existe una correlación directa y comprobada entre la exclusión social y la exclusión digital. Aquellos que son más desfavorecidos socialmente lo son también en cuanto al acceso a los recursos digitales; es siete veces menos probable que gocen los beneficios del ecosistema digital actual que aquellas personas que ven cubiertas sus necesidades básicas. Es decir, estamos ante un problema de exclusión dual. Lo mismo sucede con los negocios sociales. Los negocios sociales son emprendimientos enfocados en causas, en el cual los inversionistas pueden recuperar su fondo de riesgo, pero en el cual el objetivo principal de la compañía es puramente solucionar problemáticas sociales a través de la operación de la

empresa, desde el desarrollo de ciencia y tecnología, logrando hacer negocio de forma responsable. Muhammad Yunus, ganador del premio Nobel de la Paz en 2006, y creador de los emprendimientos sociales, mencionó que “el desarrollo de tecnología es importante, pero es lo que hacemos con ella lo que realmente importa. Busquemos cómo resolver problemas sociales con la tecnología”.

En un estudio publicado por el Oxford Internet Institute, se comprobó que más allá de dar acceso a internet, cuyo uso está condicionado por factores socioculturales que determinan la “decisión digital”, el uso de las TIC’s para un impacto social debe ser desde una plataforma de múltiples perspectivas. Esto quiere decir que, de no seguir tendencias tecnológicas, las políticas sociales serán exponencialmente ineficientes, mientras dentro de la sociedad de la información no se trabaje sobre la inclusión digital. En conclusión, mantener a los trabajadores y emprendedores sociales alejados de la tecnología, mantendrá la ineficiencia y crecimiento desigual de tecnología y desarrollo social. Además, tomando los estatutos que serán descritos en el marco teórico de este trabajo, se pretende evitar la polución informativa, y transformar en conocimiento las bases de datos manejadas en este proyecto. De nada sirve tener cantidades inmensas de datos en los exploradores, si los trabajadores y emprendedores sociales no tienen acceso a las mismas, o no saben cómo utilizarlas, siendo ellos quienes podrían transformar en resultados tangibles el conocimiento dichos datos. Es, evidentemente, totalmente necesario brindarles una herramienta que les acerque a los datos más importantes para su trabajo.

De acuerdo a un estudio publicado por el Centro para el Internet y Sociedad de la Universidad de Harvard, los patrones de exclusión social en el mundo fuera de línea, cuando se estudian en el ecosistema digital, se encuentran dentro de categorías marginales, con una cantidad de información enorme, pero aislada de la interacción y estudio dentro del internet. Si esas categorías se montan sobre un programa de predicción, se puede rehabilitar la perspectiva del uso de las TIC’s con propósitos sociales y mostrar, con evidencias y hechos científicos, por qué se debe analizar y predecir problemáticas sociales desde el desarrollo de software.

Tanto el trabajo y emprendimiento social, como las políticas públicas y el desarrollo de responsabilidad social corporativa, han seguido una línea de trabajo basada en perspectivas, no en hechos.

Capítulo 2

2.1 Antecedentes

El tratamiento de base de datos para desarrollos regionales se da dentro de un debate científico y empresarial recientemente surgido. Desde un punto epistemológico, los modelos comunes de bases de datos, no pueden trascender particularidades de patrones específicos, como en la categorización de necesidades sociales diferenciadas para un grupo demográfico específico, como los niños y adolescentes, objeto final del impacto de este trabajo de convergencia tecnológica y social. Sergio Conti, bajo la línea de investigación de la corriente surgida en el 2003, debatió que la categorización de sistemas locales eran específicos, irreversibles e irreproducibles; por lo tanto, no podían ser del todo sujetas a los paradigmas previos de clusters. Roel Rutten y Frans Boekema sugirieron, en esa misma investigación, una perspectiva de dimensión espacial, con un enfoque en el desarrollo social y económico en Italia. Esto permitió salir de un modelo empírico para contar con evidencias para la creación de algoritmos inteligentes de computación que se enfocaran en el desarrollo social.¹

El desarrollo de plataformas de software para atender e interpretar necesidades sociales específicas en base a factores demográficos, desde algoritmos de aprendizaje automatizado es, también, una narrativa tecnológica naciente, pero imparable.

Uno de los primeros antecedentes se remonta al 2007, en donde un equipo de investigadores del MIT, dirigidos por el Dr. Nathan Eagle², tuvo la iniciativa de desarrollar plataformas de software para interpretar y prevenir comportamientos sociales. En un artículo doctoral publicado en Springer-Verlag, identificaron y

¹ R.A. Boschma and R.C. Kloosterman (eds.), Learning from Clusters: A Critical Assessment, 1–15. 2005 Springer.

² MIT Media Laboratory, Massachusetts Institute of Technology, 20 AmesStreet, Cambridge, MA02139, USA.

predijeron con un 80% de exactitud los comportamientos sociales de un grupo demográfico específico, definiéndolos como eigencomportamientos e integrándolos sobre una plataforma móvil de software.³

En Mayo del 2010, Nathan Eagle, Michael Macy⁴ y Rob Claxton⁵ publicaron en la revista Science un artículo científico en donde estudiaron la relación de la estructura de las redes sociales con los patrones de marginación de los grupos sociales que no tienen acceso a las tecnologías de la información.⁶

En Septiembre de 2013, Amy Wesolowski⁷, Nathan Eagle⁸ y cinco investigadores internacionales más, publicaron un artículo científico en donde, a través de una aplicación de software, rastrearon y recolectaron datos digitales sobre los movimientos migratorios en Kenya, para identificar las relaciones entre los patrones demográficos de las regiones con grupos sociales más vulnerables a dejar sus lugares de origen por problemáticas sociales y los factores que lo provocan.⁹

Si bien la cantidad de investigadores trabajando para encontrar soluciones sociales a través de las TIC's¹⁰ ha incrementado en los últimos años, los resultados más sobresalientes y tangibles, en donde también se basan el resto de las investigaciones, son las dirigidas por el Dr. Nathan Eagle.

³ Eigenbehaviors: identifying structure in routine. Nathan Eagle & Alex Sandy Pentland. Springer-Verlag. September 2007. DOI 10.1007/s00265-009-0739-0

⁴ Department of Sociology, Cornell University, Ithaca, NY 14853, USA.

⁵ Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.

⁶ Network Diversity and Economic Development. Nathan Eagle et al. Science 328, 1029 (2010); DOI: 10.1126/science.1186605.

⁷ Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America.

⁸ Department of Computer Science, Northeastern University, Boston, Massachusetts, United States of America.

Department of Computer Science, Massachusetts Institute of Technology, Boston, Massachusetts, United States of America.

⁹ Wesolowski A, Buckee CO, Pindolia DK, Eagle N, Smith DL, et al. (2013) The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales. PLoS ONE 8(1): e52971. doi:10.1371/journal.pone.0052971

¹⁰ Tecnologías de la Información

En el MIT, el equipo dirigido por el Dr. Nathan Eagle había predecido comportamientos sociales esperando las variables en función de lo que arrojará su estudio.

En este trabajo, se consigue diferenciar dos necesidades sociales: educación y acceso a servicios, para contrastarlas y facilitar el trabajo del algoritmo. Además, se toma un fractal del universo poblacional: los niños y adolescentes.

Considerando esas necesidades sociales como clusters marginales, si se crea un algoritmo de aprendizaje automatizado capaz de interpretarlas inteligentemente, además de emitir una recomendación y una predicción, se consigue crear una propuesta para incluir esas necesidades al ecosistema digital, además de crear una herramienta que cambie la forma de trabajar bajo percepciones, y el desarrollo social sea en base a evidencias.

De seguir bajo la tendencia actual donde el desarrollo social sin una predicción a futuro, estaríamos caminando a ciegas y creando una brecha entre el imparable desarrollo tecnológico y el necesario desarrollo humano. Si se crea una convergencia, se puede utilizar, en este caso, el desarrollo de software para comprobar cuáles serán las problemáticas futuras en función de los factores actuales, y así prevenirlas. Por ello el enfoque del estudio en el sector infantil y juvenil, para mejorar su ambiente de desarrollo.

2.2 Estado del Arte

2.2.1 OECD

La OECD, a través de algoritmos de inteligencia artificial, desarrolló sobre su plataforma web un mapa interactivo de índices regionales de bienestar de los países miembros, en el que de manera visual, se pueden explorar los índices, por región, de las necesidades sociales que definen el bienestar: salud, seguridad, vivienda, acceso a servicios, compromiso civil, educación, trabajo, medio ambiente e ingreso.

En esta tesis, a diferencia del explorador de datos previamente mencionado, se consideran solamente cuatro necesidades: educación y acceso a servicios. Esto con la intención de contrastarlos y analizarlos automáticamente, y no solamente mostrar datos regionalizados, sino emitir, desde el software, una recomendación basada en evidencia demográfica y social para comprobar la razón de los índices existentes en cada región. Para esta interpretación, se consideran factores demográficos de niños y adolescentes, con la intención de brindar una herramienta que facilite el trabajo social para ese sector, y en el futuro prevenir situaciones de vulnerabilidad.

2.2.3 Declara

Declara es una compañía de tecnología basada en la región del Silicon Valley, en Palo Alto, California. Es una plataforma de crowdsourcing de soluciones sociales y conocimiento. Desde algoritmos de aprendizaje automatizado, generan una interacción de soluciones con los usuarios en función de su perfil, ofreciendo recomendaciones de contenido de calidad para facilitar el trabajo profesional, además de crear conocimiento colaborativo.

Nelson Gonzales, egresado de la Universidad de Berkeley y con un posgrado en la Universidad de Oxford, fundador de la compañía, ha sido mentor de este proyecto aún antes de ser presentado como propuesta oficial de tesis ante jurado y asesores.

2.2.4 Con Vos en la Web

Con Vos en la Web es un programa nacional del gobierno argentino, que surge de la necesidad de crear un espacio para la comunicación, la información, el asesoramiento y la participación de la población en temas vinculados con las

Nuevas Tecnologías de la Información y Comunicación (TIC), la protección de los datos personales, de la intimidad y de la privacidad. El programa se enfoca en niños y adolescentes, trabajando por la inclusión digital para que, después, se vea reflejado en la inclusión social.

Es fundamental saber qué contienen las bases de datos que se utilizarán. Al ser varias bases de datos considerablemente grandes, es imprescindible explorarlas y saber qué hay adentro, para, después, cuando a lo largo del desarrollo del proyecto titulación les dé una estructura y relación entre ellas, saber exactamente con qué se está trabajando.

2.2.5 UNICEF

La UNICEF tiene disponible un explorador abierto de datos sobre índices que miden los temas sociales entre los niños y adolescentes alrededor del mundo. Sin embargo, es posible solamente descargar los archivos, pero, a partir de ahí, se tienen que tratar con herramientas computacionales. Para alguien que domine, solamente, conceptos básicos del manejo de una computadora, es imposible transformar esos datos en conocimiento.

2.2.6 Agencia Espacial Mexicana

La AEM Utilizar la ciencia y la tecnología espacial para atender las necesidades de la población mexicana y generar empleos de alto valor agregado, impulsando la innovación y el desarrollo del sector espacial, contribuyendo a la competitividad y al posicionamiento de México en la comunidad internacional, en el uso pacífico, eficaz y responsable del espacio. En un acercamiento que se tuvo con el Director de Vinculación de la AEM, para presentar este proyecto, se mostró un interés particular en saber, con evidencias, cuáles son las necesidades sociales con más prioridad, pues, de otra forma, el desarrollo de tecnología no estaría atendiendo lo que la población mexicana necesita.

2.2.7 Singularity University

“Imaginen un híbrido entre la NASA y Teresa de Calcuta, y obtendrán Singularity University. Tecnología a lo bestia, inmensa ambición, el famoso think big estadounidense e idealismo y altruismo a raudales.” Juan Martínez-Barea

Esta institución académica, ubicada en el Silicon Valley, tiene como finalidad desarrollar proyectos científicos y tecnológicos que resuelvan las necesidades más importantes de la humanidad.

En un acercamiento que se tuvo con la representante de Singularity University en México, para pedir asesoría en este proyecto, se coincidió en la idea de que no es necesario desarrollar, dentro de las instituciones de educación superior, tecnología escandalosamente complicada y sofisticada, que al final queda en los laboratorios y no impacta ni el ambiente empresarial ni las vidas de las personas. En cambio, desarrollar ideas y métodos simples de tecnología, pero que tengan una propuesta de impacto tangible, es una de las tendencias dominantes entre muchos emprendedores.

Capítulo 3

3.1 Marco Teórico

3.1 Cualidades de la Información

Mario G. Piattini y Adoración de Miguel Castaño (1999, p. 6) sostienen que la explosión de la información – como se llama a veces a este enorme crecimiento de las necesidades de la información y a la mayor disponibilidad de este recurso – puede conducir, si no se ponen los medios para evitarlo, a una polución informativa. Fenómeno análogo al de la contaminación del aire, en el que la información, al perder sus cualidades, no puede cumplir sus objetivos, llegando incluso a ser más nociva que beneficiosa para sus destinatarios. Para evitar el peligro de la polución informativa se debe exigir a la información un conjunto de cualidades que mantengan su valor comunicativo, ya que para hacer honor a su nombre debe ser capaz de informar, es decir, de aportar un conocimiento.

Las cualidades que deben ser consideradas intrínsecas en la información, para que ésta logre su propósito fundamental son, fundamentalmente, tres:

1) Precisión

Se puede definir como el porcentaje de información correcta sobre la información total del sistema (dentro de una base de datos, por ejemplo). Si se quiere que los resultados del ordenador sean precisos, se deben suministrar datos precisos, sin pretender en los resultados una precisión superior a la que tenían los datos de entrada. Una precisión baja lleva a una falta de credibilidad del usuario hacia la información que se le proporciona.

2) Oportunidad

Se refiere al tiempo transcurrido desde el momento en que se produjo el hecho que originó el dato hasta el momento en el que la información se pone a disposición del usuario. La oportunidad depende de cada aplicación. En general, el valor de la información va disminuyendo con el transcurso del tiempo, e incluso, después de cierto momento, puede llegar a perder totalmente la relevancia que pudiera tener; la pérdida de valor será más o menos rápida dependiendo del tipo de información.

3) Compleción

Se refiere a que la información debe ser completa para cumplir sus fines. La compleción absoluta es imposible de conseguir, y lo que se suele pretender en los sistemas de información es alcanzar un nivel que se considere suficiente, el cual dependerá de dos factores: los datos existentes en el sistema de información y de los que el sistema es capaz de localizar ante una consulta concreta. Este segundo factor influirá la flexibilidad e idoneidad del lenguaje de recuperación, y el acierto en la formulación de la consulta. Así pues, la compleción no es sólo función de la información en sí misma, sino también de otros factores, tanto técnicos como humanos.

La información debe ser, también, significativa; es decir, ha de poseer el máximo contenido semántico posible, término que será tratado en el apartado posterior de este trabajo, ya que sin él la información perdería su rol fundamental de construir conocimiento. Se debe evitar proporcionar a los usuarios grandes masas de información que por su volumen no puedan ser asimiladas. Un volumen de información justo es condición indispensable para que ésta sea significativa. Cuando se diseña un sistema de información, es preciso tener en cuenta que la información suministrada debe ser, además de fácilmente interpretable, sólo la necesario y suficiente para que se cumplan los fines propuestos.

La información contenida en un sistema debe ser coherente en sí misma, además de consistente con las reglas semánticas propias del mundo real al que ha de representar lo más fielmente posible; esta cualidad, que en las bases de datos se conoce como integridad, coincide en parte con el concepto de precisión.

3.2 Concepto de Sistema de Información

Un sistema de información (SI) se diseña a fin de satisfacer las necesidades de información de una organización o una persona. El SI ha de tomar los datos del entorno y sus resultados han de ser la información que dicha organización necesita para su gestión y toma de decisiones.

Los modelos de datos soportados por los SGBD (Sistemas de Gestión de Bases de Datos) no suelen ofrecer, dado su bajo nivel de abstracción, los mecanismos

suficientes para captar la semántica del mundo real, por lo que surgen modelos conceptuales, más ricos semánticamente, que facilitan la labor del diseñador ayudándole en su comunicación con el usuario.

Entre estos modelos de datos semánticos, el modelo Entidad/Interrelación es posiblemente el de más amplia aceptación y al que se le suele dar soporte en la mayoría de las herramientas.

3.3 ¿Qué es una base de Datos?

Es necesario conocer la definición de una base de datos, para poder entender cómo es que las necesidades sociales descritas en los objetivos, podrían converger en un esquema de base de datos y, sobre todo, probar que las interacciones entre esas temáticas sociales, pueden representarse en los conceptos de bases de datos relacionables. De acuerdo a Jim Gray, autor de “The Fourth Paradigm” (El Cuarto Paradigma), un libro e investigador idolatrados por la comunidad de científicos de la computación: “Cuando la gente utiliza la palabra base de datos, fundamentalmente, lo que intentan decir, es que los datos deben describir una información y tener un esquema. Eso es lo que realmente significa una base de datos”, es decir, en mi entender de graduando de ingeniería, una base de datos es una colección de información organizada, para solventar una interpretación de datos eficiente.

En plena época del Big Data, donde está comprobado que en los últimos diez minutos, hemos generado más datos, que desde la prehistoria a antes de que iniciara la era de la información, parece necesario para este trabajo preguntar: ¿Qué problema resuelven las bases de datos? De acuerdo a Bill How, profesor e investigador de la Universidad de Washington, una base de datos tiene cuatro funcionalidades primordiales, especialmente hoy en día:

- 1) Que se pueda compartir con múltiples usuarios.
- 2) Asegurar que la aplicación tenga datos limpios y organizados.
- 3) Que sea escalable.
- 4) Que los datos se utilicen en formas disruptivas y no anticipadas.

3.4 Modelo de Base de Datos

También, para poder analizar las necesidades sociales propuestas para este trabajo, bajo la lógica de una base de datos, es necesario hacerlo desde una analogía con un modelo de datos. ¿Qué es un modelo de datos? De acuerdo al libro “Fundamentos y Modelos de Bases de Datos”, de Mario Piattini, un modelo de datos es un sistema de bases de datos, los datos, inicialmente, se organizan a fin de atender las necesidades de cada proceso; posteriormente, se intenta que respondan a un conjunto de procesos; por último, se busca una interpretación de la realidad, con el fin de captar su semántica. Para ellos, se aplica un Modelo de Datos (MD), que es un conjunto de conceptos que permiten describir, a distintos niveles de abstracción, la estructura de una base de datos, denominada, en este documento, como esquema. Este modelo puede ser externo, global o interno. Cada uno de ellos ofrece distintos elementos de descripción. Los modelos externos nos permiten representar los datos que necesita cada usuario en particular con las estructuras del lenguaje que utilice. Los modelos globales ayudan a describir los datos para el conjunto de usuarios, es decir, puede interpretarse como la información a nivel de una empresa; por último, los modelos internos, o físicos, están orientados a la máquina.

Para este trabajo, se enfocará en los modelos globales, los cuales, a su vez, se clasifican en conceptuales y convencionales. Los modelos conceptuales, también conocidos como de alto nivel, facilitan la descripción global del conjunto de información con independencia del SGBD (sistema de gestión de base de datos), lo que hace sus conceptos cercanos al mundo real (entidades, atributos, interrelaciones, etc...); es decir, son modelos de análisis, no de implementación. Los modelos convencionales, se encuentran soportados por los SGBD, y están orientados a describir los datos a nivel lógico para el SGBD (de ahí el nombre de modelos de bases de datos), sus conceptos son propios de cada SGBD (tablas o relaciones, en el caso del modelo relacional).

En pocas palabras, el modelo de datos es el instrumento que se aplica a los datos para obtener el esquema y su ocurrencia. Por esquema, entiéndase la descripción de la estructura de la base de datos; y, por ocurrencia del esquema, los datos que

se encuentran almacenados en el esquema en un determinado momento. El esquema, asociado al mundo real que describe, no varía mientras ese mundo no lo haga; en cambio, la ocurrencia del esquema, es decir, los datos contenidos en él son distintos en el transcurso del tiempo.

Es decir, un modelo de datos es un conjunto de conceptos, reglas y convenciones que nos permiten describir y manipular (consultar y actualizar) los datos, de una interpretación del mundo real que deseamos almacenar en la base.

En general, de acuerdo a George Djorgovski, investigador del Caltech, en su conferencia en el congreso científico “The Caltech-JPL Summer School on Big Data Analytics”, existen tres componentes esenciales para un modelo de datos, los cuáles se irán interpretando dentro de la propuesta de este proyecto, para proponer una aplicación a esa teoría:

1) Estructura

- a. Filas y columnas.
- b. Nodos y bordes, si es gráfico.
- c. Llaves primarias.
- d. Secuencias de bytes.

2) Reservas

- a. Las filas deben tener lógica con columnas.
- b. Todos los valores en una columna deben tener el mismo tipo.
- c. “Un niño no puede tener dos padres”. La analogía hecha por este investigador, se puede interpretar, por ejemplo, en que en mis bases de datos, México no puede tener, al mismo tiempo, 100 millones de dólares de inversión en educación en una fila y 150 millones en la otra. No tendría lógica. De ahí la importancia de presentar como trabajo de esta materia el análisis de lo que contiene cada base de datos.

3) Operadores

- a. Encontrar el valor dado una llave “x”. Por ejemplo, el ID de un país o un indicador social en la base de datos de la OECD.
- b. Encontrar las filas donde el valor es igual. Por ejemplo, encontrar los países que tienen ciertos niveles de educación de acuerdo a la prueba PISA.

3.5 Modelo Entidad Relación

El modelo de datos entidad-relación (E-R) está basado en una percepción del mundo real consistente en objetos básicos llamados entidades y de relaciones entre estos objetos. Se desarrolló para facilitar el diseño de bases de datos permitiendo la especificación de un esquema de la empresa que representa la estructura lógica completa de una base de datos. El modelo de datos E-R es uno de los diferentes modelos de datos semánticos; el aspecto semántico del modelo yace en la representación del significado de los datos. El modelo E-R es extremadamente útil para hacer corresponder los significados e interacciones de las empresas del mundo real con un esquema conceptual. Debido a esta utilidad, muchas herramientas de diseño de bases de datos se basan en los conceptos del modelo E-R.

3.6 Conjuntos de entidades

Una entidad es una «cosa» u «objeto» en el mundo real que es distinguible de todos los demás objetos. Por ejemplo, cada persona en un desarrollo es una entidad. Una entidad tiene un conjunto de propiedades, y los valores para algún conjunto de propiedades pueden identificar una entidad de forma unívoca. Por ejemplo, el D.N.I. 67.789.901 identifica unívocamente una persona particular en la empresa. Análogamente, se puede pensar en los préstamos bancarios como entidades, y un número de préstamo P-15 en la sucursal de Castellana identifica una entidad de préstamo. Una entidad puede ser concreta, como una persona o un libro, o puede ser abstracta, como un préstamo, unas vacaciones o un concepto.

Un conjunto de entidades es un conjunto de entidades del mismo tipo que comparten las mismas propiedades, o atributos. El conjunto de todas las personas que son clientes en un banco dado, por ejemplo, se pueden definir como el conjunto de entidades cliente. Análogamente, el conjunto de entidades préstamo podría representar el conjunto de todos los préstamos concedidos por un banco particular.

Una entidad se representa mediante un conjunto de atributos. Los atributos describen propiedades que posee cada miembro de un conjunto de entidades. La designación de un atributo para un conjunto de entidades expresa que la base de datos almacena información similar concerniente a cada entidad del conjunto de entidades; sin embargo, cada entidad puede tener su propio valor para cada atributo. Posibles atributos del conjunto de entidades cliente son id-cliente, nombre-cliente, calle-cliente y ciudad-cliente. En la vida real, habría más atributos, tales como el número de la calle, el número del portal, la provincia, el código postal, y la comunidad autónoma, pero no se incluyen en el ejemplo simple. Posibles atributos del conjunto de entidades préstamo son número-préstamo e importe. Cada entidad tiene un valor para cada uno de sus atributos. Por ejemplo, una entidad cliente en concreto puede tener el valor 32.112.312 para id-cliente, el valor Santos para nombre-cliente, el valor Mayor para callecliente y el valor Peguerinos para ciudad-cliente.

Para cada atributo hay un conjunto de valores permitidos, llamados el dominio, o el conjunto de valores, de ese atributo. El dominio del atributo nombrecliente podría ser el conjunto de todas las cadenas de texto de una cierta longitud. Análogamente, el dominio del atributo número-préstamo podría ser el conjunto de todas las cadenas de la forma «P-n», donde n es un entero positivo.

3.6 Conjuntos de relaciones de bases de datos

Una relación es una asociación entre diferentes entidades. Por ejemplo, se puede definir una relación que asocie al cliente López con el préstamo P-15. Esta relación especifica que López es un cliente con el préstamo número P-15. Un conjunto de relaciones es un conjunto de relaciones del mismo tipo. Formalmente es una relación matemática con $n \geq 2$ de conjuntos de entidades (posiblemente no distintos). Si E_1, E_2, \dots, E_n son conjuntos de entidades, entonces un conjunto de relaciones R es un subconjunto de:

$$\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$$

donde (e_1, e_2, \dots, e_n) es una relación.

La asociación entre conjuntos de entidades se conoce como participación; es decir, los conjuntos de entidades E_1, E_2, \dots, E_n participan en el conjunto de relaciones R . Un ejemplar de relación en un esquema E-R representa que existe una asociación entre las entidades denominadas en la empresa del mundo real que se modela. Como ilustración, el cliente individual López, que tiene D.N.I. 67.789.901, y la entidad préstamo P-15 participan en un ejemplar de relación de prestatario. Este ejemplar de relación representa que, en la empresa del mundo real, la persona llamada López cuyo número de D.N.I. es 67.789.901 ha tomado un préstamo que está numerado como P-15. La función que desempeña una entidad en una relación se llama papel de la entidad.

Debido a que los conjuntos de entidades que participan en un conjunto de relaciones son generalmente distintos, los papeles están implícitos y no se especifican normalmente. Sin embargo, son útiles cuando el significado de una relación necesita aclaración. Tal es el caso cuando los conjuntos de entidades de una relación no son distintos; es decir, el mismo conjunto de entidades participa en una relación más de una vez con diferentes papeles. En este tipo de conjunto de relaciones, que se llama algunas veces conjunto de relaciones recursivo, es necesario hacer explícitos los papeles para especificar cómo participa una entidad en un ejemplar de relación.

3.7 Correspondencia de cardinalidades

La correspondencia de cardinalidades, o razón de cardinalidad, expresa el número de entidades a las que otra entidad puede estar asociada vía un conjunto de relaciones. La correspondencia de cardinalidades es la más útil describiendo conjuntos de relaciones binarias, aunque ocasionalmente contribuye a la descripción de conjuntos de relaciones que implican más de dos conjuntos de entidades. Este apartado se centrará en conjuntos de relaciones binarias únicamente. Para un conjunto de relaciones binarias R entre los conjuntos de entidades A y B , la correspondencia de cardinalidades debe ser una de las siguientes:

- Uno a uno: Una entidad en A se asocia con a lo sumo una entidad en B, y una entidad en B se asocia con a lo sumo una entidad en A Propuesta Metodológica.
- Uno a varios. Una entidad en A se asocia con cualquier número de entidades en B (ninguna o varias). Una entidad en B, sin embargo, se puede asociar con a lo sumo una entidad en A.
- Varios a uno. Una entidad en A se asocia con a lo sumo una entidad en B. Una entidad en B, sin embargo, se puede asociar con cualquier número de entidades (ninguna o varias) en A.
- Varios a varios. Una entidad en A se asocia con cualquier número de entidades (ninguna o varias) en B, y una entidad en B se asocia con cualquier número de entidades (ninguna o varias) en A.

Figura de correspondencias uno a uno (a), y uno a varios (b).

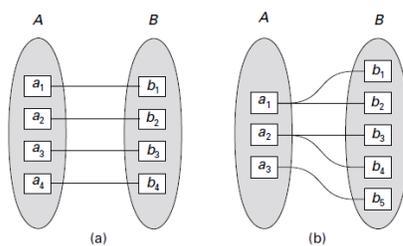
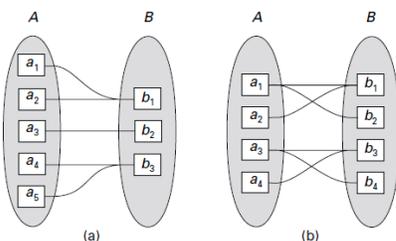


Figura de correspondencias varios a uno (a), y varios a varios (b).



3.8 Reducción de un esquema E-R a tablas

Una base de datos que se ajusta a un esquema de bases de datos E-R se puede representar por una colección de tablas. Para cada conjunto de entidades de la base de datos y para cada conjunto de relaciones de la base de datos hay una única tabla a la que se asigna el nombre del conjunto de entidades o del conjunto de relaciones correspondiente. Cada tabla tiene varias columnas, cada una de las cuales tiene un nombre único. Los modelos E-R y el de bases de datos relacionales son

representaciones abstractas y lógicas de empresas del mundo real. Debido a que los dos modelos emplean principios de diseño similares, se puede convertir un diseño E-R en un diseño relacional. Convertir una representación de bases de datos de un diagrama E-R a un formato de tablas es la base para la derivación de un diseño de bases de datos relacional desde un diagrama

E-R. Aunque existen diferencias importantes entre una relación y una tabla, una relación se puede considerar informalmente como una tabla de valores.

Las restricciones especificadas en un diagrama E-R, tales como las claves primarias y las restricciones de cardinalidad, se corresponden con restricciones sobre las tablas generadas a partir del diagrama E-R.

3.9 Representación tabular de los conjuntos de entidades fuertes

Sea E un conjunto de entidades fuertes con los atributos descriptivos a_1, a_2, \dots, a_n . Esta entidad se representa mediante una tabla llamada E con n columnas distintas, cada una de las cuales corresponde a uno de los atributos de E . Cada fila de la tabla corresponde a una entidad del conjunto de entidades E .

D_1 denota el conjunto de todos los números de préstamo y D_2 denota el conjunto de todos los saldos.

El conjunto de todas las filas posibles es el producto cartesiano de D_1 y D_2 , denotado por

$$D_1 \times D_2$$

En general, si se tiene una tabla de n columnas, se denota el producto cartesiano de D_1, D_2, \dots, D_n por $D_1 \times D_2 \times \dots \times D_{n-1} \times D_n$

3.10 Representación tabular de los conjuntos de entidades débiles

Sea A un conjunto de entidades débiles con los atributos a_1, a_2, \dots, a_m . Sea B el conjunto de entidades fuertes del que A depende. Sea la clave primaria de B el conjunto de atributos b_1, b_2, \dots, b_n . Se representa el conjunto de entidades A mediante una tabla llamada A con una columna por cada uno de los atributos del conjunto:

$$\{a_1, a_2, \dots, a_m\} \cup \{b_1, b_2, \dots, b_n\}$$

3.11 Representación tabular de los conjuntos de relaciones

Sea R un conjunto de relaciones, sean a_1, a_2, \dots, a_m el conjunto de atributos formados por la unión de las claves primarias de cada uno de los conjuntos de entidades que participan en R , y sean b_1, b_2, \dots, b_n los atributos descriptivos de R (si los hay). El conjunto de relaciones se representa mediante una tabla llamada R con una columna por cada uno de los atributos del conjunto:

$$\{a_1, a_2, \dots, a_m\} \cup \{b_1, b_2, \dots, b_n\}$$

3.12 Combinación de tablas

Considérese un conjunto AB de relaciones varios a uno del conjunto de entidades A al conjunto de entidades B . Usando el esquema de construcción de tablas descrito previamente se consiguen tres tablas: A , B y AB . Supóngase además que la participación de A en la relación es total; es decir, cada entidad a en el conjunto de entidades A debe participar en la relación AB . Entonces se pueden combinar las tablas A y AB para formar una única tabla consistente en la unión de las columnas de ambas tablas.

En el caso de relaciones uno a uno, la tabla del conjunto de relaciones se puede combinar con las tablas de cualquiera de los conjuntos de entidades. Las tablas se pueden combinar incluso si la participación es parcial usando valores nulos; en el ejemplo anterior se usarían valores nulos para el atributo nombre-sucursal para las cuentas que no tengan una sucursal asociada.

Las diferentes características del modelo E-R ofrecen al diseñador de bases de datos numerosas decisiones de cómo representar mejor la empresa que se modela. Los conceptos y objetos pueden, en ciertos casos, representarse mediante entidades, relaciones o atributos. Ciertos aspectos de la estructura global de la empresa se pueden describir mejor usando conjuntos de entidades débiles,

generalización, especialización o agregación. A menudo el diseñador debe sopesar las ventajas de un modelo simple y compacto frente a otros más precisos pero más completos.

Una base de datos que se representa en un diagrama E-R se puede representar mediante una colección de tablas. Para cada conjunto de entidades y para cada conjunto de relaciones de la base de datos hay una única tabla a la que se le asigna el nombre del conjunto de entidades o del conjunto de relaciones correspondiente. Cada tabla tiene un número de columnas, cada una de las cuales tiene un nombre único.

3.2 SQL

SQL (Structured Query Language) SQL (por sus siglas en inglés Structured Query Language) es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en ellas, manipularlas y controlarlas. Una de sus características es el manejo del álgebra y el cálculo relacional que permiten efectuar consultas con el fin de recuperar, de forma sencilla, información de bases de datos, así como hacer cambios en ellas. Es un lenguaje declarativo: sólo hay que indicar qué se quiere hacer. En cambio, en los lenguajes procedimentales es necesario especificar cómo hay que hacer cualquier acción sobre la base de datos. El SQL es un lenguaje muy parecido al lenguaje natural; concretamente, se parece al inglés, y es muy expresivo. Por estas razones, y como lenguaje estándar, el SQL es un lenguaje con el que se puede acceder a todos los sistemas relacionales comerciales.

Algunas de sus características son:

- Compatible con sistemas comerciales de bases de datos.
- Estandarizado y actualizado constantemente.
- Interactivo mediante GUI o consola.
- Basado en algebra relacional.

Para poder trabajar con bases de datos relacionales, lo primero que se tiene que hacer es definir las. Se describirán las órdenes de SQL para crear y borrar una base de datos relacional y para insertar, borrar y modificar las diferentes tablas que la componen.

La sencillez y la homogeneidad de SQL hacen que:

- 1) Para crear bases de datos, tablas, dominios, aserciones y vistas se utilice la sentencia CREATE.
- 2) Para modificar tablas y dominios se utilice la sentencia ALTER.
- 3) Para borrar bases de datos, tablas, dominios, aserciones y vistas se utilice la sentencia DROP.

La adecuación de estas sentencias a cada caso dará diferencias que se irán perfilando al hacer la descripción individual de cada una.

3.2.1 Creación y borrado de una base de datos relacional

El estándar SQL no dispone de ninguna sentencia de creación de bases de datos. La idea es que una base de datos no es más que un conjunto de tablas y, por lo tanto, las sentencias que ofrece el SQL se concentran en la creación, la modificación y el borrado de estas tablas.

En cambio, disponemos de una sentencia más potente que la de creación de bases de datos: la sentencia de creación de esquemas denominada CREATE SCHEMA. Con la creación de esquemas podemos agrupar un conjunto de elementos de la base de datos que son propiedad de un usuario. La sintaxis de esta sentencia es la que se tiene a continuación:

```
CREATE SCHEMA {[nombre_esquema]} | [AUTHORIZATION usuario]
[lista_de_elementos_del_esquema];
```

- Las palabras en mayúscula son palabras reservadas del lenguaje:

- La notación [...] quiere decir que lo que hay entre los corchetes se podría poner o no.
- La notación {A| ... |B} quiere decir que tenemos que elegir entre todas las opciones que hay entre las llaves, pero se debe poner una obligatoriamente.

La sentencia de creación de esquemas hace que varias tablas (lista_de_elementos_del_esquema)

Se puedan agrupar bajo un mismo nombre (nombre_esquema) y que tengan un propietario (usuario). Aunque todos los parámetros de la sentencia CREATE SCHEMA son opcionales, como mínimo se debe dar o bien el nombre del esquema, o bien el nombre del usuario propietario de la base de datos. Si sólo especificamos el usuario, éste será el nombre del esquema. La creación de esquemas puede hacer mucho más que agrupar tablas, porque lista_de_elementos_del_esquema puede, además de tablas, ser también dominios, vistas, privilegios y restricciones, entre otras cosas.

3.2.2 Creación de tablas

La estructura de almacenamiento de los datos del modelo relacional son las tablas. Para crear una tabla, es necesario utilizar la sentencia CREATE TABLE. Veamos su formato:

```
CREATE TABLE nombre_tabla ( definición_columna [, definición_columna...] [, restricciones_tabla] );
```

El proceso que hay que seguir para crear una tabla es el siguiente:

- 1) Lo primero que tenemos que se hace es decidir qué nombre se quiere asignar a la tabla (nombre_tabla).
- 2) Después, se da el nombre de cada uno de los atributos que formarán las columnas de la tabla (nombre_columna).

- 3) A cada una de las columnas se le asigna un tipo de datos predefinido o bien un dominio definido por el usuario. También se podrán dar definiciones por defecto y restricciones de columna.
- 4) Una vez definidas las columnas, se darán las restricciones de tabla.

3.2.3 Sentencias de manipulación

Una vez creada la base de datos con sus tablas, se puede insertar, modificar y borrar los valores de las filas de las tablas. Para poder hacer esto, el SQL ofrece las siguientes sentencias:

INSERT para insertar, UPDATE para modificar y DELETE para borrar. Una vez hemos insertado valores en las tablas, se puede consultar. La sentencia para hacer consultas a una base de datos con el SQL es SELECT FROM. Las sentencias se describen a continuación.

3.2.3.1 Inserción de filas en una tabla

Antes de poder consultar los datos de una base de datos, es preciso introducirlos con la sentencia INSERT INTO VALUES, que tiene el formato:

```
INSERT INTO nombre_tabla [(columnas)] {VALUES ({v1|DEFAULT|NULL}, ...,
{vn/DEFAULT/NULL})};
```

Los valores v1, v2, ..., vn se deben corresponder exactamente con las columnas que se han dicho que se tendrían con el CREATE TABLE y deben estar en el mismo orden, a menos que se vuelvan a poner a continuación del nombre de la tabla. En este último caso, los valores se deben disponer de forma coherente con el nuevo orden que hemos impuesto. Podría darse el caso de que se quisiera que algunos valores para insertar fuesen valores por omisión, definidos previamente con la opción DEFAULT. Entonces se podría usar la palabra reservada DEFAULT. Si se trata de introducir valores nulos, también podemos utilizar la palabra reservada NULL.

Para borrar valores de algunas filas de una tabla podemos utilizar la sentencia DELETE FROM WHERE. Su formato es el siguiente:

```
DELETE FROM nombre_tabla [WHERE condiciones];
```

3.2.3.2 Modificación de filas de una tabla

Si se quisiera modificar los valores de algunas filas de una tabla, se tendría que utilizar la sentencia UPDATE SET WHERE. A continuación se presenta su formato:

```
UPDATE nombre_tabla SET columna = {expresión|DEFAULT|NULL} [, columna = expr|DEFAULT|NULL} ...] WHERE condiciones;
```

Funciones de agregación

SQL ofrece las siguientes funciones de agregación para efectuar varias operaciones sobre los datos de una base de datos:

Función	Descripción
COUNT	Nos da el número total de filas seleccionadas
SUM	Suma los valores de una columna
MIN	Nos da el valor mínimo de una columna
MAX	Nos da el valor máximo de una columna
AVG	Calcula el valor medio de una columna

En general, las funciones de agregación se aplican a una columna, excepto la función de agregación COUNT, que normalmente se aplica a todas las columnas de la tabla o tablas seleccionadas. Por lo tanto, COUNT (*) contará todas las filas de la tabla o las tablas que cumplan las condiciones. Si se utilizare COUNT(distinct

columna), sólo contaría los valores que no fuesen nulos ni repetidos, y si se utilizase COUNT (columna), sólo contaría los valores que no fuesen nulos.

3.2.4 Subconsultas

Una subconsulta es una consulta incluida dentro de una cláusula WHERE o HAVING de otra consulta. En ocasiones, para expresar ciertas condiciones no hay más remedio que obtener el valor que buscamos como resultado de una consulta.

3.2.4.1 Predicado BETWEEN

Para expresar una condición que quiere encontrar un valor entre unos límites concretos, se puede utilizar BETWEEN:

```
SELECT nombre_columnas_a_seleccionar
FROM tabla_a_consultar
WHERE columna
BETWEEN límite1 AND límite2;
```

3.2.4.2 Predicado IN

Para comprobar si un valor coincide con los elementos de una lista se utilizará IN, y para ver si no coincide, NOT IN:

```
SELECT nombre_columnas_a_seleccionar
FROM tabla_a_consultar
WHERE columna [NOT] IN (valor1, ..., valorN);
```

3.2.4.3 Consultas con agrupación de filas de una tabla

Las cláusulas siguientes, añadidas a la instrucción SELECT FROM, permiten organizar las filas por grupos:

La cláusula GROUP BY sirve para agrupar filas según las columnas que indique esta cláusula.

La cláusula HAVING especifica condiciones de búsqueda para grupos de filas; lleva a cabo la misma función que antes cumplía la cláusula WHERE para las filas de toda la tabla, pero ahora las condiciones se aplican a los grupos obtenidos.

Presenta el siguiente formato:

```
SELECT nombre_columnas_a seleccionar
FROM tabla_a_consultar [WHERE condiciones]
GROUP BY columnas_según_las_cuales_se_quiere_agrupar
[HAVING condiciones_por_grupos]
[ORDER BY columna_ordenación [DESC] [, columna [DESC]...]];
```

Se puede percibir que en las sentencias SQL se van añadiendo cláusulas a medida que la dificultad o la exigencia de la consulta que lo requiere.

3.3 Matlab

Matlab es el lenguaje de alto nivel y con un entorno interactivo usado por millones de ingenieros y científicos alrededor del mundo. Permite explorar y visualizar ideas y colaborar a través de distintas disciplinas, incluyendo procesamiento de señales e imágenes, comunicaciones, sistemas de control, bases de datos y finanzas computacionales.

Usando Matlab, se pueden analizar bases de datos, desarrollar algoritmos y crear modelos y aplicaciones.

El lenguaje, herramientas y funciones permiten explorar múltiples acercamientos y lograr una solución a problemas de ingeniería, más rápida que hojas de datos o lenguajes de programación tradicionales, como C/C++ o Java. Más de un millón de ingenieros y científicos, tanto en la industria como en la investigación académica, usan MATLAB.

MATLAB ofrece un rango de métodos numéricos para analizar bases de datos, desarrollo de algoritmos y creación de modelos. Incluye funciones matemáticas que soportan operaciones comunes de ciencia e ingeniería. Las funciones matemáticas principales usan librerías que ofrecen una ejecución rápida de operaciones y cálculos.

MATLAB ofrece herramientas para adquirir, analizar y visualizar bases de datos, permitiendo ganar perspectiva sobre las bases de datos en una fracción de tiempo de la que tomaría usando hojas de datos o lenguajes de programación tradicionales. Se puede, también, documentar y publicar los resultados de las gráficas usando código de MATLAB.

MATLAB permite acceder a datos desde archivos, otras aplicaciones, bases de datos y dispositivos externos. Se pueden leer datos desde formatos populares como Microsoft Excel; texto plano o archivos binarios; imagen, sonido, archivos de videos; y archivos científicos como netCDF y HDF. Funciones de entrada y salida de funciones permiten trabajar con datos en cualquier formato.

Usando MATLAB, se puede adquirir data de dispositivos de hardware, como un puerto serial de la computadora, así como transmisiones en vivo, y bases de datos directamente cargada a MATLAB para su análisis y visualización.

Importando archivos de bases de datos a MATLAB, se generan automáticamente una función para importar, automáticamente, el programa.

3.4 C#

C# es un lenguaje de programación orientado a objetos desarrollado y estandarizado por Microsoft como parte de su plataforma .NET.

Su sintaxis básica deriva de C/C++ y utiliza el modelo de objetos de la plataforma .NET, similar al de Java, aunque incluye mejoras derivadas de otros lenguajes.

C# es un lenguaje de programación moderno que combina la potencialidad de C, C++ y Java, con la agilidad y la velocidad para el desarrollo propios de Visual Basic. Sin dudas, es el lenguaje más importante de la plataforma .NET. Fue diseñado para

crear sistemas en el mundo actual, donde la necesidad de comunicarse con otras aplicaciones y servicios mediante una red de información –como Internet– es moneda corriente, donde conviven bases de datos de distintos fabricantes, donde los sistemas son construidos a partir de componentes locales o remotos, y donde se requieren soluciones integrales para diversos tipos de plataformas. Con C# podremos crear sistemas de una gran diversidad, simples o complejos; aplicaciones de consola, de escritorio o para la Web; programas para computadoras personales o para dispositivos móviles. C# La guía del programador. Diego Ruiza

Aunque C# forma parte de la plataforma .NET, ésta es una API, mientras que C# es un lenguaje de programación independiente diseñado para generar programas sobre dicha plataforma. Ya existe un compilador implementado que provee el marco Mono - DotGNU, el cual genera programas para distintas plataformas como Windows, Unix, Android, iOS, Windows Phone, Mac OS y GNU/Linux

Capítulo 4

4.1 Marco Metodológico

4.1.1 Introducción a la Metodología

De acuerdo a lo que se ha expuesto en los capítulos anteriores, y de acuerdo a la experiencia obtenida, la metodología computacional para predecir problemáticas sociales es la que se describirá a continuación.

Basándonos en el trabajo de investigación hecho por la investigadora Marta Cubero García, lo que se busca mediante el empleo, en el caso de este proyecto, de herramientas computacionales, es describir y formalizar la realidad, empleando modelos flexibles que interpreten las leyes que rigen el comportamiento humano y las relaciones entre los hombres.

Para describir esa realidad incierta, tanto en el orden de lo social como en el de lo natural, se utilizarán polinomios matemáticos.

La condición humana está marcada, afirma Cubero, por dos grandes incertidumbres: la incertidumbre cognitiva y la incertidumbre histórica. Conocer y

pensar no es llegar a una verdad totalmente cierta, es dialogar con la incertidumbre. Esa incertidumbre puede representarse con una fórmula matemática.

Si bien, en este proyecto, se parte de bases de datos, se pueden representar en fórmulas, apoyándonos en la idea de Cubero, que afirma que “el tratamiento tradicional de los datos no utiliza conceptos complejos como el de la incertidumbre, que aplicado al tratamiento informático de los datos muestra unas representaciones lineales objetivas para comprender las dispersiones y los agrupamientos mostrados por los diagramas de barras de los gráficos obtenidos con el método tradicional. Los investigadores buscan nuevos campos de aplicación de esta técnica” (La investigación en Trabajo Social con técnicas de lógica difusa, 2010)

También, como analizando uno de los artículos citados en la justificación de este trabajo, se puede percibir que el autor identificó varios comportamientos sociales, que después consideró como variables dentro de una ecuación matemática, para analizar los movimientos en los índices de crimen, definidos en una variable x , en función del resto de las variables. Sin embargo, como se menciona en la justificación, requería de un conocimiento avanzado de matemáticas.

Esta metodología es utilizada para describir la realidad social en fórmulas matemáticas, como , pero desde herramientas computacionales y el tratamiento de base de datos, como ; ambos conceptos, serán fácilmente leídos e interpretados por cualquier persona en una interfaz de software simple de entender, mejorando la desventaja que tiene, en el caso de , utilizar términos sobre especializados, reduciendo el valor agregado de las bases de datos de índices sociales, perdiendo uno de los principios de la información establecidos por Piatti.

4.1.2 Bases de Datos

La OCDE, siguiendo su línea de trabajo, que establece la imperiosa necesidad de cuantificar los fenómenos sociales, para poder trabajar en ellos de forma profesional y científica, tiene disponible un explorador de bases de datos, en dos modalidades distintas:

- a) Open Data: Base de datos enteras, pero, solamente de algunos de los temas que la OCDE trabaja. Sin embargo, en esta modalidad, no están disponibles

todas las bases de datos, artículos, publicaciones y recomendaciones de la organización. (OCDE)

- b) Explorador Privado: Acceso, en su totalidad, a todas las bases de datos de la organización. Numerosos artículos, publicaciones, recomendaciones y, evidentemente, bases de datos, son accedidas, solamente, desde esta modalidad. El precio de las bases de datos, según su utilidad, varía entre \$20 y \$90 dólares por archivo, debido a que son archivos que fueron auditados bajo los más rigurosos estándares internacionales y son, en consecuencia, los más confiables y utilizados por científicos de datos e investigadores de temas que la OCDE trabaja. (I-Library)

Para el desarrollo de este proyecto, se consiguió un patrocinio de la OCDE, gestionado por la editorial Innovación Educativa durante los meses de Abril y Mayo de 2015, que proporcionó al graduando una cuenta, con acceso ilimitado a su explorador y biblioteca digital, durante dos meses: Julio y Agosto de 2015.

4.1.2.1 Exploración de datos

El explorador de bases de datos está organizado en 20 categorías, cada una con otro número de subcategorías; cada categoría y subcategoría tiene una gran cantidad de archivos de bases de datos, publicaciones e investigaciones. Si bien este explorador es una prueba fehaciente del fenómeno del Big Data, descrito en este trabajo, permitía al usuario consultar su contenido a través de “Queries” (consultas de bases de datos), para encontrar más fácilmente la información buscada.

Explorar cada categoría demoraba, en promedio, 8 horas; al ser bases de datos grandes, las consultas para acceder a la información, permitidas por el explorador, demoraban, para terminar de procesarlas, mucho más que las trabajadas durante los cursos de asignaturas como “Lenguajes de Internet” e “Ingeniería de Software”. Adicional, los archivos, al ser muy grandes, requerían un tiempo considerable de descarga.

Ante esta situación, y considerando que el tiempo de acceso al explorador era, solamente, de dos meses, siguiendo las comprobaciones de la justificación de este

trabajo, se decidió trabajar sobre los siguientes temas, que en el explorador estaban organizados, hasta el 31 de Agosto del 2015, de esta manera:

4.1.2.2 Estadísticas Generales.

Estadísticas de cada uno de los países, que representan los indicadores de temas demográficos, económicos, ambientales y sociales, pero, solamente, de forma general, sin mayor rigor.

- Perfil estadístico de los países: contenía 32 bases de datos por cada país de la OCDE.
- Indicadores económicos de corto plazo: Por cada país de la OCDE, se tienen 10 indicadores; cada uno, representando entre 2 y 6 unidades de medición. Consecuentemente, a través de las consultas que el explorador permitía, se podían explorar entre 20 y 60 bases de datos por país.
- Factbook 2014: Es el más utilizado a nivel global, especialmente por agencias de noticias, para realizar análisis de los países. Por cada indicador, se tiene una base de datos por país (34), tiene más de 80 indicadores, teniendo, en consecuencia, más de 80 bases de datos por país.

4.1.2.2.1 Población y Demografía

Índices de las características sociales, económicas, educativas y demográficas de cada uno de los países miembros de la OCDE.

- Estadísticas de Migración: Tiene otras 10 subcategorías más. Todas las subcategorías suman un total de más de 150 unidades de indicadores. Por tanto, por cada país de la OCDE (34) se podían tratar más de 150 bases de datos.
- Estadísticas de Población: Tiene 3 subcategorías más que, al tratar las bases de datos de este tema en función a una cronología histórica, desde 1961 hasta 2014, tenía 35 indicadores de medición por año, y por país de la OCDE (34).

4.1.2.2.2 Desarrollo Social

Son, básicamente, indicadores sobre los flujos de inversión destinados a solucionar los problemas primordiales de la población, alrededor de los países miembros de la organización.

- Flujos de inversión por país benefactor: Tiene 3 subcategorías. Cada una de ellas, tenía más de 40 indicadores por país (34). En consecuencia, se tenían más de 120 bases de datos para explorar por país.
- Flujos de inversión por país beneficiado. Tiene 14 subcategorías. Cada una de ellas, con más de 45 indicadores por país (34). En consecuencia, aquí se tenían 630 bases de datos por país. Esto, considerando el promedio de un rango de años, y no todos y cada uno de los años, lo que multiplicaría 53 veces las bases de datos.
- Flujos de inversión por proyecto individual de cada país. Tiene 4 subcategorías. Cada una, mostraba los flujos de inversión hacia proyectos en cada país, teniendo 81 sectores de inversión, alojados en bases de datos. Estos flujos están considerados en países no miembros de la OCDE, y en los miembros menos favorecidos de la misma (México, Chile y Sudáfrica). En total, eran 72 países disponibles. Como se puede observar, se tenían más de 360 bases de datos por cada país.

4.1.2.2.2 Educación

Indicadores de desempeño académico, infraestructura, acceso a educación y cuerpo docente.

- Una mirada a la educación: 11 indicadores, cada uno, tenía entre 4 y 17 sectores de medición, por cada país, y dividido entre 4 sectores de edad (niños, adolescentes, jóvenes y adultos). En consecuencia, por cada país, se tenían 187 bases de datos.
- Educación y Habilidades: 23 indicadores, cada uno, tenía entre 4 y 9 sectores de medición, por cada país, y dividido entre 4 sectores de edad (niños, adolescentes, jóvenes y adultos). En consecuencia, por cada país, se tenían, aproximadamente, 207 bases de datos.

- Sondeo Internacional de Enseñanza y Aprendizaje: 3 bases de datos sobre el sondeo.
- Medición de la Innovación en la Educación: 2 indicadores, cada uno, tenía 5 sectores de medición; es decir, se tenían 10 bases de datos por cada país.

4.1.2.2.2 Ciencia y Tecnología

Indicadores precisos sobre el desarrollo de ciencia, tecnología y patentes en cada uno de los países de la OCDE, así como su desempeño e impacto en el desarrollo económico, social y ambiental de los países.

- Mirador de ciencia y tecnología de la OCDE: 9 indicadores, con más de 24 unidades de medición, por cada país. En total, se tenían más de 216 bases de datos, por país.
- Estadísticas de patentes: 17 indicadores, con más de 9 unidades de medición, por cada país. En total, se tenían más de 153 bases de datos, por país.
- Estadísticas de investigación y desarrollo: 28 indicadores, con más de 7 unidades de medición, por cada país. En total, se tenían más de 196 bases de datos, por país.
- Indicadores de ciencia y tecnología: 1 solo indicador, con 71 unidades de medición, es decir, 71 bases de datos por país.

4.1.2.2.2 Protección Social y Bienestar

Indicadores que muestran un panorama preciso de los índices de bienestar y protección social de los países de la OCDE y, en consecuencia, del nivel de vida de sus habitantes.

- Protección Social: 15 indicadores, con más de 8 unidades de medición, por cada país. En total, se tenían más de 120 bases de datos, por país.
- Distribución del ingreso y pobreza: 4 indicadores, con 17 unidades de medición, por cada país. En total, se tenían más de 68 bases de datos, por país.
- Estadísticas de crimen: 8 indicadores, con 9 unidades de medición, por cada país. En total, se tenían más de 72 bases de datos, por país.

- Distribución de la riqueza: 4 indicadores, con 5 unidades de medición, por cada país. En total, se tenían 40 bases de datos, por país.
- Salarios e impuestos: 2 indicadores, con 7 unidades de medición, por cada país. En total, se tenían 14 bases de datos, por país.
- Índices de calidad de vida: 1 solo indicador, con 24 unidades de medición, por cada país. En total, se tenían 24 bases de datos, por país.
- Género: 61 indicadores, con más de 24 unidades de medición, por cada país. En total, se tenían más de 1424 bases de datos, por país.
- Familia: 9 indicadores, con 10 unidades de medición, por cada país. En total, se tenían más de 90 bases de datos, por país.
- Bienestar de la niñez: 1 solo indicador, con 21 unidades de medición, por cada país. En total, 21 bases de datos, por país.

4.2 Extracción de bases de datos

En función de las categorías descritas en el paso anterior, y con el tiempo limitado para utilizar el explorador, fueron descargadas más de 490 bases de datos que, en conjunto, tenían indicadores e información de los panoramas estadísticos del paso anterior.

El explorador permite exportar las bases de datos, en format CSV(Comma Separated Value), que guarda registros de bases de datos en texto plano. Cada registro, consiste en uno o más campos, separados por comas, de ahí su nombre.

Cuando se exporta, desde el explorador, una base de datos de una categoría, Educación, por ejemplo, no siempre es posible para un solo indicador o unidad de medición, como el desempeño en Matemáticas de la prueba PISA, por ejemplo. Se descargan, por default, todos y cada uno de los indicadores y unidades. En consecuencia, se terminan descargando, sin pretenderlo, más bases de datos de las necesarias.

En ocasiones, cuando se pretendía acceder a la información de algún indicador específico, era necesario descargar un archivo que tenía varias bases de datos

enteras, las cuales no estaban indicadas por nombre una vez que finalizaba la descarga, ni tampoco tenían un orden adecuado.

Para solucionar este inconveniente, desde el lenguaje SQL, con las consultas descritas en el marco teórico de este trabajo, se fueron creando tablas de cada una de las necesidades sociales que se pretenden trabajar desde este proyecto, para, de esta forma, organizar la información, evitar la polución informativa, y facilitar el diseño de la base de datos final.

4.3 Base de Datos Final

De las tablas organizadas en el paso anterior, se decidió, mantener, por tabla, los siguientes campos:

Educación

Desempeño de la prueba PISA en Matemáticas, Lectura y Ciencias Básicas.
Tasa de asistencia a la escuela.
Tasa de término de estudios de educación básica.
Desempeño en prueba PISA de los niños de nivel socioeconómico más bajo.
Desempeño en prueba PISA de los niños de nivel socioeconómico medio.
Desempeño en prueba PISA de los niños de nivel socioeconómico alto.
Diferencia en el desempeño PISA entre los niños de niveles socioeconómicos bajo y alto.
Años dentro del sistema educativo.
Porcentaje del PIB destinado a la educación.

Ciencia y Tecnología

Investigadores por cada 100,000 habitantes.
Porcentaje del PIB destinado al desarrollo de ciencia y tecnología.

Desarrollo Social

Viviendas sin servicios básicos.
Índice de Desigualdad.
Porcentaje del PIB destinado a la salud.
Porcentaje del PIB destinado al gasto social.

Niñez

Niños explotados laboralmente.
Desnutrición.
Menores de 5 años muertos por año.

Para cada tabla, con consultas de SQL, basadas en las teorías de conjuntos, se extraía lo que se deseaba de alguna base de datos, y se adjuntaba en una tabla creada por el graduando, que pudiera alojar los registros necesarios para el análisis deseado en este trabajo.

4.4 Modelado Matemático

Matlab, dentro de sus numerosas y útiles funcionalidades, ofrece dos que fueron utilizadas en este trabajo: el trabajo con bases de datos y la función “curve fitting”. De esta forma, las bases de datos pueden ser traducidas en polinomios matemáticos, en los cuales, variando los valores de las incógnitas, se puede conseguir una predicción de las mismas.

4.4.1 Bases de Datos en MATLAB

Los pasos de la metodología para trabajar las bases de datos en Matlab son los siguientes:

- 1) En una sola tabla, en formato .txt o .xlsx, guardar todos los datos. Dicho archivo deberá ser alojado en la carpeta raíz en la que se trabajará en MATLAB.
- 2) Desde la función “Import Data”, ubicada en la barra de herramientas, importar el archivo con el que se desea trabajar.
- 3) Se tendrá que elegir, ahora, qué registros se importarán. Por lo tanto, la cardinalidad y el grado, descritos en el marco teórico de este trabajo, quedará definida en función de este paso. En consecuencia, los atributos tendrán una

cantidad de registros que varían en función de la selección hecha por el usuario.

- 4) Se elige el formato en el que se quiere importar la base de datos, para que sea leída por MATLAB. Se puede exportar como vector de columnas, matriz, arreglo de celdas o tablas. En este caso, se utilizó la opción de tablas, seleccionando todos los campos de la base de datos final, descrita en la sección “Base de Datos Final” de este trabajo. El usuario puede seleccionar la base de datos que crea conveniente, según la aplicación de la metodología. Se importa la selección hecha.
- 5) En el “workspace” de MATLAB, se tendrá una cantidad de variables igual a la cantidad de atributos (véase Marco Teórico) seleccionados.
- 6) Cada una de las variables guardadas en el paso anterior, puede ser utilizada como parte de un polinomio que describa una combinación de las mismas. Para lograrlo, es necesario utilizar la aplicación “Curve Fitting”, que ofrece MATLAB. Cada variable representa un atributo de la base de datos.

4.4.2 Curve Fitting

Se encuentra en el Toolbox “Curve Fitting”. Los Toolbox en MATLAB son paquetes que ofrecen funcionalidades adicionales. Esta funcionalidad es bastante útil para representar ecuaciones representadas en funciones matemáticas $f(x,y)$. Tiene una interfaz interactiva que facilita el trabajo del usuario. Los pasos utilizados en esta metodología, son los siguientes.

- 1) En la sección que permite elegir las variables de la ecuación (X,Y,Z), seleccionar alguna de las variables importadas desde las bases de datos. Recordemos que cada variable representa un atributo. Así, se pueden seleccionar tres atributos de la base de datos, que serán representados como X, Y o Z, según sea el caso.
- 2) Se podrá ver una gráfica en 3D, donde cada eje representa a una de las variables del paso anterior.

- 3) El modelo de la gráfica dependerá del tipo de ecuación que se quiera generar. Esta herramienta ofrece 4 opciones:
- a. Interpolación. Tiene 5 métodos disponibles: vecino más cercano, lineal, cúbico, bi-armónico y spline.
 - b. Ecuación personalizada. El usuario puede introducir la ecuación de su conveniencia.
 - c. "Lowess". El usuario puede elegir un polinomio lineal o cuadrático, pero con un rango en porcentaje.
 - d. Polinomio. Genera un polinomio matemático, en el que una variable se representa en X, y las otras dos se representan en una relación en la variable Y.

Para efectos de esta metodología, se utilizara la función Polinomio, para poder generar una ecuación de primer o segundo grado, según sea el caso, por cada selección de atributos.

- 4) Se pueden elegir los grados de la ecuación. Debido a que la gráfica, en su mayoría, no cambia tanto en función de los grados, para esta metodología se recomienda generar ecuaciones de primer o segundo grado.
- 5) En la sección de resultados del Curve Fitting, se mostrará el polinomio resultante, así como algunas estadísticas que indicaran la confiabilidad del mismo, como el error cuadrático medio.
- 6) Para utilizar, después, de una forma más eficiente el trabajo de los puntos previos, se puede generar un código en MATLAB que realice lo descrito anteriormente.
- 7) Ahora, desde la opción "Generate Code" ubicada en la pestaña "File", se puede generar el código explicado en el paso anterior, y guardar en una función, que después, solamente, será necesario llamar.
- 8) La función generada en el paso anterior puede trabajar, también, con otras variables guardadas en el paso 5 de la sección "Bases de Datos" de este apartado. De esta forma, podemos mandar llamar la función y solamente

cambiar las variables; así, MATLAB generará el polinomio correspondiente a las variables introducidas.

9) Las ecuaciones quedan generadas, entonces, del siguiente modo:

$$Ax + By + C = 0$$

Donde A, B y C, son valores numéricos que ya da la función de MATLAB; en contra parte, “x” y “y” son las incógnitas.

4.5 Modo de predicción

Considerando “x” como una variable fija, cuyo valor se puede obtener de la base de datos, tendrá un valor que se pueda sustituir. En el caso de uso que se describirá después en este trabajo, “x” representa al promedio de asesinatos anuales en México, por cada 100 mil habitantes, cuyo valor es de 13.1; para cada caso de aplicación, evidentemente, el valor cambiaría, pero se puede sustituir, dejando, de esta forma, solamente “y” como incógnita, de la forma siguiente:

$$y = \frac{-(Ax + C)}{B}$$

Y representará la relación entre las otras dos variables comparadas versus la primera. De este forma, tendríamos un valor numérico de “y”. Por ejemplo, como se verá en el caso de uso, “y” puede representar la relación de los índices de desnutrición en niños y los índices de asistencia escolar.

Es necesario calcular el valor de “y” para cada polinomio del cual se quiera hacer la predicción.

Ese valor numérico de “y”, puede subir o bajar. Consecuentemente, la modificación del mismo, ya sea a causa de una adición o una sustracción, modificaría el valor de “x”. Es decir, si la representación de las variables de la incógnita “y” (desnutrición y asistencia escolar, por ejemplo) baja en su valor, consecuentemente, el valor de “x” se modificaría; de la misma forma, si el valor de “y” sube, “x”, también, se modificaría. Esto, daría como resultado el valor de “x” (asesinatos por cada 100 mil habitantes) modificado en función del crecimiento o decrecimiento de otras variables, como las ejemplificadas para “y”. Evidentemente, las variables “x” y “y” pueden representar los atributos que el usuario decida.

Así, se tendría la predicción de una problemática social, ya modelada y comprobada matemáticamente. Ahora, es necesario mostrar los resultados en una interfaz.

4.6 Interfaz

El objetivo de la interfaz es, fundamentalmente, mostrar los resultados de la predicción en una forma simple y fácil de entender. En ella, el usuario podrá modificar, en los rangos propuestos, el crecimiento o decrecimiento de la variable y , para poder ver la variación de la variable x . Como se mencionó antes, la intención es que cualquier trabajador social pueda transformar en conocimiento tangible la información que vea en la interfaz; por tanto, al usuario, no se le muestra variables “ x ” o “ y ”, per se. En cambio, se muestran los nombres de los atributos que representan las variables; es decir, los índices de desarrollo de la OCDE y la UNICEF. A la interfaz se le pueden añadir las variables de los polinomios que sean necesarios, para hacer cualquier predicción deseada, apoyándose en la metodología propuesta.

La interfaz desarrollada será usada, entonces, para mostrar el resultado final de la predicción. En ella, se pueden percibir los siguientes elementos.

4.4.3 Combo Box

En un Combo Box de C#, se puede mostrar una lista de texto, en la que cada línea es un caso distinto. Cada caso es elegido por el usuario y, en función de esa selección, el programa realizará la acción necesaria.

Para esta Metodología, en el Combo Box, cada uno de los polinomios estará representado por un caso, pero, el usuario, solamente verá el nombre de los índices de la OCDE o de la UNICEF, para facilitar la interpretación de la información.

4.4.4 Button

Para que el usuario modifique el valor de los indicadores mostrados, se tiene un botón que, al ser presionado, modificará el valor de “ y ” y, por tanto, el de x , mostrando la variación descrita, pero de los nombres de los índices.

4.4.5 Programa

Switch

Cada uno de los casos del combo box, representa un caso dentro de un switch. Por tanto, por n líneas disponibles para el usuario en el combo box, habrá n-1 casos para el switch.

En cada uno de estos casos, se declaran como variables cada uno de los valores de la ecuación (A, B, C, X y Y) que fueron encontrados como se describió en la sección de “modo de predicción”. Para cada polinomio se hará este proceso, y cada polinomio, tal cual se mencionó, tendrá sus valores alojados en cada caso.

Así, la siguiente ecuación tendría, para cada elemento, valores numéricos:

$$Ax + By + C = 0$$

En cada uno de los casos del switch, se mandará a llamar las funciones que se describirán a continuación, para conseguir modificar el valor de “y” y observar el comportamiento de “x”. Para ello, hay una variable, que recibe el valor de “y” modificado por el usuario, y es enviado a las funciones.

Funciones

Despejando “y”, como se describe en “modo de predicción”, se tiene:

$$y = \frac{-(Ax + C)}{B}$$

En el paso anterior, se dijo que el valor de “y” es modificado por el usuario, por tanto, ahora, se debe recalcular el valor de “x”, que, despejándola, queda de la siguiente manera:

$$x = \frac{-(By + C)}{A}$$

Reiterando, “y” tendrá un nuevo valor, modificado por el usuario. Ese nuevo valor dará un valor distinto para x.

De esta forma, la función tendrá el valor de x para la modificación respectiva de y; ese valor, será enviado al switch, para ser mostrado en la interfaz.

Capítulo 5

5.1 Caso de estudio: México

Realizando consultas en SQL, se obtuvieron las 20 regiones del mundo con mayor cantidad de asesinatos, dentro de los países miembros de la OCDE. De esas 20 regiones, 18 correspondían a estados de la República Mexicana.

De la misma forma, como se mostró en la introducción, los índices de desarrollo social y de ciencia tecnología, arrojaban como resultado una lista, en la cual México ocupaba, siempre, uno de los últimos 3 lugares. Por ello, se decidió experimentar qué tanto modificarían estos índices, a las cifras de asesinatos.

De la base de datos final, mostrada en la metodología, se extrajo, únicamente, los registros correspondientes a México, eliminando el resto de los países.

Se realizó en MATLAB el proceso descrito en la metodología, para traducir, esa base de datos con índices correspondientes a México, en funciones matemáticas, obteniendo 20 polinomios; es decir, 20 comparaciones de índices de desarrollo versus los asesinatos en México.

En cada polinomio, se tendrá como variable “x” el número de asesinatos anuales en México, por cada 100 mil habitantes. En la variable “y” se tendrán los índices que a continuación se muestran para, precisamente, ver el efecto que tendría, a futuro, la modificación de los mismos. Se listarán los atributos representados por la variable “y”, considerando, de facto, que la variable “x” será, siempre, la mencionada. Además, “ $x = 13.1$ ”, lo cual se obtuvo haciendo un promedio en la base de datos; es decir, de cada 100 mil habitantes, en México se asesinan a 13.1. Ese valor estará fijo para realizar los cálculos pertinentes. El valor de x será multiplicado por 10, y después por 120. Esto, a causa de que representa un valor por cada 100mil habitantes, al multiplicarlo 10, tenemos 1 millón de habitantes, y por 120, 120 millones de habitantes, cifra oficial de población en México. Se tendrá, así, cuántos asesinatos menos, o más, habrá en todo México, en función de que tanto mejoren o empeoren los índices propuestos.

5.2 Resultados

Trabajo infantil

$$f(x, y) = Ax + By + C$$

Donde

$$A = 480$$

$$x = 13.1$$

$$B = 84.36$$

$$y = 2.53$$

$$C = 213.7$$

Cada vez que el usuario presione el botón “subir”, $y = 2.53$, subirá 0.1.

Cada vez que el usuario presione el botón “bajar”, $y = 2.53$, bajará 0.1.

De esta forma, con cualquiera de las dos opciones anteriores, el valor de los asesinatos por cada 100 mil habitantes (x), se verá modificado en función de qué tanto mejore o empeore la situación del trabajo infantil en México (y).

Trabajo infantil - Desempeño académico en función del estrato socioeconómico de los niños

$$f(x, y) = Ax + By + C$$

Donde

$$A = -4.568e6$$

$$x = 13.1$$

$$B = 2.34e - 7$$

$$y = -8.42e - 7$$

$$C = 19.71$$

Cada vez que el usuario presione el botón “subir”, $y = -8.42e - 7$, subirá 0.1.

Cada vez que el usuario presione el botón “bajar”, $y = -8.42e - 7$, bajará 0.1.

De esta forma, con cualquiera de las dos opciones anteriores, el valor de los asesinatos por cada 100 mil habitantes (x), se verá modificado en función de qué tanto mejore o empeore la situación del trabajo infantil en México (y).

Desempeño académico en función del estrato socioeconómico de los niños – Desnutrición

$$f(x, y) = Ax + By + C$$

Donde

$$A = 1.53$$

$$x = 13.1$$

$$B = -0.624$$

$$y = -6.43$$

$$C = 17.19$$

Cada vez que el usuario presione el botón “subir”, $y = -6.43$, subirá 1.

Cada vez que el usuario presione el botón “bajar”, $y = -6.43$, bajará 1.

De esta forma, con cualquiera de las dos opciones anteriores, el valor de los asesinatos por cada 100 mil habitantes (x), se verá modificado en función de qué tanto mejore o empeore la situación del trabajo infantil en México (y).

Desempeño académico en función del estrato socioeconómico de los niños - Años en la escuela

$$f(x, y) = Ax + By + C$$

Donde

$$A = 5.016$$

$$x = 13.1$$

$$B = 0.00014$$

$$y = -4.32$$

$$C = 5.06$$

Cada vez que el usuario presione el botón “subir”, $y = -4.32$, subirá 1.

Cada vez que el usuario presione el botón “bajar”, $y = -4.32$, bajará 1.

De esta forma, con cualquiera de las dos opciones anteriores, el valor de los asesinatos por cada 100 mil habitantes (x), se verá modificado en función de qué tanto mejore o empeore la situación del trabajo infantil en México (y).

Viviendas sin servicios básicos - Desempeño académico en función de la desigualdad social

$$f(x, y) = Ax + By + C$$

Donde

$$A = -4.556$$

$$x = 13.1$$

$$B = -1.035$$

$$y = -9.08$$

$$C = 9.46$$

Cada vez que el usuario presione el botón “subir”, $y = -9.08$, subirá 5.

Cada vez que el usuario presione el botón “bajar”, $y = -9.08$, bajará 5.

De esta forma, con cualquiera de las dos opciones anteriores, el valor de los asesinatos por cada 100 mil habitantes (x), se verá modificado en función de qué tanto mejore o empeore la situación del trabajo infantil en México (y).

Gasto en educación como porcentaje del PIB

$$f(x, y) = Ax + By + C$$

Donde

$$A = -4.615$$

$$x = 13.1$$

$$B = -0.1784$$

$$y = 33.68$$

$$C = 5.06$$

Cada vez que el usuario presione el botón “subir”, $y = 33.68$, subirá 5.

Cada vez que el usuario presione el botón “bajar”, $y = 33.68$, bajará 5.

De esta forma, con cualquiera de las dos opciones anteriores, el valor de los asesinatos por cada 100 mil habitantes (x), se verá modificado en función de qué tanto mejore o empeore la situación del trabajo infantil en México (y).

5.3 Conclusiones

A través de la metodología, y como se puede observar, tanto en la interfaz como en los resultados, el desempeño de México en algunos indicadores sociales, sí tendría un impacto directo en la reducción, o incremento, de asesinatos en el país. Analicemos algunos casos.

Las bases de datos tenían incontables inconsistencias. Al ser auditadas rigurosamente por la OCDE, algunos países, México, especialmente, tenían valores nulos en muchos de los indicadores de desarrollo. En un principio, si se encuentran varios registros nulos, quien siga la metodología puede pensar que está cometiendo algún error con las consultas de SQL, pero, en realidad, se puede concluir que algunos países no envían todas las estadísticas, para no contradecir sus discursos políticos locales con la realidad de sus países.

Es fundamentalmente importante utilizar el sentido común cuando se manejan bases de datos. Aún si las consultas se arrojan de manera correcta, en base a la experiencia de este trabajo, puedo concluir que, para interpretar la realidad que las bases de datos intentan describir, hay que contrastar los resultados arrojados por las mismas. Por ejemplo, cuando se estaba creando la base de datos final, se podía observar que el PIB de México era, relativamente, alto; pero, en contra parte, el índice de desigualdad era el más alto de los países de la OCDE. Por tanto, se puede concluir que, uno de los objetivos fundamentales de las bases de datos es, como se menciona en el marco teórico de este trabajo, transformar los datos en conocimiento, usando el sentido común. De nada sirve crear algoritmos poderosos o hacer miles de consultas, si al final los datos no representan fielmente la realidad.

Si el trabajo infantil bajara un 50%, habría cerca de 146 asesinatos menos, por año, en México. De la misma forma, si el trabajo infantil se incrementara en un 50%, habría 167 asesinatos más, por año, en México. Con esta perspectiva, se puede desatar un debate que interprete las cifras. Si bien, como se mencionó al principio de este trabajo, una gran parte del trabajo social se hace en función de perspectiva,

y no de evidencias, con estos números, quienes se dedican a atender el trabajo infantil, ya sea desde una ONG o desde alguna institución gubernamental, tendrían evidencia que guíe y sustente su trabajo.

En el ambiente empresarial y social, el IPN ha sufrido un rezago considerable en los últimos años. El 77% de los CEO's de la totalidad de las empresas mexicanas, son graduados de tres universidades: Anáhuac, ITESM y Universidad Iberoamericana. Menos del 2% de los CEO's del país son egresados del IPN. Si bien la narrativa de la mayoría de mis profesores durante mi vida académica en la institución, se reduce a desarrollar tecnología, pero sin una visión global e integral, que permita, realmente, crear un impacto tangible con esa tecnología fuera de las aulas y laboratorios, a través de este proyecto, se propone cambiar la estrechez de ese criterio.

Durante mi experiencia trabajando en proyectos de desarrollo en República Checa, Brasil y Argentina, percibí que, en esos países, varias universidades siguen la tendencia de Silicon Valley, influenciada fuertemente por la Singularity University, que pretende crear empresas de tecnología que realmente propongan soluciones a los problemas de la humanidad. En otras instituciones, se realizan proyectos de menor complicación y sofisticación técnica, comparados con los que se desarrollan en el IPN. Al final, a pesar de su complejo desarrollo, los proyectos del IPN difícilmente aparecen en los círculos emprendedores del país y, consecuentemente, de la región de Latinoamérica. Es evidente la brecha que se debe aprovechar.

Además de ser, como se mostró en la justificación, el país con menos investigadores por habitantes, dentro de la OCDE, también es el que menos patentes realiza al año. En conclusión, no hay una correlación directa entre la complejidad técnica de una investigación, y el valor agregado real que podría tener en círculos empresariales.

Por tanto, la conclusión que más aprendizaje me ha dejado a lo largo de este proyecto, más allá del desarrollo técnico del mismo, es la imperante necesidad de desarrollar ideas disruptivas de tecnología, simples, especialmente, para que puedan ser utilizadas por personas que realmente le den un valor agregado. En el

caso de este proyecto, en acercamientos extra oficiales que se tuvieron con emprendedores e inversionistas, se mostró un interés en una idea similar para ser trabajada, de forma independiente, a futuro.

Bibliografía

1. Organización para la Cooperación y Desarrollo Económicos. *www.oecd.org*. [En línea]

2. United Nations Children's Fund. *www.unicef.org*. [En línea]

I-Library, OCDE. <http://www.oecd-ilibrary.org/>. [En línea]

La investigación en Trabajo Social con técnicas de lógica difusa. **García, Marta Cubero. 2010.** 48, Madrid : s.n., 2010, Vols. ISSN 1133-6552.

OCDE. <http://stats.oecd.org/>. [En línea]