CENTRO DE CIENCIAS BASICAS

DEPARTAMENTO DE SISTEMAS DE INFORMACIÓN

TRABAJO PRÁCTICO

UTILIZACIÓN DE LINGÜÍSTICA COMPUTACIONAL PARA LA CODIFICACIÓN DE PRODUCTOS DE LA ENCUESTA ENGASTO

PARA OBTENER EL GRADO DE

MAESTRIA EN INFORMATICA Y TECNOLOGIAS COMPUTACIONALES

PRESENTA:

LI. HÉCTOR ONCHI VÁZQUEZ

TUTOR:

DR. CARLOS ARGELIO ARÉVALO MERCADO

COMITÉ TUTORAL:

DR. JAIME MUÑOZ ARTEAGA

M.C. LIZETH ITZIGUERY SOLANO ROMO

AGUASCALIENTES, AGS., 6 de Junio del 2012

UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

HECTOR ONCHI VÁZQUEZ
ALUMNO DE LA MAESTRÍA EN INFORMÁTICA Y TECNOLOGÍAS COMPUTACIONALES
P R E S E N T E

Estimado Sr. Onchi:

Por medio de este conducto me permito comunicarle a usted, que el trabajo tesis o caso práctico titulado "**Utilización de lingüística computacional para la codificación de productos de la encuesta ENGASTO**", está autorizado y será bajo la dirección del Dr. Carlos Argelio Arévalo Mercado; y como revisores el Dr. Jaime Muñoz Arteaga y la M en C Lizeth Itziguery Solano Romo, para que pueda obtener así el grado de Maestría en Informática y Tecnologías Computacionales.

Sin otro particular me permito saludarle (a) muy afectuosamente.

A T E N T A M E N T E
Aguascalientes, Ags., 1 de junio del 2012
"SE LUMEN PROFERRE"
LA DECANO

MTRA. MARTHA CRISTINA GONZÁLEZ DÍAZ

universidad autonoma de aguascalientes
CENTRO DE CIENCIAS BÁSICAS

VoBo M en C JORGE EDUARDO MACÍAS LUEVANO
SECRETARIO TÉCNICO MAESTRÍA EN INFORMÁTICA Y TECNOLOGÍAS COMPUTACIONALES

c.c.p. Dr. Alejandro Padilla Díaz.- Secretario de Investigación y Posgrado.
c. c .p. M en C Jorge Eduardo Macías Luevano.- Secretario Técnico de la Maestría en Informática y Tecnologías Computacionales.
c. c. p. Dr. Dr. Carlos Argelio Arévalo Mercado
c. c. p. Dr. Jaime Muñoz Arteaga
c. c. p. M en C Lizeth Itziguery Solano Romo
c. c. p. Archivo.

UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES

FORMATO DE CARTA DE VOTO APROBATORIO

M. en C. MARTHA CRISTINA GONZÁLEZ DÍAZ
DECANO (A) DEL CENTRO DE CIENCIAS BÁSICAS
P R E S E N T E

Por medio del presente como Tutor designado del estudiante **HÉCTOR ONCHI VÁZQUEZ** con ID 10953 quien realizó el trabajo práctico titulado: **UTILIZACIÓN DE LA LINGÜÍSTICA COMPUTACIONAL PARA LA CODIFICACIÓN DE PRODUCTOS DE LA ENCUESTA ENGASTO**, y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia, me permito emitir el **VOTO APROBATORIO**, para que el pueda proceder a imprimirlo, y así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

A T E N T A M E N T E
"Se Lumen Proferre"
Aguascalientes, Ags., a 05 de Junio de 2012.

Dr. Carlos Argelio Arévalo Mercado
Tutor de trabajo práctico

c.c.p.- Interesado
c.c.p.- Secretaría de Investigación y Posgrado
c.c.p.- Jefatura del Depto. de Sistemas de Información
c.c.p.- Consejero Académico
c.c.p.- Minuta Secretario Técnico

Por este conducto autorizamos al tesista:

L.I. HÉCTOR ONCHI VAZQUEZ

La impresión de su documento final de Tesis, ya que cumple con los requisitos de contenido y forma exigidos en la Universidad Autónoma de Aguascalientes.

Asesor

Dr. Carlos Argelio Arévalo Mercado

Sinodales

Dr. Jaime Muñoz Arteaga

M. en C. Lizeth Itziguery Solano Romo

INDICE

ÍNDICE DE TABLAS.

ÍNDICE DE FIGURAS.

RESUMEN.

El presente trabajo trata sobre el uso de la Lingüística Computacional aplicada a la codificación de los productos utilizados en la encuesta ENGASTO levantada por el INEGI. La lingüística computacional es una rama de la inteligencia artificial que trata sobre el procesamiento del lenguaje humano para diversos fines, entre los que destacan el mejoramiento de las interfaces entre el ser humano y la computadora. La Encuesta ENGASTO proporciona información detallada acerca de los rubros en los que se aplica el gasto en los hogares de México. En este estudio se llevó a cabo una prueba piloto mediante un Sistema de Información que utilizó la Lingüística Computacional para optimizar el proceso de codificación (asignar una clave, o código a un concepto de gasto para facilitar su procesamiento) de la mencionada encuesta y se observó que se lograron reducciones significativas en el tiempo de codificación y se redujeron errores (precisión) al eliminar pasos innecesarios. De tal suerte que se propone que las estadísticas ofrecidas por esta encuesta y tentativamente encuestas similares pueden ser más precisas y consistentes, si se aplica la tecnología de lingüística computacional.

ABSTRACT.

This study shows how the use of computational linguistics, applied to the coding process of a national statistics survey called ENGASTO. Computational Linguistics is a branch of Artificial Intelligence that deals with human language processing, mostly to create better human-computer interfaces. The “ENGASTO” National Survey gives detailed information about the economic topics in which Mexican Families spend their money. A pilot study was performed using an Information System that applied Computational Linguistics principles to the process of “coding” (that is, assigning a numerical code to a spending concept, to facilitate its processing) of such survey. Significant time and error reductions were reported, mostly by elimination of unnecessary steps and by automation of the coding process. It is proposed that similar national surveys can benefit from this technology and consequently, data given to the public could be more consistent and precise.

1. INTRODUCCIÓN

La Ingeniería Lingüística tiene como principal objetivo proporcionar medios para ampliar y mejorar la utilización de la Lengua, haciendo de ella una herramienta más eficaz. Desde el punto de vista tecnológico, la Ingeniería Lingüística ayuda a mejorar la utilización de la lengua en los sistemas informáticos, asimilando, analizando, seleccionando y presentando la información con el objetivo de que las máquinas lleguen a "entender" el lenguaje natural y, de esta forma, se satisfagan las necesidades de información de los usuarios con mayor precisión y se contribuya a superar el problema de exceso de información(Jurafsky& Martin, 2008). La información podrá estar alimentada y disponible para cualquier encuesta que quiera ser uso de ella.

Se planteará la forma de hacer la codificación con la tecnología lingüística o bien conocidas como tecnologías para el lenguaje humano (HLT) de tal manera que todos los escenarios que existan para un producto puedan ser reconocido y codificado bajo el estándar de la Encuesta, utilizando la filosofía de las herramientas de corrección ortográfica y traductores, tal el caso de Google traductor solo que en lugar de regresarnos una corrección o una traducción, nos regresará una posible palabra de la búsqueda y un código.

Se desarrolló un sistema web bajo los fundamentos de HLT que se denomina SEBS (Sistema de Entendimiento de Bienes y Servicios), es un sistema que permite la captura, almacenamiento e información de los productos.

El sistema tiene como finalidad la integración y compartición de información entre diferentes áreas del instituto, promoviendo con ello la colaboración entre cada una de las áreas involucradas en dicha encuesta.

SEBS desarrollado en java busca su independencia de cualquier tipo de sistema operativo, evitando ser un problema el sistema operativo del que sea que ocupen, además de que se pueda compartir la información entre las diferentes oficinas en las que se lleva a cabo la encuesta.

No obstante el sistema cuenta con un módulo de exportación de información para que esta pueda ser utilizada por otros sistemas de análisis de información, como por ejemplo sistemas estadísticos, los cuales se encargan de transformar esa información en indicadores económicos de suma importancia para la toma de decisiones en el país.

Es un sistema desarrollado en 3 capas, la interfaz del cliente, la aplicación como tal encargada de dar respuesta a estas peticiones e interactuar con la base de datos en la cual se almacena toda la información.

El sistema está dividido en dos perfiles :Administrador del catálogo y Responsable del proceso, estas dos figuras de acuerdo a sus responsabilidades son los encargados de canalizar la información grabada por los usuarios finales dando consistencia al sistema, cada uno de ellos tiene tareas y actividades muy específicas dentro del sistema , estas actividades podría dividirse de la siguiente manera :

Perfil del Responsable del proceso:

Este perfil es el encargado de asignar las cargas de trabajo a los administradores, basado en el número de responsables, no asigna estas cargas “manualmente” ya que el sistema cuenta con un algoritmo que en base a diferentes variables como por ejemplo usuarios activos, peticiones, usuario con mayor carga de trabajo, se encarga de distribuir estas cargas de manera uniforme entre los administradores, con lo cual el solo tiene que presionar un botón para que las cargas de trabajo sean asignadas.

Perfil del Administrador:

Una de las actividades más importantes de este perfil dentro del sistema , es la captura de la información, para cual cuenta con una interfaz muy sencilla que le permite clasificar esa captura en 3 tipos : productos, sinónimos y dialectos, cada uno de estos es capturado en una pantalla donde se encuentran previamente cargada la información mediante la cual se puede asignar la clave del producto, sinónimo o dialecto respectivamente, esta información es presentada en listas desplegables en las cuales el responsable va clasificando cada producto.

Además de capturar esta información el responsable cuenta con una herramienta para corregir el conocimiento capturado, es decir editar la primera captura de la información en caso de que sea necesario por algún motivo, al corregir esta información, el sistema la almacenara adecuadamente dependiendo el tipo de objeto que se esté capturando.

Un módulo muy importante del sistema es el buscador de productos, el cual puede tener acceso no solo por los dos perfiles mencionados anteriormente si no por cualquier usuario, puesto que no es necesario firmarse en el sistema para poder consultar la

información, esta consulta de información es clasificada de acurdo a la clave de cada producto a buscar, fue desarrollado mediante un algoritmo de HLT y busca darle la mayor utilidad posible a la información capturada.

El buscador es utilizado por los responsables del proceso de codificación y responde a peticiones de usuario no solo por nombre del producto si no basado en descripciones, y sonidos similares a al producto que se está buscando, se facilita el proceso de codificación ya que la respuesta del sistema es muy rápida dado el algoritmo utilizado en su desarrollo.

Todo esto provoca una reingeniería de proceso, algunos autores definen a la reingeniería de proceso como “el método mediante el cual una organización puede lograr un cambio radical de rendimiento medido por el coste, tiempo de ciclo, servicio y calidad, mediante la aplicación de técnicas y herramientas enfocadas en el negocio como una serie de procesos del producto principal del negocio, orientados hacia el cliente en lugar de una serie de funciones organizacionales”(Ferrero & Alda, 2007). El proceso evoluciona de ser manual, propenso a errores y tedioso, a un proceso automatizado, preciso y eficiente, que sea compartido con todo el negocio esto quiere decir, a cada proceso que intervienen en este, permitiendo ser la fuente de alimentación de los códigos de los bienes y servicios de la encuesta ENGASTO.

2. PROBLEMÁTICA

La encuesta ENGASTO (Encuesta Nacional de Gasto) se crea en el año 2011.Esta es formada por la experiencia y conocimientos que se obtuvieron de otras encuestas del INEGI tales como la ENIGH (Encuesta Nacional de Ingresos y Gastos del Hogar) y ENOE (Encuesta Nacional de Ocupación y Empleo). ENGASTO es una encuesta que permite medir el consumo anual y trimestral de los hogares y los componentes multidimensionales de la pobreza. Recaba por primera vez mediciones de kilos, litros o watts que se consumen en la República Mexicana.

El proceso de recolección de datos de la encuesta ENGASTO consiste en el levantamiento de información por medio de cuestionarios en una determinada zona muestra geográfica que se determina mediante un sistema interno de muestreo. La encuesta consta de 8 cuestionarios (Sociodemográfico, Hogares, Población, Gastos Diarios, Gastos Mensuales, Gastos trimestrales, Gastos Anuales, Gastos de la Vivienda). Cada uno de ellos capta información con el fin de obtener índices de Gastos de las viviendas y hogares. Una parte de los datos son recolectados por los entrevistadores y otra es captada en los propios Hogares con cuestionarios impresos que se llenan de forma manual.

Los cuestionarios son redactados de tal forma que los entrevistadores e integrantes de los hogares los comprendan. Esto implica que no existe una sola forma de mencionar los productos (todos los bienes y servicios que adquieren los hogares)en lo que se está gastando, sin embargo entran en una clasificación única con las recomendaciones emitidas por la ONU en la Clasificación del Consumo Individual por Finalidades (CCFI) que es aplicada en esta encuesta, dando origen a un cuadernillo con todos los productos posibles existentes con un respectivo código, cada producto pertenece a una familia mayor de productos(catálogo de productos) dichas familias son creadas y/o adaptadas al CCIF de la ONU.

Los clasificadores son las categorías en que los productos son ubicados para su estudio permitiendo la elaboración del catálogo de la ENGASTO se consideran:

- Divisiones ordenadas y agrupadas gradualmente mediante el marco del CCIF.
- Títulos en las divisiones y grupos simples.

- Clases limitadas pero exhaustivas.
- Clases mutuamente excluyentes.
- El clasificador considera rangos libres para incluir códigos en caso necesario de una futura revisión o expansión.
- Ofrece información para que se consideren los elementos que se incluyen y aquellos que se excluyen.

El catálogo es fundamental para la estructura de la ENGASTO (ver Figura 1) debido a:

1. Permite la generación de información sólida toda vez que delimita, ordena y clasifica las variables.
2. Asegura un tratamiento uniforme de los datos estadísticos.
3. Se apega a las recomendaciones emitidas por la ONU.
4. Ofrece un marco homogéneo de bienes y servicios, permitiendo el análisis y comparación a nivel internacional.

Figura 1 Catálogo de Bienes y Servicios

2.1 Estructura del catálogo.

El catálogo de bienes y servicios del consumo de la ENGASTO, está integrado por 12 divisiones, las cuales abarcan los gastos de consumo individual de los hogares, además de un rubro que considera los impuestos de la vivienda y los vehículos del hogar.

Las divisiones del catálogo (ver Figura 2) son:

01 Alimentos y Bebidas no alcohólicas.

02 Bebidas alcohólicas, tabaco y estupefacientes.

03 Prendas de vestir y calzado.

04 Vivienda, agua, electricidad, gas y otros combustibles.

05 Muebles, artículos para el hogar y para la conservación ordinaria del hogar.

06 Salud.

07 Transporte.

08 Comunicaciones.

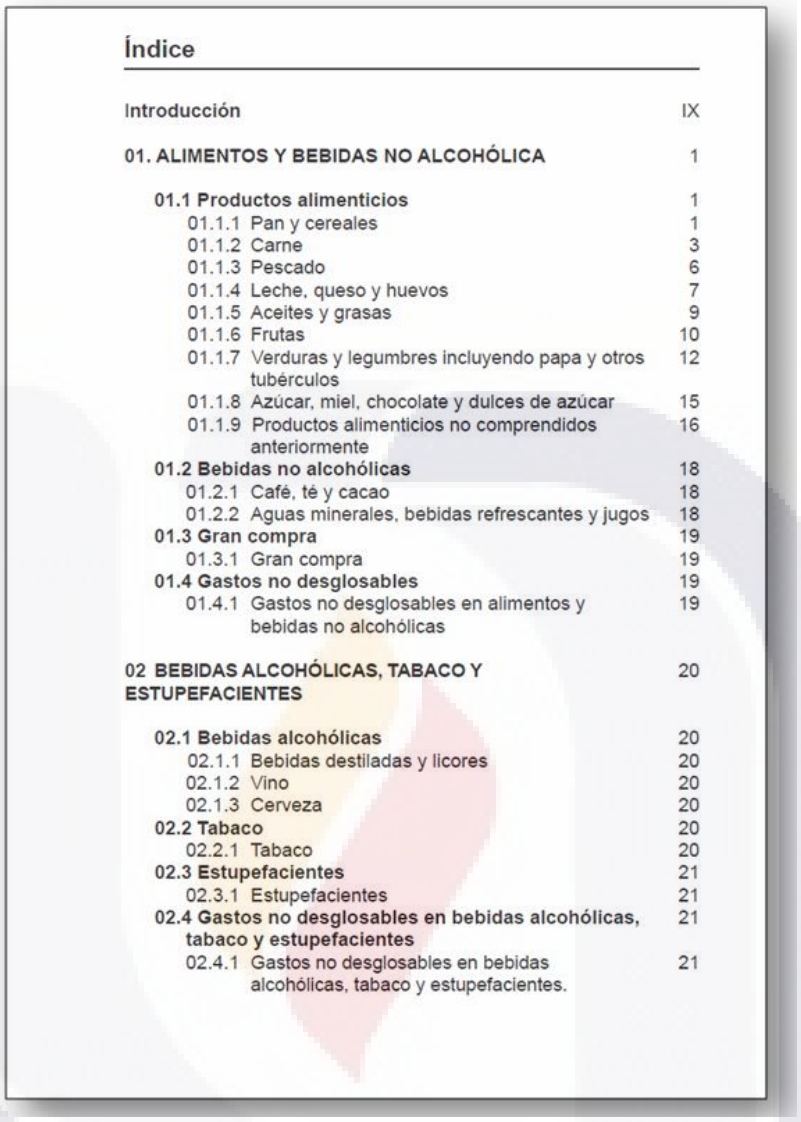
09 Recreación y cultura.

10 Educación.

11 Restaurantes y Hoteles.

12 Bienes y servicios.

20 Impuestos a la vivienda y a los vehículos de los integrantes del hogar.



Índice

Figura 2 Divisiones del catálogo

Cada una de las divisiones contiene:

1. Grupos: clasificación a 3 dígitos (01.1).
2. Clases: clasificación a 4 dígitos (01.1.1).
3. Subclases: clasificación a 5 dígitos (01.1.1.1).
4. Productos: clasificación a 6 dígitos y nivel más detallado (01.1.1.1.1)

La clasificación de los cuestionarios se realiza a nivel producto (Figura 3), a 6 dígitos. En los cuestionarios de gasto mensual, trimestral, anual ya hay códigos pre-codificados, solo

se completa el último número. En los demás cuestionarios, la codificación se hace total ya que no existen códigos pre-codificados (casillas vacías en el cuestionario)

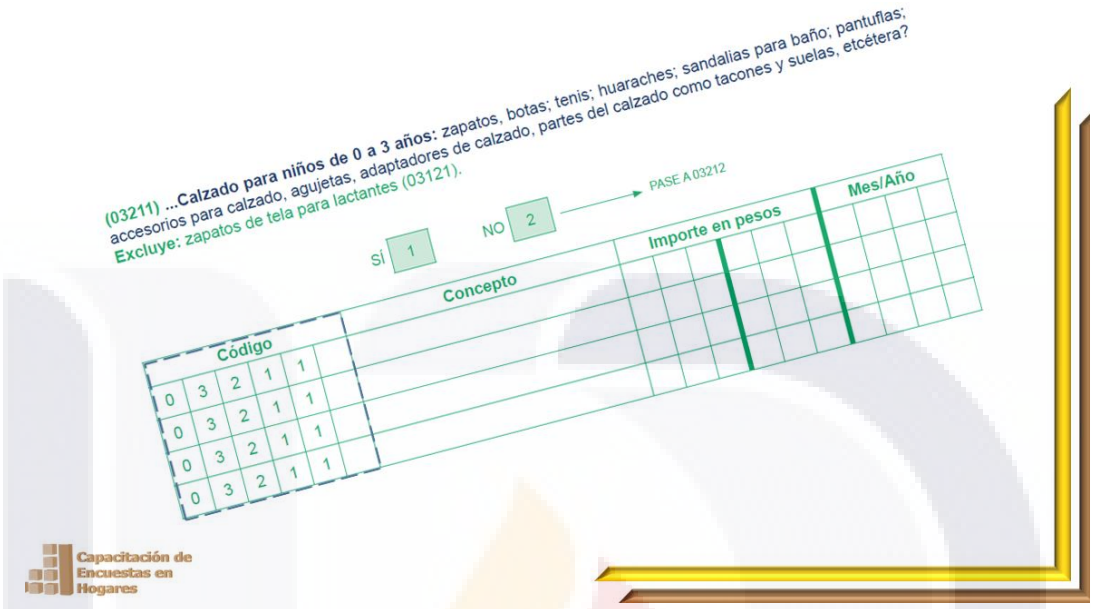


Figura 3 Codificación en cuestionarios

2.2 Concepto de codificación.

La codificación es la acción del entrevistador en donde hace la relación de los productos que fueron captados en los cuestionarios y lo refleja en los mismos cuestionarios pero con números, números que son la clasificación de los productos como se menciona anteriormente del catálogo (ver Figura 2).

Dado este nivel de complejidad, se encuentra que un producto puede ser nombrado de 1 - n diferentes maneras por ejemplo: es común que cierto extracto de población llame al “refresco” como “soda” o bien hasta con un nombre de marca como “coca”, ninguna de las tres palabras es tachable y nos dan un mismo significado que indirectamente nos lleva al mismo código de productos. Esto provoca que al codificar se generen códigos diferentes de un producto que pertenece a un solo código, y esto es debido a como se mencionó anteriormente, el producto tiene *n* formas de identificarse provocando que se codifique al conocimiento del captador y en donde se corre el riesgo de una mala interpretación de los datos al momento de generar la información estadística. Caso similar ocurre cuando las zonas de entrevistas son rurales, nos encontramos con lenguas indígenas (dialecto), por lo tanto, el entrevistador no siempre tiene el dominio,

provocando que al codificar se encuentre en el problema de codificar los productos cuando vienen con ésta situación.

Con base en la observación que se realizó en la ciudad de Mérida, Yucatán, y tomando 4 tipos diferentes de cuestionarios (cuestionario de gasto trimestral, cuestionario de gasto mensual, libreta de gastos individuales, cuaderno de gastos del hogar) al azar, se ha medido que un 1.68% de cuestionarios del hogar, el 3.37% de las libretas individuales, 1.2% del cuestionario de gastos mensuales y 4.4% del cuestionario de gastos trimestrales están siendo mal codificados, teniendo una incongruencia de información con la realidad (ver Tabla 1).

TIEMPO MIN/HOMBRE PROCESO MANUAL				
	Cuaderno de Gastos del Hogar	Libreta de Gastos individuales	Cuestionario de Gastos Mensuales	Cuestionario de Gastos Trimestrales
Códigos Incorrectos	9/954	3/208	1/83	7/250
Concepto fuera de Catálogo	2/954	0	0/83	4/250
Concepto con modismo	5/954	4/208	0/83	0/250
Concepto en dialecto	0	0	0/83	0/250
Total codificación no acertada	**16/954**	**7/208**	**1/83**	**11/250**
Porcentaje de error	**1.68**	**3.37**	**1.20**	**4.4**

Tabla 1. Análisis de errores de los cuestionarios.

2.3 Proceso de Codificación.

El proceso de codificación (ver Figura 5) está diseñado para realizarse en dos días, en los que se deben realizar las actividades de:

1. Reunir los cuestionarios que están contestados en la sección de sus “conceptos”.
2. Revisar concepto por concepto de cada uno de los cuestionarios e interpretar a lo que se está refiriendo.
3. Hacer la búsqueda en el catálogo impreso del concepto ubicado en los cuestionarios.
4. Escribir el código en el cuestionario conforme a lo que se interpretó y encontró en el catálogo.
5. Si no encuentra relación concepto-código, escribe y pregunta al jefe superior inmediato (supervisor de zona), de igual manera si la duda no puede resolverse se va en cadena burocrática ascendente hasta llegar a escribir en un Foro virtual de la encuesta ENGASTO sobre ese producto.
6. Capturar la información codificada en los sistemas de captación de datos.

2.4 Descripción del proceso

El proceso de codificación, como se ha mencionado, es manual y actualmente queda abierto a la interpretación y experiencia del entrevistador. Cuándo el entrevistador lee el concepto del cuestionario es cuando la interpretación y el conocimiento entran en juego porque debido a la gran variedad de productos de bienes y servicios existe una alta probabilidad de error. A pesar de que intervienen diversos roles en el apoyo de la codificación, la parte crítica queda sujeta a interpretaciones y conocimientos de los involucrados (ver Figura 4).

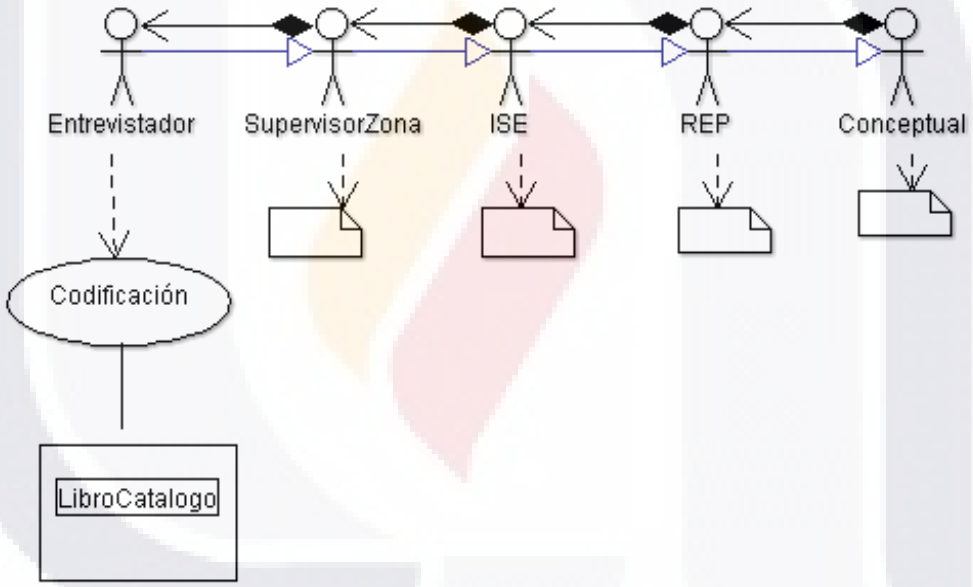


Figura 4 Proceso de Codificación

Cuando el entrevistador interpreta y comprende el producto, comienza la búsqueda exhaustiva por todo el libro del catálogo de bienes y servicios, está búsqueda es tardada y tediosa, provocando una inversión considerable de tiempo

El Entrevistador cuando no logra interpretar y/o encontrar un bien o servicio hace una petición al Supervisor de zona para aclararlo, quien al no ser aclaradas, hace una petición al ISE para nuevamente tratar de solucionar la codificación. Al momento que éste no tiene una respuesta, pasa hacer petición al REP para que pueda atender su duda, si la

duda no puede ser solventada, escribe un correo a oficinas centrales para que el área conceptual le solucione dicha petición. Una vez que ya fue resuelta, la respuesta fluye en escalera de igual manera como fue haciéndose (Área conceptual-REP-ISE-Supervisor-Entrevistador) (ver Figura 5).

2.5 Flujo detallado del proceso de codificación (actual)

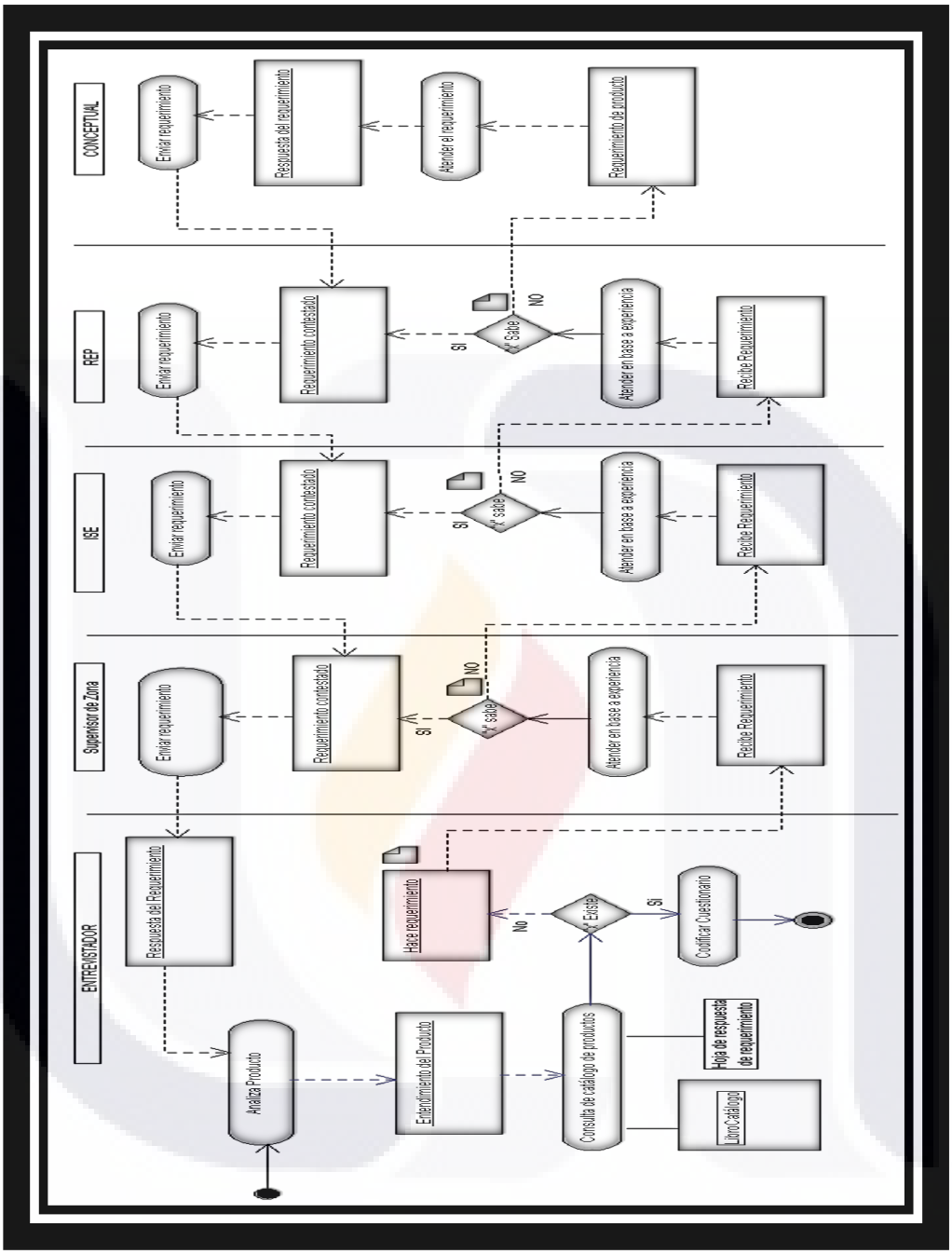


Figura 5 Flujo del proceso de codificación

3. JUSTIFICACIÓN

Al realizar la codificación se presentan situaciones que limitan nuestro proceso de la ENGASTO, con la observación y el análisis que se hizo en la ciudad de Mérida, tenemos el caso que para hacer una codificación en los cuestionarios del hogar (cuestionarios sin pre-codificación) están tardando un tiempo promedio de 43 min (ver Tabla 2) donde la mayor parte se centra en la búsqueda del producto, es aún más tardado que registrarlo en el cuestionario (ver Figura 6).

Tiempos Promedios (min/hombre) por producto	
Tiempo Promedio Análisis de Búsqueda	1.30
Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	5.83
Tiempo promedio de captura del código	1.41
Tiempo promedio por cuestionario	43.70

Tabla 2. Análisis de tiempo del cuaderno de gastos del hogar

Tiempos Promedios (min/hombre) por producto	
Tiempo Promedio Análisis de Búsqueda	0.94
Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	9.51
Tiempo promedio de captura del código	1.82
Tiempo promedio por cuestionario	13.5

Tabla 3. Análisis de tiempo del cuaderno de gastos individuales.

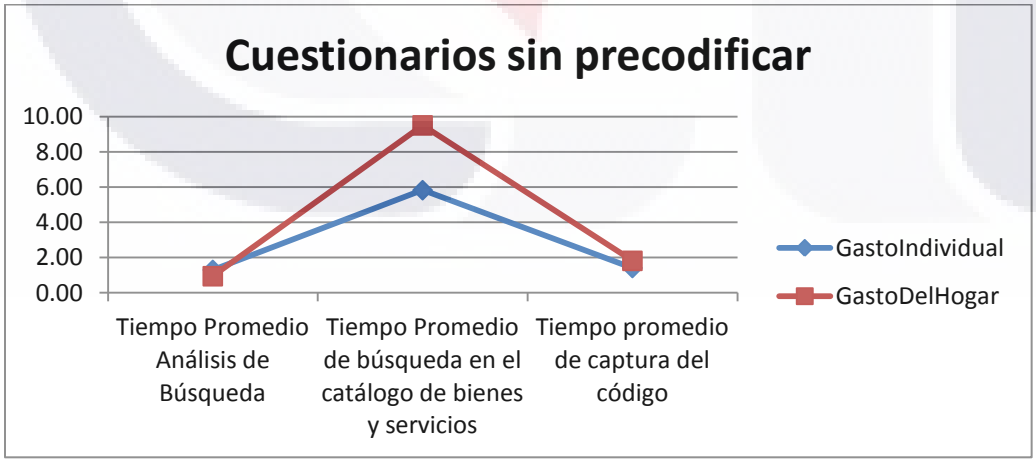


Figura 6 Gráfica del cuello de botella del proceso de codificación.

Los errores en el proceso de codificación como se menciona anteriormente (ver Tabla 1) pueden generar inconsistencias en la información estadística, ya sea tanto para índices de gastos, tabulados, ingresos, etc. Ya que todo los procesos anteriores caen en la codificación que es registrada en el proceso de captura, a pesar de que se registra el concepto del producto (ver Figura 3) lo que se usa para el análisis es el código del producto, debido que para los análisis de la información hacen caso omiso al concepto y toman el código como dato.

El alto volumen de información generado por este codificador (sistema basado en el entendimiento de bienes y servicios) como ya se había mencionado, permitirá la precisión del código del producto dado que tendrá el lenguaje de los productos de manera como la sociedad los conoce.

Los servicios generados con este estudio, tienen el potencial de ser utilizados en diferentes procesos de otras encuestas del INEGI, tanto en el sistema de captura para una posible automatización de captura de la codificación, a la etapa de validación para cotejar que el producto sea el esperado y en explotación para generar los índices de productos nacionales.

4. OBJETIVOS

Esta investigación servirá para diseñar y aplicar un servicio informático especializado para la codificación de los productos de la encuesta ENGASTO, de tal manera que sea escalable y crezca tanto como lo es la lingüística Nacional, teniendo la capacidad de codificar el mayor número de productos con un menor margen de error.

Este servicio de codificación será implementado en la Encuesta ENGASTO logrando con ello una gran evolución, trasladándonos de un proceso manual a un proceso automatizado, facilitando el trabajo a todos los procesos que hacen uso de ello, en específico a todos aquellos que codifican y su tiempo es breve para lograr el fin de sus actividades.

4.1 Objetivos Generales

Diseñar y probar un Sistema de Información basado en el rediseño del proceso de codificación de la encuesta ENGASTO que busque mejorar sus indicadores de operación (tiempo de ejecución, número de errores) utilizando elementos de lingüística computacional.

Demostrar que la implementación de un servicio de codificación basado en técnicas de lingüística computacional puede mejorar el proceso de codificación de encuestar los productos respecto al desempeño de un codificador humano.

4.2 Objetivos Específicos

- Habilitar un componente basado en tecnología de lenguaje humano para la codificación de productos dentro del proceso general de producción de la encuesta ENGASTO.
- Implementar un servicio de codificación en la encuesta ENGASTO, para que los entrevistadores que participan en el proceso, pasen de un proceso manual a uno automatizado.
- Disminuir el porcentaje de error en el proceso de codificación.
- Compartir la información de la lingüística Nacional para los procesos que estén interesados.

5. MARCO TEÓRICO.

5.1 Tecnología de Lenguaje

5.1.1 Lingüística computacional.

La lingüística computacional podrá entenderse como la disciplina que abarca tanto el procesamiento del lenguaje como el del habla desde una perspectiva general o desde un punto de vista teórico (Gómez, 2000a, b; Grishman, 1986; Jurafsky y Martin, 2000; Sidorov, 2001; Uszkoreit, 2000), aunque en ocasiones se encuentra esta denominación empleada como sinónimo de “procesamiento del lenguaje natural”(Llisterri, 2003).

En contexto se utiliza también el término de "Tecnología del Lenguaje Humano (HLT) es un área de estudio que trata de cómo el lenguaje humano se procesa en productos de tecnología de información, como ordenadores, teléfonos móviles y otros dispositivos digitales para diversas aplicaciones tales como el aprendizaje de idiomas y (la Web) la enseñanza”.

HLT y la naturaleza de los datos lingüísticos.

Con grandes cantidades de información disponible a través de la computadora y el Internet, el lenguaje es el medio a través del cual se puede transmitir esta información. Los datos del lenguaje, entonces, sería el recurso más esencial en el que especialistas de HLT debería trabajar. La naturaleza de los datos lingüísticos, sin embargo, es muy complicada. Lawler y Dry describen una serie de características de los datos lingüísticos, que también puede explicar por qué el conocimiento experto en lingüística es tan esencial en el proceso de desarrollo de los recursos y herramientas de HLT (Lawler & Dry, 1998):

1. Los Datos son Multilingüe.

2. Los Datos en texto se despliegan secuencialmente.

3. Los Datos se estructuran jerárquicamente.

4. Los Datos son multidimensionales.

5. Los Datos están muy integrados.

Una de las tareas más importantes para los especialistas en HLT es desarrollar las tecnologías pertinentes para que las personas que utilicen lenguas diferentes que tengan el mismo acceso a la información en esta era de la globalización. Como la tecnología se hace accesible a todas las partes de la aldea global, es el momento de abordar de manera más que nunca la cuestión de llevar esta tecnología aplicable a todo tipo de lenguas, culturas y visiones del mundo(Bodomo, 2006).

5.1.2 La Lingüística Computacional como rama de la Informática

Si el lenguaje es el vínculo entre la Lingüística Computacional (LC) con la lingüística, el empleo de los ordenadores como herramienta fundamental de trabajo conecta la LC con la Informática. No se trata solo de estudiar el lenguaje y las lenguas, sino de hacerlo con la ayuda que suponen hoy los ordenadores. La lingüística en programas con algún fin práctico.

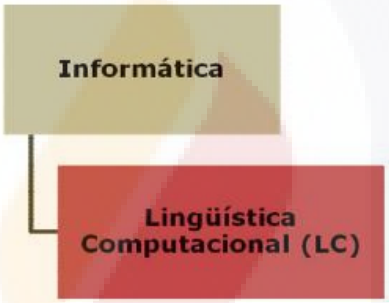


Figura 7 La LC como parte de la Informática

Para alcanzar esta pretensión también le resulta imprescindible las aportaciones de la informática, en especial una de sus subdisciplinas, la inteligencia Artificial (IA), que precisamente estudia todas las conductas inteligentes del ser humano, entre las que ocupa un lugar destacado del lenguaje. De esta forma, tanto la informática y la IA proporciona a la LC, técnicas, estrategias, formalismos de representación y otras herramientas que puedan contribuir, desde una orientación eminentemente aplicada, a ese objetivo de lograr ordenadores capaces de “hablar”(Villayandre Llamazares, 2010).

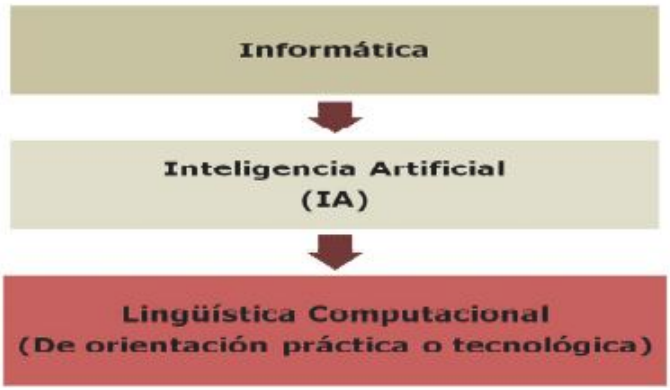


Figura 8 La LC como rama de la IA.

5.1.3 Breve historia.

El término “lingüística computacional” comenzó a usarse en los años sesenta, pero ya a finales de la segunda guerra mundial se había estado trabajando en este campo. De hecho, uno de los primeros usos que se les dio a las computadoras fue en el área de del procesamiento del lenguaje humano. La primera demostración de un sistema de traducción automática tuvo lugar en 1954.

A finales de los años 80´s comenzaron a ser utilizados con una creciente frecuencia para el análisis morfológico y sintáctico, para la traducción automática y para muchas otras áreas.

Los años 90’s han visto la revolución en internet y una consecuente necesidad de perfeccionar las tecnologías de procesamiento automático del lenguaje. Actualmente en todo el mundo industrializado numerosas empresas y centros académicos trabajan en el área. La LC aún está en su infancia, pero su desarrollo es cada vez más acelerado.

Áreas de la Lingüística Computacional.

Se trabaja para desarrollar aplicaciones para el análisis automático de la fonética, la fonología, la morfología, la sintaxis, la semántica y la pragmática. La generación de lenguaje puede implicar desde métodos para transformar conceptos complejos en representaciones semánticas fácilmente procesables por máquinas, hasta la transformación de un texto en lenguaje concreto y con convenciones muy particulares en una voz de apariencia humana(Domínguez Burgos, 2002).

5.1.4 Niveles de la Lingüística computacional.

Se puede tratar de desarrollar un modelo de lenguaje completo, sin embargo, es preferible dividir el objeto en partes y construir modelos más pequeños, y por ello más simples, de partes del lenguaje. Para esto se usa el concepto de *niveles del lenguaje*. Tradicionalmente el lenguaje se divide en 6 niveles:

1. Fonética/Fonológica.
2. Morfológica.
3. Sintaxis.
4. Semántica.
5. Pragmática.
6. Discurso.

No existe criterio exacto para la separación de cada uno de los niveles; de hecho, las diferencias entre los niveles se basan en el enfoque de análisis de cada uno. Por eso pueden existir traslapes entre niveles sin presentar contradicción alguno. Por ejemplo, existen fenómenos relacionados tanto con fonología como con morfología, digamos, alteraciones de raíces, *acordar-acuerdo, dirigir-dirijo,* entre otros casos.

5.1.4.1 Fonética/Fonológica

Es la parte de la lingüística que se dedica a la exploración de las características del sonido –que es un elemento substancial del lenguaje. Esto determina que los métodos de fonética sean en su mayoría físicos; por eso su posición dentro de la lingüística es bastante independiente.

A la fonología le interesa los sonidos pero desde otro punto de vista. Su interés se enfoca a la posición del sonido en relación con otros sonidos de algún idioma, es decir, las relaciones con los demás sonidos dentro del sistema y sus implicaciones. Por ejemplo, ¿Por qué los japoneses no pueden distinguir entre los fonemas [l] y [r]? la respuesta es que, en los idiomas nativos no existe oposiciones entre los fonemas mencionados y, por lo tanto, las diferencias que parecen muy notables en algunas lenguas, son insignificantes en otras.

5.1.4.2 Morfológica.

El área morfológica es la estructura interna de las palabras y el sistema de categorías gramaticales de los idiomas (género, número, etc.).

El objetivo de hacer un análisis morfológico automático es llevar a cabo la clasificación morfológica de una forma específica de palabra. Por ejemplo, el análisis de la forma *gatos* resulta en

Gato+sustantivo+masculino+singular

Que nos indica que se trata de un sustantivo plural con género masculino y que su forma lema es gato.

La variedad de designaciones a que aluden los dos géneros y la arbitrariedad de la asignación del género (masculino o femenino) a los sustantivos impiden, en muchos casos, determinar con exactitud lo que significa realmente el género. Es preferible identificarlo como un rasgo que clasifica los sustantivos en dos categorías diferentes, sin que los términos masculinos y femeninos provoquen prejuicio en algún sentido concreto.

No existe un modelo de reglas para la reflexión de género en sustantivos. Por lo tanto, en el programa se almacena todas las formas de sustantivos singulares en el diccionario (*por ejemplo gato y gata*).

5.1.4.3 Los Sinónimos.

La idea principal del método es permitir la búsqueda por palabras parecidas (relacionadas) en lo que respecta su sentido.

Para realizar la búsqueda se realiza el enriquecimiento de la petición. Para cada palabra el diccionario en este caso nuestras bases de datos contienen la lista de palabras relacionadas.

El diccionario para de sinónimos más cercanos permite encontrarlos de las palabras que se teclearon por el usuario en el campo de búsqueda: por ejemplo, para la petición *senado* se encontrará *asamblea*.

El concepto sinónimos comprende también, técnicamente, el concepto de las formas de palabras. Por esta razón se incluyeron las formas morfológicas de las palabras, además de sus sinónimos(Gelbukh& Sidorov, 2006).

5.1.5 Maternés.

Haciendo una conexión entre la Psicolingüística y la lingüística computacional hemos encontrado en un campo de investigación de la psicolingüística que, *a priori* podría

parecer alejado de la Lingüística Computacional, a saber, las investigaciones y estudios relacionados con el desarrollo del lenguaje. Y el caso es que un concepto clave de estos estudios es el de *maternés.* Según se recoge en OWENS, el maternés es la modalidad lingüística que utilizan los adultos (v.gr.: las madres, de ahí la denominación de maternas) cuando interactúan con los bebés/niños para facilitar la comprensión del mensaje, y por tanto, ayudar implícitamente en el desarrollo del lenguaje. Esta modalidad se caracteriza, léxicamente, por un vocabulario restringido temáticamente y no complicado. Se ha observado que, para las distintas aplicaciones en las que interviene el tratamiento del habla y, especialmente, el reconocimiento del habla, los locutores (quizá al considerar que las máquinas "no son tan maduras como el locutor adulto") utilizan una modalidad del lenguaje próxima al maternés, solo que, en este caso, el mensaje no va dirigido a un niño sino a un sistema computacional. Por ello, dado el paralelismo apuntado, nos gustaría denominar a dicha modalidad lingüística como el *computernés*.(Tordera Illescas, 2009)

5.2 Clasificación del consumo individual por finalidades.

1. En el Sistema de Cuentas Nacionales 1993 (SCN 1993) se incluyen cuatro clasificaciones de gastos por finalidades. Esas cuatro clasificaciones se incluyen en la presente publicación. Son las siguientes:

CFG: Clasificación de las funciones del gobierno;

CCIF: Clasificación del consumo individual por finalidades;

CFISFL: Clasificación de las finalidades de las instituciones sin fines de lucro que sirven a los hogares;

CGPF: Clasificación de los gastos de los productores por finalidades.

2. La CFG, la CCIF, la CFISFL y la CGPF tienen tres niveles de detalle que se denominan como sigue:

01 División (o nivel de dos dígitos);

01.1 Grupo (o nivel de tres dígitos);

01.1.1 Clase (o nivel de cuatro dígitos).

3. Las clasificaciones se definen ahora al nivel de clase o de cuatro dígitos. Sin embargo, en el SCN 1993 sólo se esbozan las estructuras de las clasificaciones: a nivel de dos dígitos para la CFISFL, y a nivel de tres dígitos para la CFG, la CCIF y la CGPF. Las estructuras se basan en las clasificaciones anteriores: la Clasificación de las funciones del gobierno publicada en 19803; la Clasificación de enseres y servicios domésticos del SCN 19684; la Clasificación de las finalidades de las instituciones sin fines de lucro que sirven a los hogares del SCN 1968, y la Clasificación por finalidades de los desembolsos de las industrias publicada en 19755.

4. La tarea de reestructuración y definición de las clasificaciones fue emprendida por la Organización de Cooperación y Desarrollo Económicos (OCDE) y la División de Estadística de las Naciones Unidas. La OCDE, en estrecha colaboración con la Oficina de Estadística de las Comunidades Europeas (Eurostat), se encargó de la CFG, la CCIF y la CFISFL. La División de Estadística de las Naciones Unidas se encargó de la CGPF. Las partes II a V de este documento recogen los detalles de la CFG, la CCIF, la CFISFL y la CGPF.

Finalidad y función

5. “Finalidad” y “función” se usan de modo indistinto en el SCN 1993; “objeto” se usó en el SCN 1968.Las tres palabras se utilizan con el mismo significado, es decir, los “objetivos socioeconómicos” de las instituciones al realizar diversos tipos de desembolsos.

6. Las cuatro clasificaciones están destinadas principalmente a clasificar las operaciones realizadas por los hogares, las instituciones sin fines de lucro que sirven a los hogares (CFISFL), los gobiernos y los productores que dan lugar a “sumas adeudadas”, es decir, las sumas de dinero abonadas o adeudadas por la adquisición de bienes corrientes y de capital o de mano de obra y otros servicios, por la adquisición de activos financieros o por la extinción de obligaciones financieras.

Más en concreto:

— La CFISFL y la CFG se utilizan para clasificar diversas operaciones, incluidos los desembolsos relacionados con los gastos de consumo final, el consumo intermedio, la formación bruta de capital y las transferencias de capital y transferencias corrientes realizadas por las instituciones sin fines de lucro que sirven a los hogares (ISFLSH) y por el gobierno general, respectivamente;

— La CCIF se usa para clasificar sólo un único tipo de desembolso, a saber, los gastos de consumo individual de los hogares, las ISFLSH y el gobierno general;

— La CGPF se utiliza para clasificar el consumo intermedio y los desembolsos de capital de las empresas constituidas en sociedad y no constituidas en sociedad de carácter financiero y no financiero.

7. En el capítulo IV, “Unidades y sectores institucionales” (SCN 1993) se dan definiciones completas de los sectores institucionales a los que se refieren las clasificaciones, por ello no se repiten en este documento.

Utilización de las clasificaciones de gastos por finalidades

8. En el capítulo XVIII, “Clasificaciones funcionales”, del SCN 1993, se describen las tres aplicaciones de esas clasificaciones.

9. La primera aplicación se relaciona concretamente con la CFG. Los servicios del gobierno pueden beneficiar a los hogares ya sea de manera individual o colectiva. La CFG se utiliza para diferenciar entre los servicios individuales y colectivos prestados por el gobierno general. Los gastos relacionados con los diversos servicios son tratados como transferencias sociales en especie. Del monto total de los gastos de consumo finales del gobierno se deducen las transferencias sociales en especie para obtener el consumo final efectivo del gobierno (o consumo colectivo efectivo), y para obtener el consumo final efectivo de los hogares (o consumo individual efectivo) se suman las transferencias sociales en especie a los gastos de consumo final de los hogares y de las ISFLSH.

10. La segunda aplicación tiene por objeto proporcionar una amplia variedad de estadísticas relacionadas con los gastos del gobierno, los hogares, las ISFLSH y los productores, de los cuales la experiencia ha demostrado que son de interés general y susceptible de ser utilizados en una amplia variedad de aplicaciones analíticas.

Por ejemplo, la CFG indica los gastos oficiales relacionados con la salud, la educación, la protección social y la protección ambiental, así como con los asuntos financieros y fiscales, las relaciones exteriores, la defensa y el orden público y la seguridad; la CCIF indica los gastos de los hogares en concepto de alimentación, ropa, vivienda, salud y educación, siendo todos ellos importantes indicadores del bienestar nacional; la CGPF

puede proporcionar información sobre la “externalización” de los servicios empresariales, es decir, sobre la tendencia cada vez más pronunciada de los productores a contratar fuera de la empresa los servicios de comidas, limpieza, transporte, auditoría y de otra índole que eran prestados anteriormente dentro de la empresa con carácter de actividades auxiliares.

11. La tercera aplicación de las clasificaciones es ofrecer a los usuarios medios para reestructurar agregados importantes del Sistema para determinados tipos de análisis. Por ejemplo:

— En estudios sobre la productividad de la mano de obra, los investigadores suelen necesitar una medición del “capital humano”, que normalmente se obtiene a partir de información sobre los gastos de educación de períodos anteriores. Las cuatro clasificaciones de los gastos por finalidades identifican los gastos de educación realizados por los hogares, las instituciones sin fines de lucro, el gobierno y los productores.

— Al estudiar el proceso de crecimiento económico, los investigadores prefieren a veces tratar una parte o la totalidad de los gastos en investigación y desarrollo y como formación de capital y no como consumo intermedio. En la CFG, la CFISFL y la CGPF se señalan los gastos de investigación y desarrollo en forma separada.

— En estudios sobre el gasto y el ahorro en los hogares, algunos investigadores consideran que es más útil estudiar los gastos en bienes de consumo duraderos como gasto de capital y no como gastos corrientes. En la CCIF se desglosan los gastos en bienes de consumo duraderos.

— En estudios sobre la repercusión del crecimiento económico en el medio ambiente, los investigadores suelen necesitar información sobre los gastos realizados en concepto de reparación o para la prevención de los daños ocasionados al medio ambiente. En la CFISFL, la CFG y la CGPF se señalan los gastos relacionados con la protección del medio ambiente.

Finalidades comunes

12. En el cuadro 1.1 que figura abajo se enumeran algunos de los objetivos socioeconómicos comunes a dos o más de las cuatro clasificaciones. Una “x” significa que la finalidad es considerada pertinente a un determinado sector institucional y, por

consiguiente, se señala en la clasificación correspondiente a ese sector; un guión (—) significa que la finalidad no es pertinente al sector correspondiente o que se considera que su valor es cuantitativamente insignificante en la mayoría de los países, y por ello no se consigna.

13. Las finalidades de gastos señaladas en estas clasificaciones son las consideradas importantes en la mayoría de los países en los últimos años del siglo XX. Tal vez ciertos países asignen gran prioridad a finalidades no indicadas en el cuadro 1.1 y con el transcurso del tiempo algunas finalidades señaladas en él serán posiblemente reemplazadas por otras que no pueden ser previstas en la actualidad. La necesidad de adaptar las clasificaciones de gastos según la finalidad a fin de contemplar las necesidades nacionales y de revisarlas para ajustarse a la evolución de las circunstancias es común a todas las clasificaciones internacionales.

Clasificaciones conexas

14. Las cuatro clasificaciones de gastos por finalidades a que se refiere el presente documento se vinculan entre sí, así como con otras clasificaciones internacionales enumeradas en los capítulos en que se trata concretamente de la CCIF, la CFISFL, la CFG y la CGPF. Una clasificación internacional común a todas estas clasificaciones, que no se menciona en estos capítulos, es la clasificación internacional uniforme de la educación 1997 (ISCED-97)6. Sin embargo, en el proceso de consulta muchos países solicitaron que se mantuviera la división de ISCED-767 Servicios de enseñanza no definidos por nivel. Esta división ha vuelto a utilizarse en todas esas clasificaciones(Naciones Unidas, 1999).

Finalidad del gasto	*Hogares CCIF*	*Instituciones sin fines de lucro que sirven a los hogares CFISFL*	*Gobierno general CFG*	*Empresas constituidas en sociedades y no constituidas en sociedades CGPF*
Salud	x	x	x	x
Recreación	x	x	x	x
Cultura	x	x	x	x
Educación	x	x	x	x
Protección social	x	x	x	x
Protección del medio ambiente	—	x	x	x
Investigación y desarrollo	—	x	x	x
Vivienda	x	x	x	—
Transporte	x	—	x	x
Comunicaciones	x	—	x	x
Socorro en casos de desastre	—	x	x	—
Ayuda económica al exterior	—	x	x	—
Religión	—	x	x	—

Figura 9 Ejemplos de finalidades de gastos comunes a más de una clasificación

5.3 Dialectos en México.

Las lenguas indígenas objeto del presente trabajo son aquellas que cuentan con hablantes vivos que adquirieron alguna de ellas como lengua materna y que la hablan con fluidez. Las fuentes que sustentaron la catalogación de las familias, agrupaciones y variantes lingüísticas aquí consideradas, así como la demarcación de las referencias geoestadísticas de sus respectivos asentamientos históricos, además de la Cartografía INALI 2005, fueron: **a)** los censos generales de población y vivienda 1990 y 2000, y los conteos de población y vivienda 1995 y 2005, realizados por el Instituto Nacional de Estadística, Geografía e Informática(INEGI); en particular, lo correspondiente a la información proporcionada por las personas que declararon hablar alguna lengua indígena; **b)** los conocimientos más recientes sobre genealogía, dialectología y sociolingüística generados por los investigadores de las lenguas indígenas de México y de sus países vecinos, accesibles mediante publicaciones, o proporcionados al INALI a través de consultas; y **c)** la información producto de las consultas que el INALI realizó a hablantes de las distintas lenguas indígenas nacionales.

Las 11 familias lingüísticas indoamericanas consideradas son:

I. Álgica.
II. Yuto-nahua.
III. Cochimí-yumana.
IV. Seri.

V. Oto-mangue.
VI. Maya.
VII. Totonaco-tepehua.
VIII. Tarasca.
IX. Mixe-zoque.
X. Huave

Estas familias lingüísticas se presentan de acuerdo con la distribución geográfica que tienen de norte a sur en el continente. Cabe aclarar que la integración de estas familias lingüísticas, lo que esencialmente significa el conjunto de agrupaciones lingüísticas comprendidas en cada una de ellas, corresponde, en algunos casos, a la propuesta que incorpora los análisis lingüísticos más recientes y más exhaustivos, y, en los otros casos, a la más aceptada entre los estudios de genealogía lingüística de las lenguas de la región. Para mayor información respecto de la ubicación geográfica del territorio donde se hablan las agrupaciones lingüísticas de las familias contenidas en este trabajo, sobre otras relaciones genealógicas, tanto de estas familias lingüísticas, como de las agrupaciones lingüísticas que las integran, y en torno a otros nombres dados a dichas familias(Instituto Nacional de Lenguas Indígenas, 2008).Identificando las familias lingüísticas en la república Mexicana:

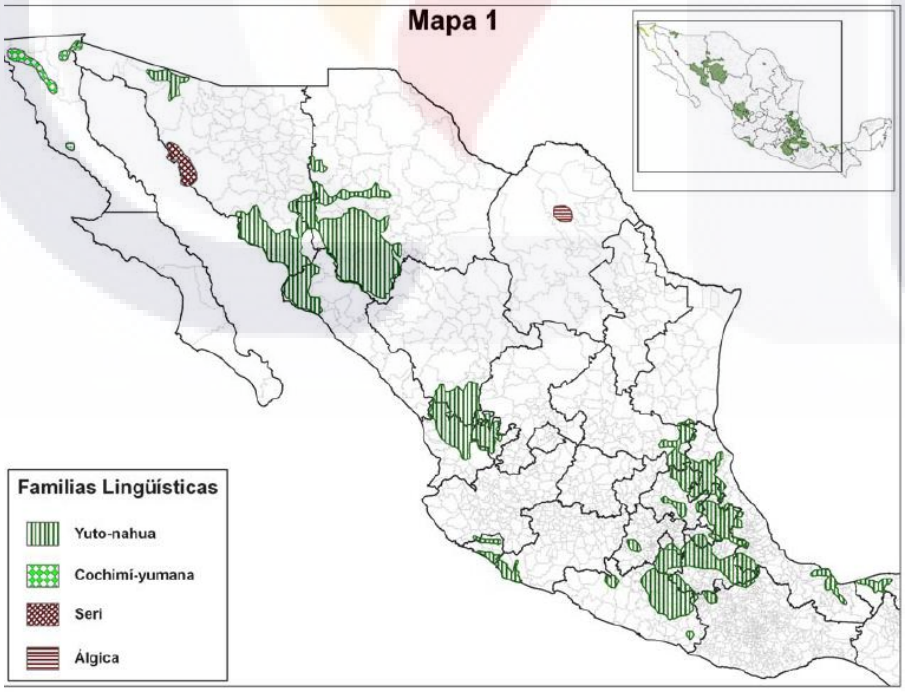


Figura 10 Familias lingüística Yuto-nahua, Cochimi-Yumana, Seri, Álgica.

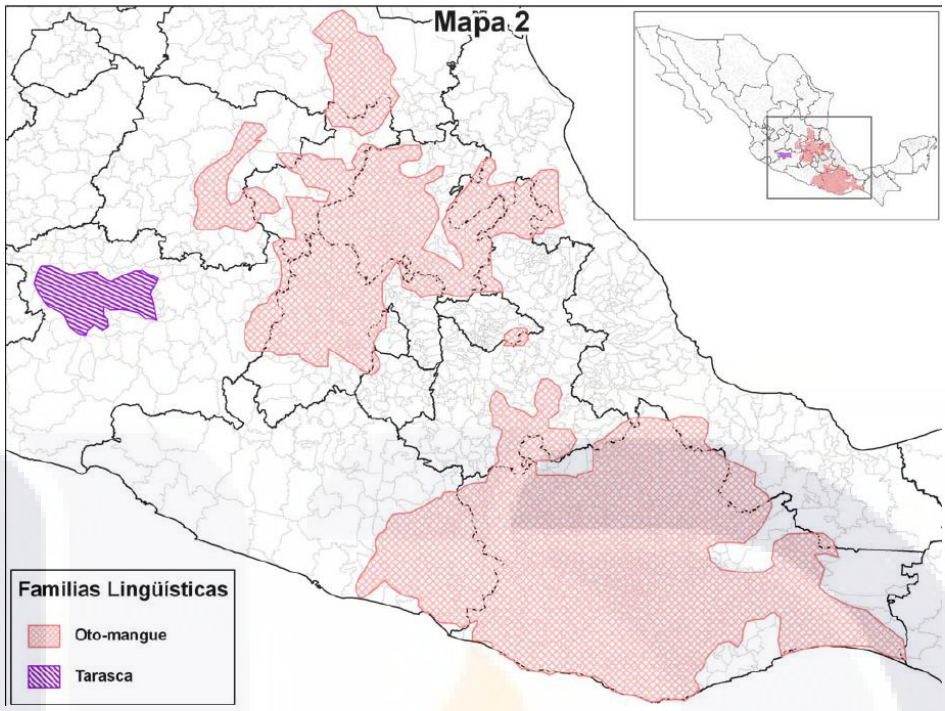


Figura 11 Familias lingüística Oto-mangue, Tarasca.

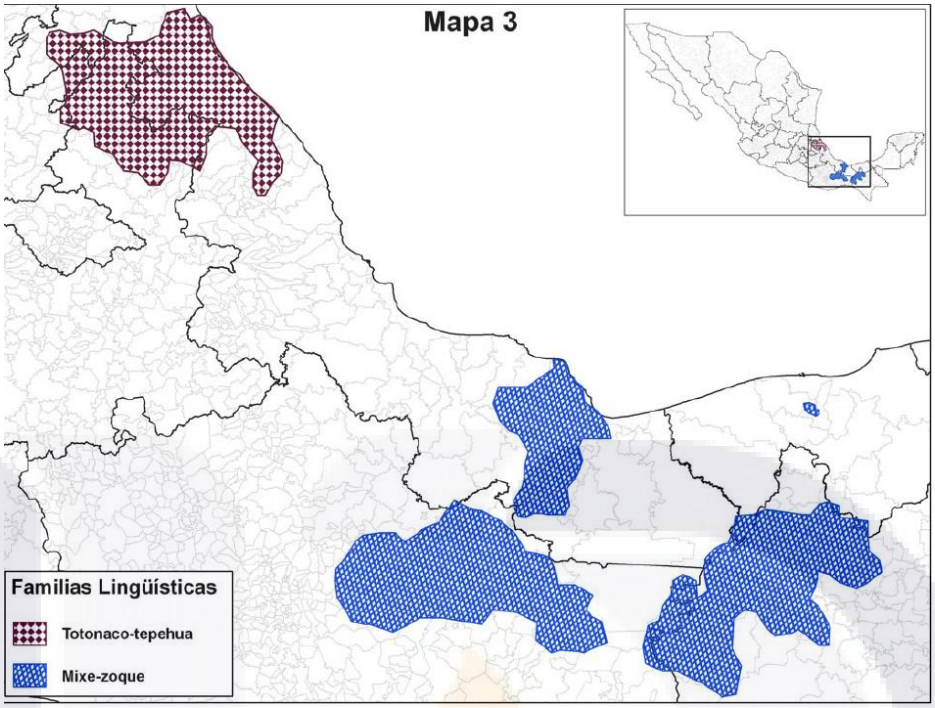


Figura 12 Familias lingüística Totonaco-Tepehua.

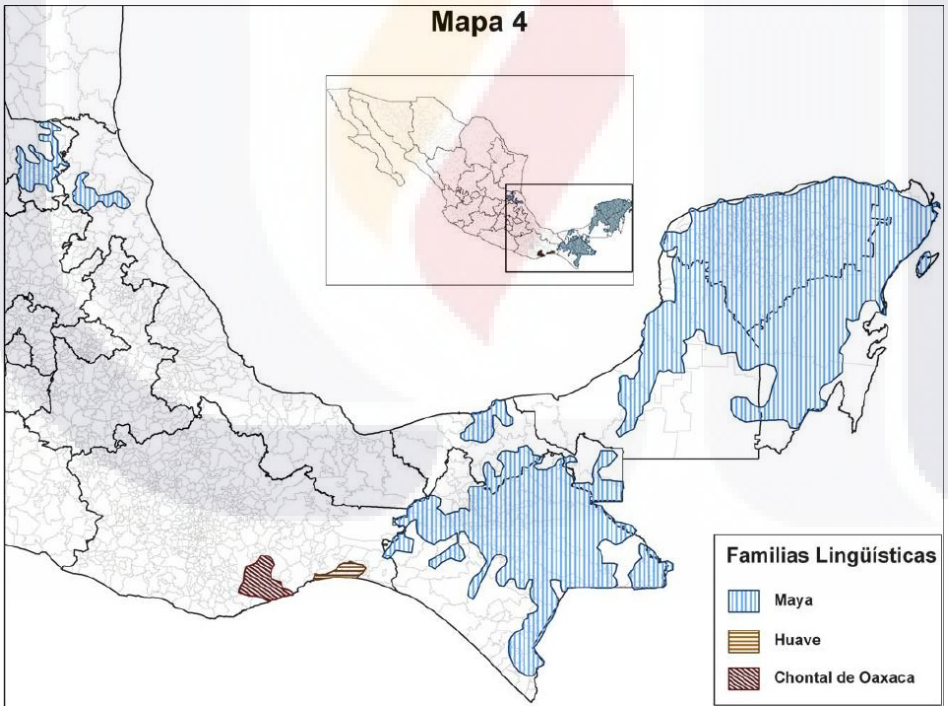


Figura 13 Familias lingüística Maya, Huave, Chontal de Oaxaca.

5.4 Reingeniería de Procesos.

Reingeniería es el rediseño rápido y radical de los procesos estratégicos del valor agregado –y de los sistemas, las políticas y las estructuras organizacionales que lo

sustentan- para optimizar los flujos de trabajo y la productividad de una organización(Manganelli& Klein, 2004).

Un proceso es una serie de actividades relacionadas entre sí que convierten insumos en productos. Los procesos se componen de tres tipos principales de actividades: las que agregan valor (actividades importantes para los clientes); actividades de traspaso (las que mueven el flujo de trabajo a través de fronteras que son principalmente funcionales, departamentales u organizacionales); y actividades de control (las que sea crean en su mayor parte para controlar los traspasos a través de las fronteras mencionadas)(Manganelli& Klein, 2004).

Mediante un rediseño rápido y radical modificamos no todos los procesos dentro de una organización sino solo aquéllos que son *a la vez* estratégicos y de valor agregado.

Los tipos de procesos de una organización son “Estratégicos” y “No Estratégicos”. Los “Estratégicos” son la más importante e indispensable para los objetivos, las metas, el posicionamiento y la estrategia declarada de una compañía; los procesos estratégicos son parte integrante de la manera como la compañía se define a sí misma. Los de valor agregado son los procesos indispensables para satisfacer los deseos y las necesidades del cliente, y por los cuales se está dispuesto a pagar; suministran o producen que él aprecia como parte del producto o servicio que se le ofrece(Manganelli& Klein, 2004).

5.4.1 Análisis de Procesos. Interacción entre procesos de la Empresa.

Es recomendable hacer el estudio del proceso de forma simplificada para facilitar su estudio, pero con suficiente detalle para no dejar en el aire cuestiones significativas. En cada estudio es necesario definir hasta qué nivel de detalle conviene llegar.

A) Elaboración de un diagrama de flujo.

Existen múltiples sistemas para realizar el análisis de un proceso. Uno de los más gráficos y de gran difusión es la realización de un diagrama de flujo, en los que se representa gráficamente cada una de las actividades de un proceso y se dibujan las relaciones entre ellas.

La realización de un diagrama de flujo es recomendable porque obliga a diseñar una secuencia lógica de realización de las operaciones que es uno de los objetivos de la reingeniería de los procesos.

B) Medición del proceso (indicadores).

Toda acción de mejora necesita partir de una cuantificación de sus parámetros representativos con los siguientes objetivos:

- Para poder proponer objetivos y medir si se han alcanzado, es necesario partir de datos numéricos.
- Proponer la realización de inversiones y analizar su rendimiento requiere también de valores numéricos.
- El uso de indicadores numéricos elimina la subjetividad en la evaluación del estado previo de los resultados.
- Para conocer los puntos débiles del proceso (excesos de inventario, cuellos de botella, baja productividad, mala calidad, etc.).

En cuanto a la cantidad de indicadores y su precisión, se recomienda que sean indispensables, de forma que el esfuerzo realizado en las medidas compense con los resultados obtenidos: un exceso de información tiene el efecto de dificultar el análisis de saturación.

Lo recomendable es que las medidas sean:

- *Sencillas*: El tiempo disponible para hacer mediciones está limitado, por lo que incrementa la precisión más de lo necesario va a consumir recursos necesarios para tareas de análisis.
- *Oportunas*: Para que describan la situación actual.
- *Objetivas*: Siempre basada en datos numéricos, no opiniones. Incluso cuando miden percepciones se deben expresar numéricamente.
- *Comprensibles*: Fáciles de interpretar.
- *Adecuadas*: Deben medir lo que se pretende.
- *Precisas*: Dentro de las limitaciones de economía de la medida.

Los indicadores más adecuados son aquellos que cuantifican de alguna forma la eficiencia (valor generado/recursos consumidos) o la eficiencia (porcentaje de éxito en la consecución de los objetivos)(Centros Europeos de Empresas Innovadoras de la Comunidad Valenciana (CEEI CV), 2008).

6. METODOLOGÍA

6.1 Rediseño del Proceso de Codificación.

En el presente trabajo, se plantea un rediseño del proceso de codificación para eliminar los pasos burocráticos que aumentan los tiempos de producción (ver Figura 15).

El rediseño del proceso resalta que los roles críticos son solamente dos: el entrevistador y el área conceptual. Dentro del área conceptual, se opta por dejar personal especializado en la atención de alimentar al sistema, llamándose "administradores del catálogo", logrando con ello una conexión 1 a 1 (ver Figura 14) donde el canal de comunicación es directo.

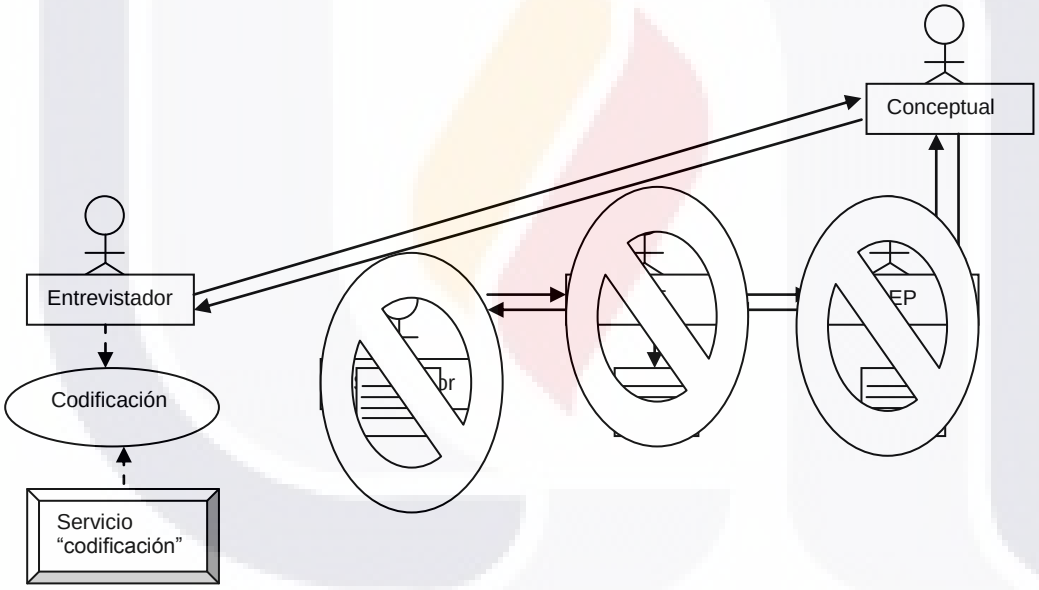
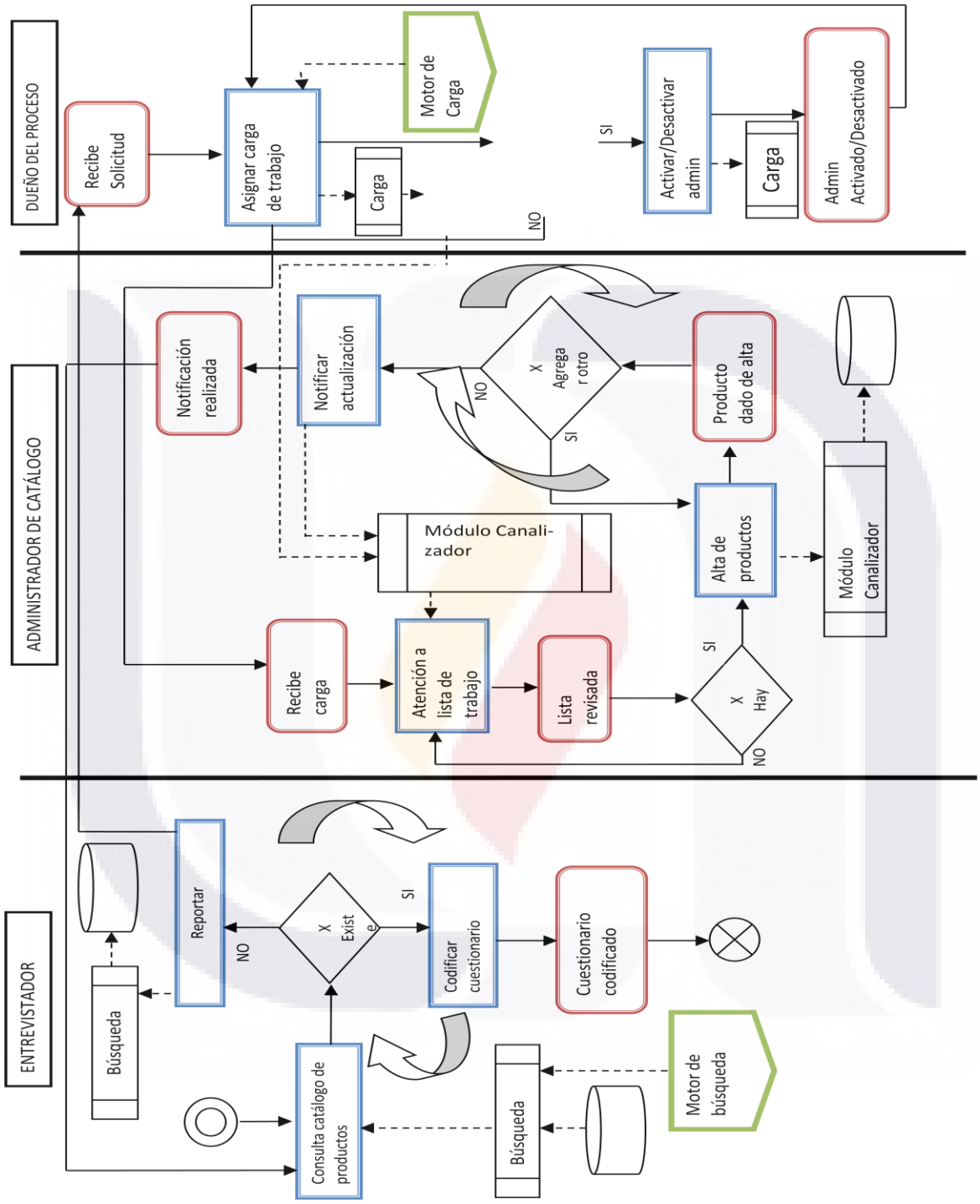


Figura 14 Relación Entrevistador - Área Conceptual

Se elimina la duplicidad de actividades y se busca simplicidad en el proceso, que para el entrevistadores tan fácil como activar solo una casilla.

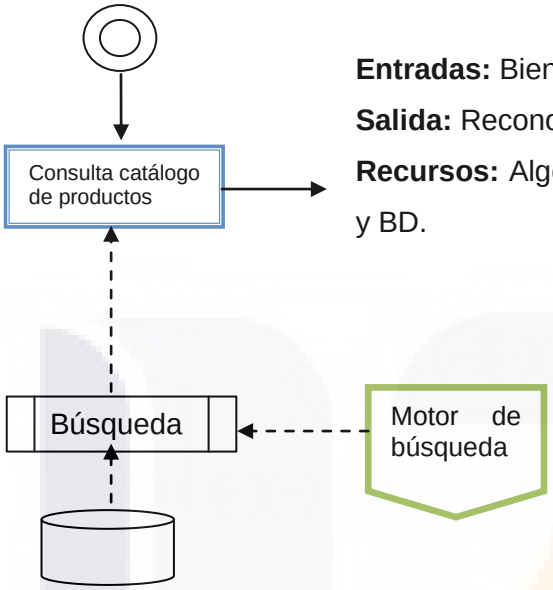
6.2 Flujo detallado del proceso de codificación (protocolo)



(Ver símbolos Anexo No. 3)

Figura 15 Rediseño del proceso

Descripción de actividad



Entradas: Bien o Servicio.

Salida: Reconocimiento de la búsqueda

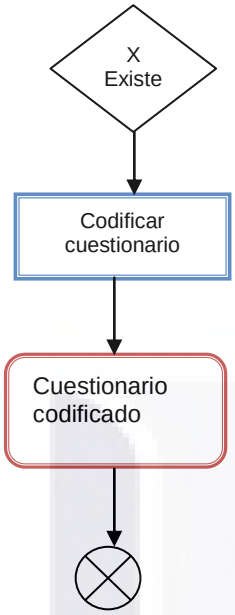
Recursos: Algoritmos basados en la lingüística computacional y BD.

Reglas del negocio:

1. Toda consulta se hará por nombres del producto.
2. La consulta guiará rumbo de acción (existe o no)
3. Toda consulta es por producto.

Procedimientos para la búsqueda.

1. Entrar al módulo de Búsqueda.
2. Ingresar el nombre del bien o servicio.
3. Presionar el botón de consultar.



Entrada: Producto existente

Salidas: Cuestionario codificado.

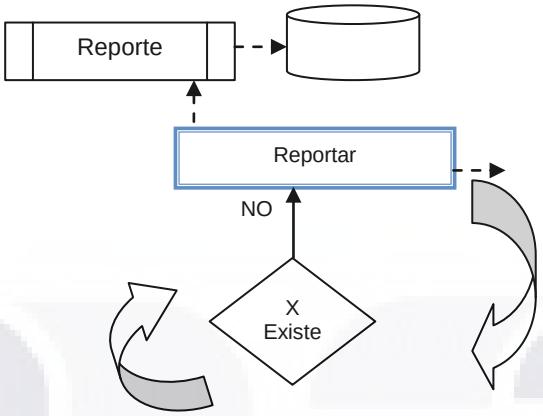
Recursos: Módulo de búsqueda.

Reglas del negocio:

1. Todo producto en existencia del sistema será reflejado en los cuestionarios.
2. Ningún producto del cuestionario se quedará en blanco y/o sin codificar.
3. A todo producto corresponde una codificación.
4. Toda codificación será a imagen de lo que el sistema muestra.
5. Todo cuestionario genera un ciclo de funciones y/o tareas que queda concluido al terminar la codificación del cuestionario.

Procedimiento para codificar cuestionario.

1. Copiar código que arrojó el sistema de codificación.
2. Consultar productos del sistema hasta completar todos los productos captados en el cuestionario.



Entrada: Producto y datos generales

Salidas: Solicitud de reporte.

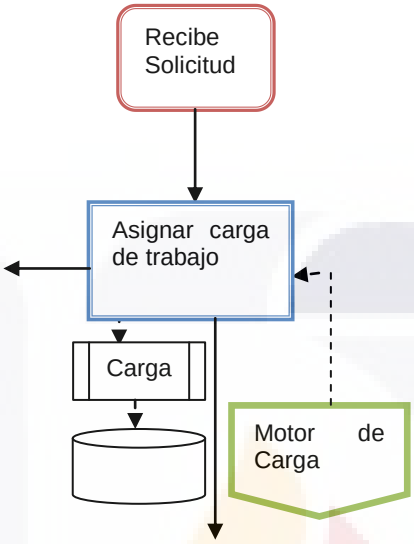
Recursos: Módulo de Reporte

Reglas del negocio:

1. Todo reporte debe tener un producto no existente de los resultados.
2. Escribirá en todos los campos obligatorios (*)

Procedimiento para reportar.

1. Abrir módulo de reporte.
2. Escribir nombre del entrevistador (opcional)
3. Escribir nombre de la Entidad y municipio donde se encuentran (opcional).
4. Escribir nombre del producto NO identificado.
5. Escribir algún comentario (opcional).



Entrada: Solicitud de reporte

Salidas: Asignación de carga de trabajo entre los administradores activos.

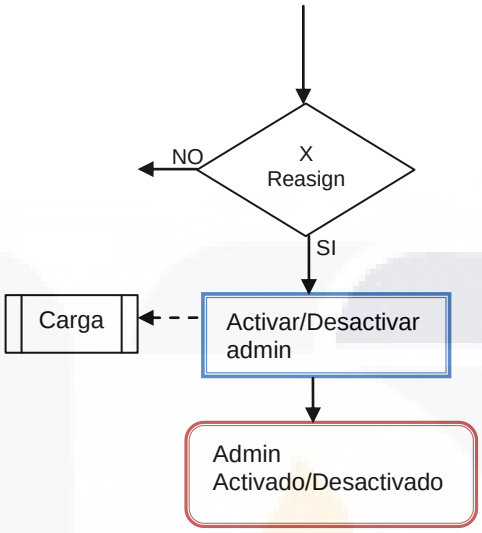
Recursos: Motor de carga (Algoritmos de reparto de cargas de trabajo), Modulo de Carga, BD.

Reglas del negocio:

1. Todo reporte debe tener un producto no existente de los resultados.
2. Escribirá en todos los campos obligatorios (*)

Procedimiento para asignar carga de trabajo.

1. Abrir módulo de motor de carga.
2. Identificarse como dueño del proceso.
3. Ejecutar la opción de carga de trabajo.



Entrada: Reporte de producto

Salidas: Reasignar carga de trabajo con los administradores modificados.

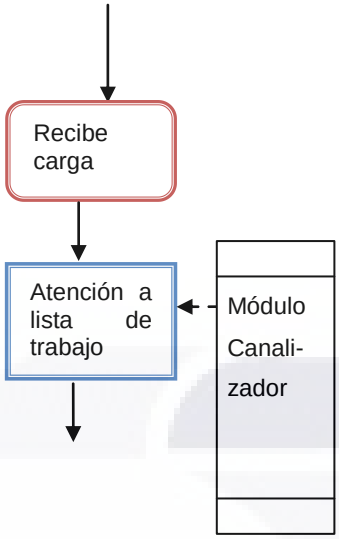
Recursos: Motor de carga, Modulo de Carga, BD.

Reglas del negocio:

1. Todo usuario debe ser activado o desactivado conforme lo que se necesite.
2. Todo cambio de Actividad o Inactividad se hace una reasignación de la cargas de trabajo.

Procedimiento para reasignar carga de trabajo

1. Abrir módulo de carga.
2. Identificarse como dueño del proceso.
3. Activar o desactivar los administradores.
4. Reasignar las cargas de trabajo.



Entrada: Carga de trabajo

Salidas: Atención al requerimiento

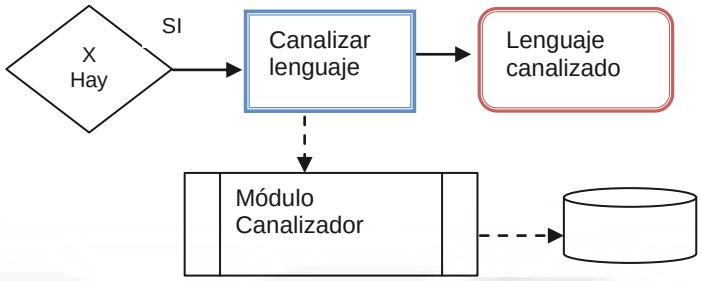
Recursos: Módulo de lista de trabajo, BD

Reglas del negocio:

1. Todo listado debe ser atendido.
2. El revisado del listado debe ser en orden en cómo fue solicitado (orden por día de solicitud).
3. Todo listado debe ser reportado en la situación en donde se encuentra la solicitud.

Procedimiento para la atención de lista de trabajo.

1. Abrir módulo Canalizador.
2. Identificarse como administrador del catálogo.
3. Reportar etapa de resolución en que se encuentra la solicitud.
4. Seleccionar un registro del listado para su atención.



Entrada: Atención de requerimiento

Salidas: Producto canalizado

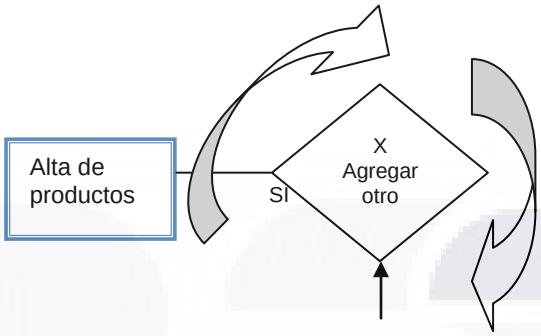
Recursos: Módulo Canalizador, BD

Reglas del negocio:

1. Todo listado debe ser atendido.
2. Todo producto de bien y servicio debe tener un nombre.
3. Todos los aparatados deben ser analizados y completados conforme las reglas de la lengua española.

Procedimiento para canalizar lenguaje.

1. Abrir módulo Canalizador.
2. Identificarse como administrador del catálogo.
3. Escribir el bien o servicio solicitado.
4. Registrar el análisis del bien o servicio.
5. Guardar conocimiento en el sistema.



Entrada: Solicitud del administrador

Salidas: Producto canalizado

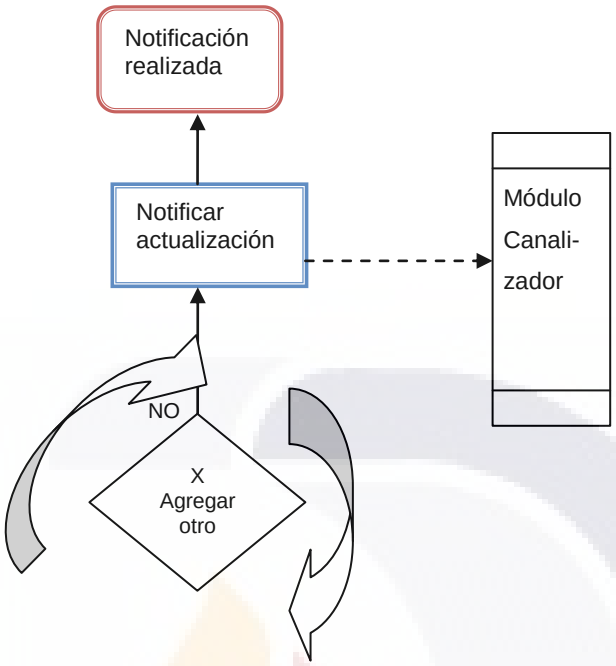
Recursos: Módulo Canalizador, BD

Reglas del negocio:

1. Todo producto de bien y servicio debe tener un nombre.
2. Todos los aparatados deben ser analizados y completados conforme las reglas de la lengua española.

Procedimiento para agregar otro producto sin lista de trabajo.

1. Abrir módulo Canalizador.
2. Identificarse como administrador del catálogo.
3. Escribir el bien o servicio solicitado.
4. Registrar el análisis del bien o servicio.
5. Guardar conocimiento en el sistema.



Entrada: Conocimiento guardado.

Salidas: Notificación de la actualización.

Recursos: Módulo Canalizador, BD

Reglas del negocio:

1. Todo producto de bien y servicio debe tener un nombre.
2. Todos los aparatados deben ser analizados y completados conforme las reglas de la lengua española.

Procedimiento para publicar actualización.

1. Abrir módulo Canalizador.
2. Identificarse como administrador del catálogo.
3. Publicar archivo actualizado.

REGLAS

1. Todo proceso debe tener entrada y salida.
2. Todo evento activa una función
3. Toda condicional activa una o más funciones.
4. Un conjunto de tareas y/o funciones es un proceso.
5. El rol es el responsable de ejecutar las funciones, eventos que están en su línea.
6. Todo proceso tiene un inicio y fin.
7. El flujo es apoyado por los conectores (Xor, Or, And).
8. Las funciones tienen la capacidad de llevar mensajes al rol
9. No se puede tener 2 roles en una misma calle
10. El Xor se cumple en un caso o en otro pero no en ambos.

6.3 Sistema basado en el entendimiento de Bienes y Servicios (SEBS)

En el servicio de codificación que se pretende brindar se vio la necesidad de desarrollar un sistema web (disponible por ahora en la intranet del INEGI: http://10.22.4.155:8989/SBS2-ViewController-context-root/faces/index.jspx)basado en las tecnologías del lenguaje con el fin de hacer el objetivo específico del negocio, que en esencia es un motor de búsquedas, tomando la filosofía de las tecnologías del lenguaje. Estas tecnologías tratan de generar lenguaje humano, para ello realizan un tratamiento automático. Para lograr este objetivo incorporan modelos teóricos, métodos y técnicas de diferentes disciplinas: lingüística, filosofía e ingeniería, ya que todas ellas están implicadas o pueden resultar útiles para tratar los diferentes procesos que envuelven el lenguaje natural.

Para asistir a este codificador es necesario hacerlo comprender lo que se le está pidiendo, para lograr ello se entiende las siguientes fases(Vallez, 2009):

- Reconocer
- Comprender
- Interpretar
- Generar

Para los dialectos se utilizará el plurilingüismo compuesto, que son diferentes lenguas (dialectos mexicanos) implicadas para acceder al significado de las palabras de cada lengua; las palabras de las diferentes lenguas remiten a un único significado cognitivo que es compartido, es decir, se podría decir que existe una interlingua en la que hayan su significado las palabras del resto de las lenguas (ver Figura 16). (TorderaIllescas, 2009)

Figura 16 Plurilingüismo compuesto en dialectos de México

En base a lo que nos indica el Instituto de Lenguas Indígenas, los dialectos que se hablan en México son 11 (Instituto Nacional de Lenguas Indígenas, 2008), con ello se generó una base de datos, quedando en el entendido que dicha información de igual manera que los productos se va a ir cargando conforme lo vayan solicitando los entrevistadores y canalizándolo el "administrador del catálogo" a un código del que pertenezca.

Los dialectos son mayormente usados por los entrevistadores que les toca encuestar en zonas rurales, ya que como se menciona en el documento, hay algunos cuestionarios que son llenados por los integrantes del hogar y otros en donde el entrevistador anota tal cual como el informante (dueño del hogar) se lo mencione, de tal

manera que si mencionan a un producto como ellos lo conocen (dialecto), este tipo de modelo del plurilingüismo compuesto, sea capaz de procesarlos para mostrarnos el código al que pertenece.

Con el siguiente análisis se muestran las distintas fases del análisis lingüístico al que es sometido un texto para obtener una representación de su significado, en este caso, obtener el código fijado conforme lo requerido (ver Figura 17). Cada fase tiene subprocesos que se ejecutarán para lograr una interpretación precisa y que no sea muy robusta la búsqueda, logrando así una codificación exhaustiva, eficiente y eficaz

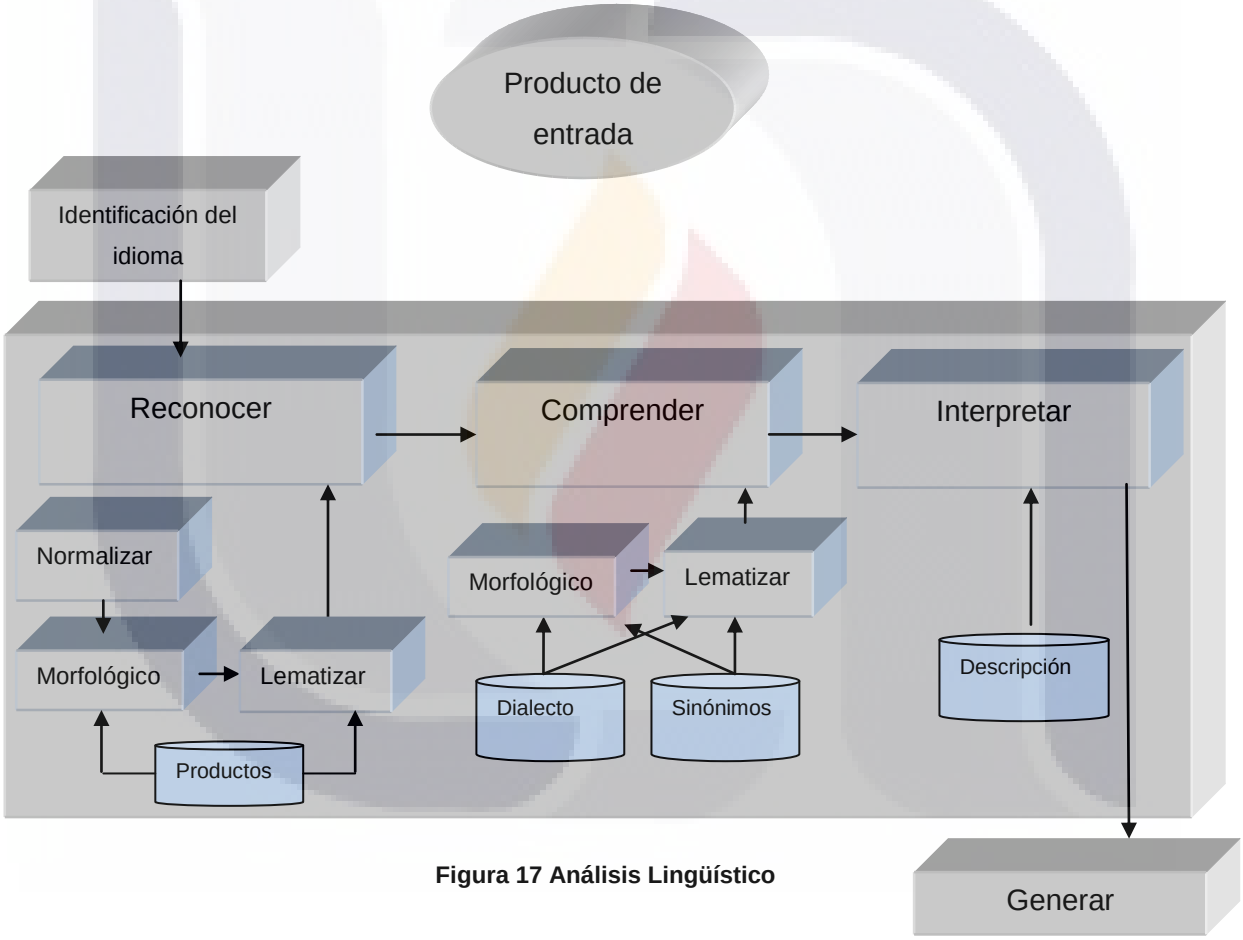


Figura 17 Análisis Lingüístico

6.3.1 Reconocer.

Es la primera fase lanzada por el sistema que pretende que el sistema haga el reconocimiento de los productos en su forma pura, donde reconoce las palabras *maternés*

y muestra todas las palabras que sean exactas o con algún patrón parecido a la búsqueda que se haya realizado.

La búsqueda se realiza por la tabla *maternés* que concentra los productos y sus propiedades morfológicas, lemas, añadiendo la etapa única de normalizar, si en esta etapa encuentra coincidencia, entonces, se encontrarán al inicio de los resultados.

Cuando el hilo es lanzado hace los subprocesos de:

***Normalizar*.**

Establece los términos y límites de la palabra, elimina el contenido que no aporta nada, homogeniza entre mayúsculas y minúsculas. De esta manera la búsqueda se inicializa con una pequeña filtración de posibles caracteres que no dan aportación a la consulta.

Análisis Morfológico.

Establece la relación de los productos en base a la lingüística, como el género (masculino, femenino), número (singular y plural).

***Lematizar*.**

Obtener el lema y poder representarlo en diferentes variantes, por ejemplo “vehículo” es el lema de “carro”, “coche”, “automóvil”.

6.3.2 Comprender

Es la segunda fase que es lanzada por el sistema, en esta fase trata de comprender que es lo que se está pidiendo, esto pasa cuando un producto es llamado de diferentes formas dependiendo del estrato social donde se encuentren.

La búsqueda transcurre por la información derivada de los productos *maternés,* en este proceso, es la oportunidad para hacer la búsqueda en los dialectos de México y obtener el código único al producto *matern* al que pertenece.

Cuando el hilo es lanzado a este subproceso, las palabras ya fueron normalizadas de lo cual este paso ya no es necesario repetirlo.

Cuando el hilo es lanzado hace los subprocesos de:

Análisis Morfológico.

Establece la relación de los productos y las ***n*** formas de llamarles (sinónimos) en base a la lingüística, como el género (masculino, femenino), número (singular y plural)

***Lematizar*.**

En este subproceso cambia de obtener el lema a obtener la etiqueta de las palabras derivadas (sinónimos) de los productos, que pareciera ser completamente diferente de lematizar los productos, no lo es así, la mecánica es exactamente igual, solo que ya son etiquetas lo que contiene dichas características, por ejemplo el sinónimo “Coca Cola Company” se relaciona con las etiquetas de “Coca”, “Coke”.

Análisis Fonético.

El análisis fonético a pesar que se centra en la captación de sonidos como entrada de datos al ordenador, lo que se hace es una búsqueda por todas aquellas formas en que los productos se escuchan igual independientemente de la letra que se esté usando (tal como el usuario se expresa), este se basa bajo el principio del sonido y no en el texto.

El análisis Fonético es el subproceso que se hace presente cuando la búsqueda no encuentra coincidencia en los dos primeros subprocesos (Reconocer y Comprender) esto con el fin de captar que el usuario esté expresando su conocimiento de sonidos y no de escritura.

6.3.3 Interpretar

Es el tercer y última fase en ser lanzada por el sistema, este consiste en hacer una búsqueda exhaustiva en todos los productos, el sistema comienza a interpretar por medio de su significado contextual palabras que estén relacionadas con el término que están buscando.

Este proceso es responsabilidad de los tutores (administradores del catálogo) de enseñar la parte contextual de los productos que está aprendiendo el sistema, esta parte contextual permite al sistema llegar más allá que un catálogo y poder abrir ese canal de comunicación entre el usuario (entrevistador para este caso) y el SEBS.

En base a la interpretación, pueden lograr conclusiones los entrevistadores con los resultados que están siendo mostrados por el sistema.

6.3.4 Generar.

Es la parte final de la interactividad entre el usuario y el SEBS, es donde se reflejan los resultados obtenidos después de todo el procesamiento (comunicación del ordenador). En la generación se puede saber si el conocimiento del sistema logró un resultado satisfactorio de lo contrario, capta la idea y lo deja preparado para su aprendizaje futuro.

6.4 Desarrollo del SEBS

6.4.1 Interfaz

Se creó una simplicidad de la modificación y puesta al día del conocimiento lingüístico del sistema, es decir, que el conocimiento lingüístico pueda almacenarse de tal modo que los lingüistas y otros expertos no informáticos puedan acceder a él y modificarlo fácilmente. Una interfaz que no requiera tampoco conocimiento informático experto para su utilización, y que ofrezca un acceso a la información rápido, sofisticado desde el punto de vista lingüístico.

La interfaz fue desarrollada en Java haciendo la conectividad con la base de datos Oracle por medio del driver de JDBC, permitiendo acceder a la base de datos independientemente de su categoría y del sistema operativo subyacente(Read& Bárcena, 2000). Una alternativa definitiva para almacenar el conocimiento en base de datos.

```
publicbooleanconectar(){ String url,pass,nom,reg; url
= "jdbc:oracle:thin:@localhost:1521:xe";
```

Es un sistema desarrollado en java con sus tecnologías javaServerPage, lo cual permite que su mantenimiento, instalación y uso sean sumamente sencillos.

Java puede desempeñar múltiples tareas a través de mecanismos de entretejido, y cada tarea puede dividirse en hilos para aumentar la velocidad del procesamiento. Con ello SEBS, lanza hilos para su procesamiento a las diferentes sub-procesos del análisis lingüístico.

En inicio lanza los hilos a los subprocesos de la primera fase de “Reconocer” (morfológico, lematizar).Cuando se lanza el hilo de esta fase es la posibilidad de encontrar una respuesta almacenándolo en una lista, caso contrario, al no encontrar nada, el sistema está programado para lanzar los hilos a los subprocesos de la segunda fase “Comprender” (morfológico, lematizar), de igual manera que la primera fase, espera la posibilidad de encontrar respuesta almacenándola en la lista, donde la lista es actualizada eliminando los casos de todos aquellos productos que estén duplicados)esta es la única fase que también sucede cuando el usuario quiere hacer la búsqueda en algún dialecto ejecutándose todos los subprocesos de dicha fase (ver Figura 18). Aun cuando no se encuentre un resultado en los subprocesos de las fases, es lanzado el último hilo de dichas fases y es el de análisis fonético tratando de encontrar respuesta al producto que se consulta, si es así, mostrará en pantalla el producto y su código. Por último si no se logró respuesta en la búsqueda, entra a la última fase de “Interpretar” lanzando el último hilo, buscando un patrón en todo el texto de la descripción de cada producto que esté almacenado en la lingüística (ver Figura 18).

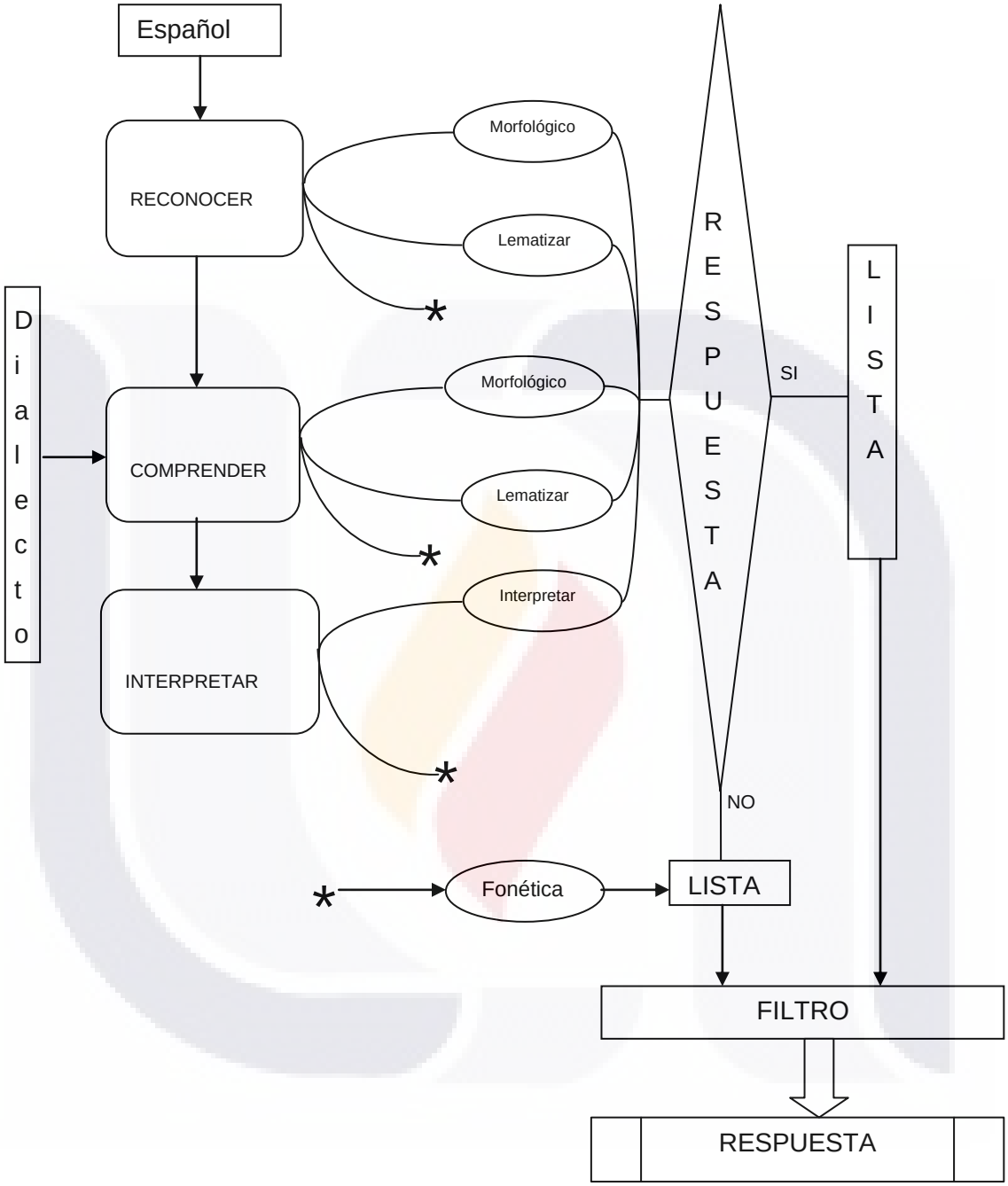


Figura 18 Procesamiento del reconocimiento del producto

6.4.2 Base de datos

La comunicación entre el administrador y la información es bidireccional, ya que es tan escalable como sea necesario. Edita, Elimina y Conserva los productos, sinónimos y dialectos que sufren cambios a través del tiempo, esto se torna a una aspecto parecido a lo que un data Warehouse hace por esencia: *Un Data Warehouse es, por definición, una base de datos histórica o temporal. A pesar de bases de datos OLTP, en casos excepcionales, podrán incluir piezas que capturan la historia, hay una diferencia significativa entre los dos, y esta diferencia se debe a la diferencia de perspectiva histórica: las bases de datos OLTP rara vez se capturan más de unas pocas semanas o meses de la historia, mientras que almacenes de datos debe ser capaz de capturar 3 a 5 hasta 10 años de historia, en el fondo de todos los datos que se registran en el almacén de datos. En un sentido, una base de datos histórica es una base de datos de dimensiones, la dimensión es el tiempo. En ese sentido, un modelo de datos históricos puede ser desarrollado utilizando un enfoque de modelado tridimensional*(Ballard, Herreman, Schau, Bell, & Kim, 1998).

La base de datos del lenguaje del sistema es una OLTP por el tipo de transacciones que realiza (agrega, modifica, elimina, etc.) se va a centrar en la transformación del lenguaje a través de los años, almacenando y haciendo un cierre de ciclo por año de los productos y pueda generar un histórico al estilo OLAP (On-Line AnalyticalProcess).

El sistema almacenará toda la información de los diferentes años del acontecimiento para recordarla en los momentos que sea necesario (explotación de datos de años anteriores, evolución del lenguaje, etc.), sin embargo, las búsquedas serán centradas por el año al que está en curso y las actualizaciones que se realizan agregarán los años en que lo hacen, todas aquellas palabras intactas, se entiende que no sufren modificaciones y se quedan en memoria y disponibles para el año en curso (ver Figura 19).

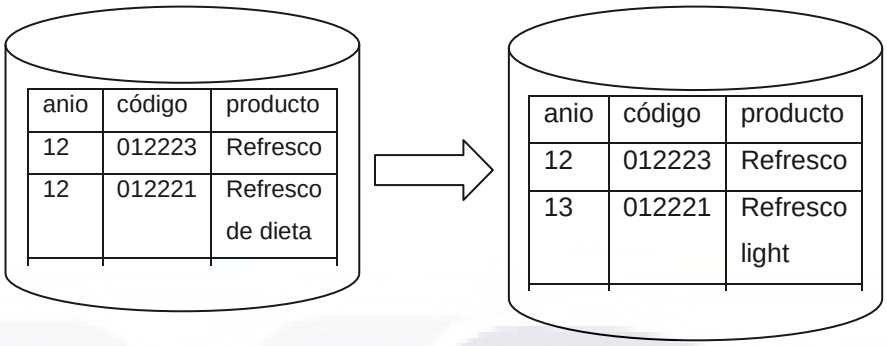


Figura 19 Base histórica del conocimiento lingüístico.

6.5 Interactividad con SEBS.

6.5.1 Búsqueda-Interactividad.

Rol: Usuarios interactivos (entrevistadores, REP, sociedad, etc.)

Módulo del sistema que permite la interactividad con el usuario (entrevistador) que por medio de la comunicación escrita (barra de búsqueda) hace la consulta del bien o servicio que desea saber su código (ver Figura 20).

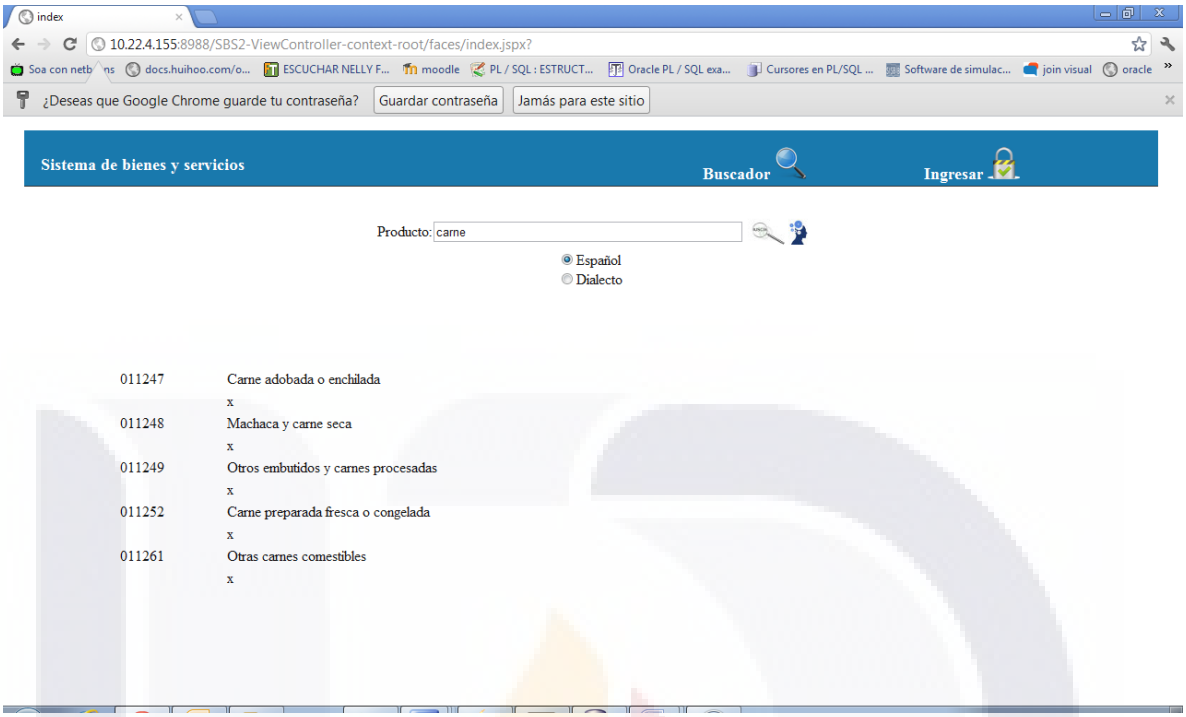


Figura 20 Interfaz de búsqueda e Interactividad.

La barra de búsqueda se puede apoyar de un diccionario integrado por el navegador, esto servirá para que recorra todas las palabras de dicho texto, una a una y en orden creciente y decreciente, buscándola primero en el diccionario principal (versión abreviada del diccionario académico del español) y después y luego en el secundario (palabras que no figuran en el diccionario principal y que cada usuario va introduciendo según sus necesidades). Se trata, por tanto, cotejar o confrontar el texto con lo que contiene sus diccionarios(Ariza García & Tapía Poyato, s. f.).

Internamente el sistema funciona a base de hilos, aplicando los algoritmos de Tecnologías del Lenguaje, acudiendo a la Base de Datos (lenguaje del sistema) para devolverle una petición del usuario, mostrándole en pantalla todas las posibles coincidencias que este encuentra.

Si no encuentra peticiones en la búsqueda, nos sugiere opciones de productos que lo relaciona con palabras similares a la que está consultando.

Cuando no encuentra el resultado esperado por el usuario, es cuando entra la interactividad con el usuario, ya que él mismo es el que enseña la palabra al sistema para canalizar la palabra por los administradores. Al enseñar al sistema la palabra, pide datos

extras para una mejor comprensión y canalizarla de la mejor manera, esto derivado de que cada entidad tiene terminología propia (ver Figura 21).

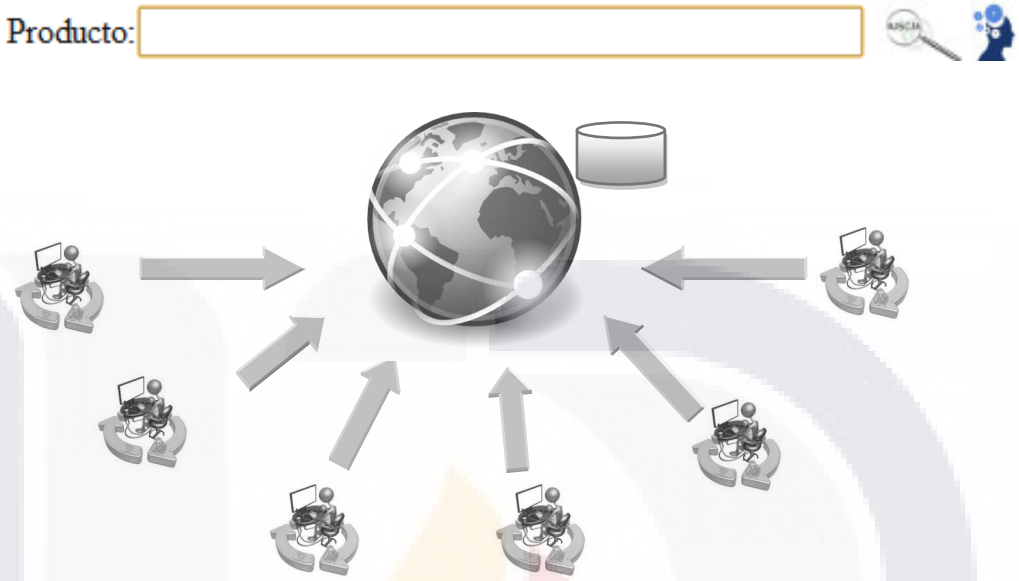


Figura 21 Interactividad con usuarios

6.5.2 Herramienta Administrativa.

Rol: Dueño del Proceso

La herramienta administrativa es la que permite el funcionamiento de la distribución de los administradores del catálogo y sus cargas de trabajo, tales como inactividad de ellos ya sea por enfermedad, inasistencia, etc., de esta manera permite que las cargas de trabajo se distribuyan con los que están presentes. También permite que los inactivos con cargas de trabajo puedan quitarles sus cargas de trabajo para que la distribuya entre los administradores activos (ver Figura 26).

6.5.2.1 Motor de Carga de Trabajo.

Carga Automática:

Será una carga equitativa entre los Administradores que estén activos y el acumulado de cada hora de las solicitudes entrantes (ver Figura 23).

Carga de trabajo= Número de solicitudes/usuarios activos.

Repartir carga de trabajo

Si número de solicitudes **- (**Carga de trabajo*usuarios activos**)** !=0 **entonces**

Escoge a la diferencia en usuarios (ordenados por el menor carga de trabajo) para repartir la carga equitativamente

Figura 22 Pseudocódigo de la carga automática.

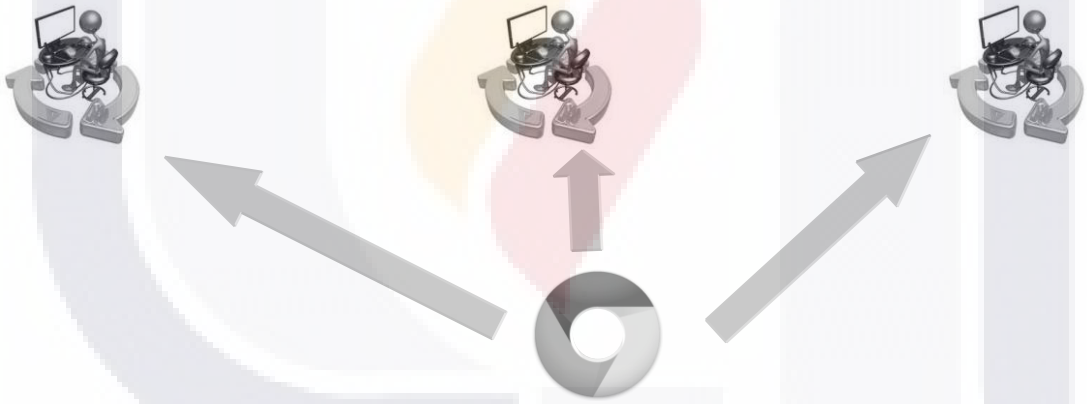


Figura 23 Asignación de carga automática.

La carga automática puede ser activa de forma manual, esto es debido a que si la asignación de cada hora es insuficiente por la demanda pueda hacerse cargas antes de dicho intervalo de tiempo.

Reasignación de Carga:

Para hacer una reasignación de carga de trabajo de todos aquellos usuarios inactivos que quedaron con alguna carga pendiente. La carga será equitativa entre los usuarios activos (ver Figura 25).

Reasignación= Carga de trabajo de usuarios inactivos/usuarios activos.

Repartir carga de trabajo

Si número de solicitudes - (Reasignación*usuarios activos) !=0 **entonces**

Escoge a la diferencia en usuarios (ordenados por el menor carga de trabajo) para repartir la carga equitativamente

Figura 24 Pseudocódigo de la reasignación de carga.

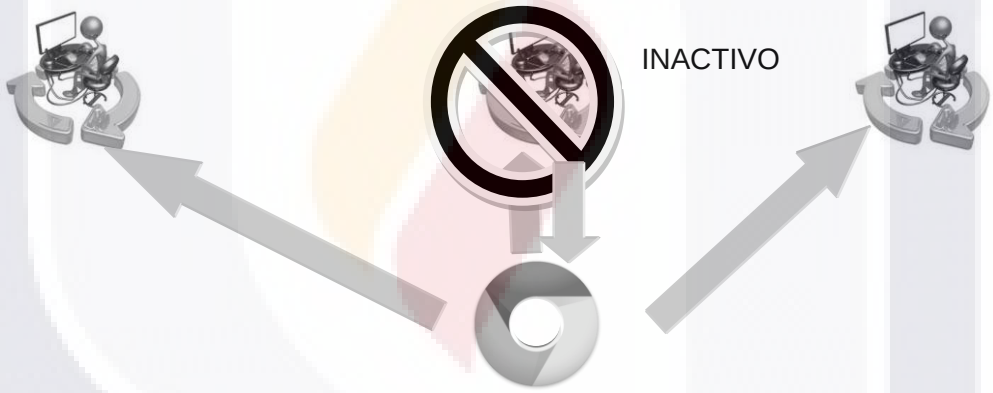


Figura 25 Reasignación de carga de trabajo.

Número de solicitudes: conteo de la tabla “TR_SOLICITUD”

Usuarios Activos: Los usuarios con id_status=1 (BD: TR_TRABAJADOR, TC_STATUS)

Carga de Trabajo: Las cargas que se hacen (BD: TR_CARGA_TRABAJO)

Usuarios Inactivos: Los usuarios con id_status=2 (BD: TR_TRABAJADOR, TC_STATUS)

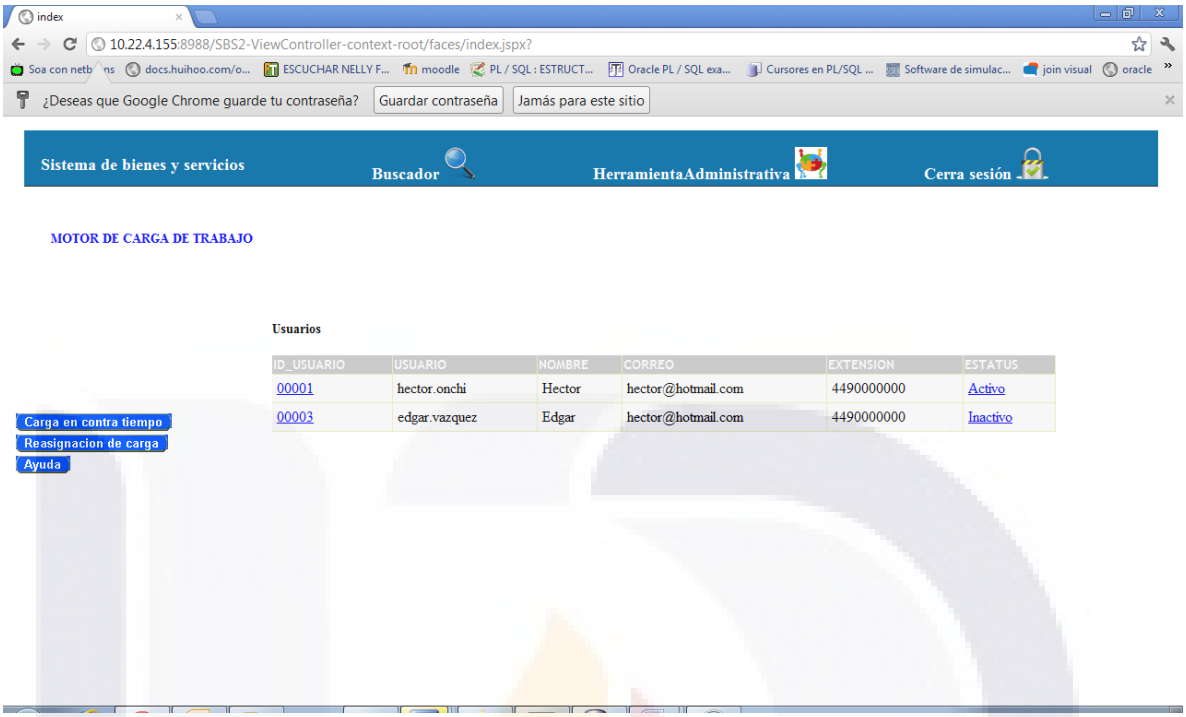


Figura 26 Interfaz Administrativa del SEBS.

Status: Podrá ser Activado o Desactivado en la misma tabla de trabajo de tal manera que puede hacer juego de hacer cargas inhibiendo usuarios o también haciendo reasignación de cargas.

6.5.3 Canalizador.

Este módulo, es generado como la influencia de los padres a sus hijos sobre lo que va aprendiendo en el entorno, de lo que es verdad, y lo que no, cuales palabras son las adecuadas y enseñarle que significan las que son nuevas, de igual manera el modulo del canalizador es la interfaz que permitirá a los administradores del catálogo hacer esa parte del tutor hacia el sistema que enseñara en qué consiste cada producto de bien y servicio (ver Figura 27).

Figura 27 Interfaz del Canalizador.

Rol: Administrador del catálogo.

Actualizar: En esta sección se actualizará el catálogo para todos lo que usan su información compartida (XML) y queda disponible en la web para su explotación (ver Figura 28).

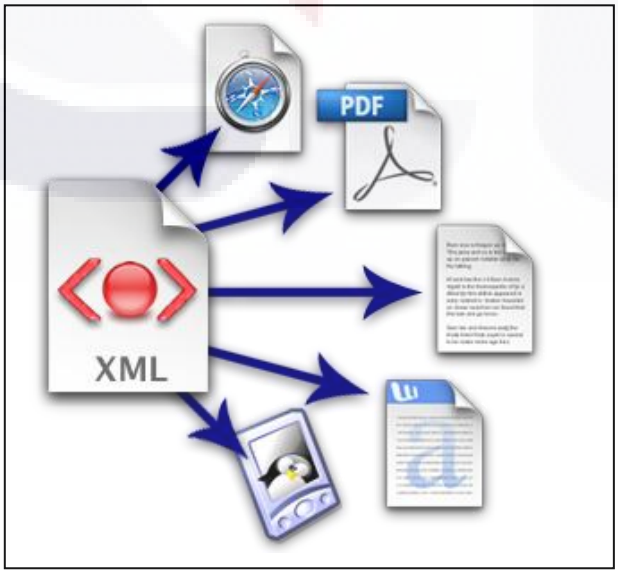


Figura 28 Compatibilidad de XML.

El archivo XML podrá ser utilizado por las tecnologías como:

- Java.
- .NET Microsoft.
- HTML
- PHP
- iOS
- Android
- Blackberry, etc.

De tal manera que es la opción de intercambiar la información entre diferentes áreas de trabajo interesadas en dicha información.

Con este archivo se podrá llegar a niveles tan extensos desde un nivel de procesos, como hasta los niveles de comparativita internacional, de tal manera que sea útil para los interesados en la información y sea una referencia válida y confiable con el fundamento apropiado (ver Figura 29).

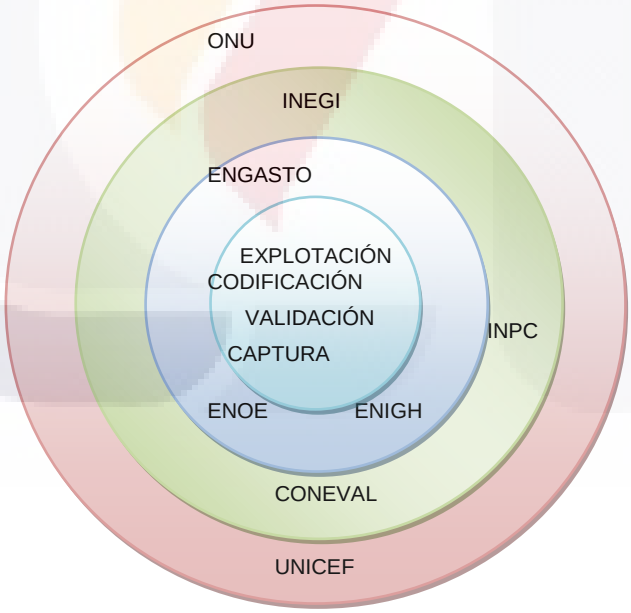


Figura 29 Expansión del conocimiento del SEBS

Buscar: Accede al buscador normal sin la interacción de alimentar el sistema.

Ayuda: En la ayuda encontrará apoyo para manipular el sistema, junto con observaciones que ellos sugieren.

Status: El estatus será un combo para que pueda ser modificado conforme el administrador del catálogo lleve su seguimiento.

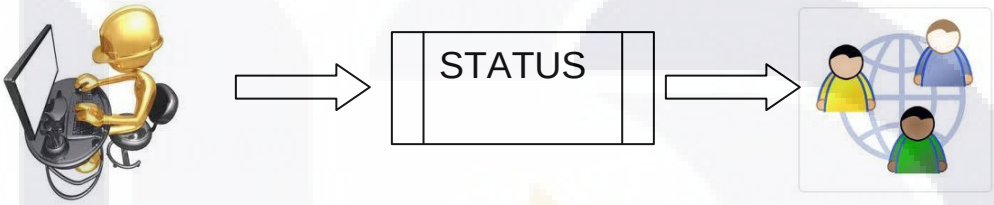


Figura 30 Publicación del estatus a los usuarios

El status es publicado en la Web a los entrevistadores para que conozcan en que proceso de conocimiento del lenguaje está la solicitud (ver Figura 30). Los estatus de atención por parte del administrador son:

1. ***Reportado***: el momento que el usuario alimenta al sistema del producto que no encontró.
2. ***En carga de trabajo***: Cuando el motor de carga de trabajo asigna a cada administrador de catálogo sus responsabilidades por atender.
3. ***En proceso de atención***: El administrador de catálogo anuncia que está por atender dicho producto reportado.
4. ***Sin respuesta***: Los administradores del catálogo están sin conocimiento del producto que se reporta y entrarán en una investigación futura.

6.5.3.1 Desarrollo del Lenguaje.

En esta sección dota al sistema del conocimiento. Ubica los nuevos productos, sinónimos y dialectos que quiere que el sistema tenga en su lenguaje (ver Figura 31).

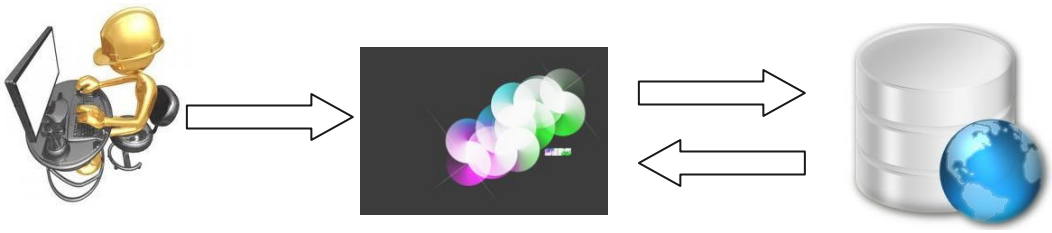


Figura 31 Desarrollo del Lenguaje.

Al dar clic en una de las secciones del desarrollo del Lenguaje, entraremos a la interfaz que nos permitirá canalizar en que parte va, si pertenece a un “Producto” del maternés, si pertenece a un “sinónimo” (otra forma de llamar al producto, alguna marca comercial) o si es un dialecto (ver Figura 32).

Figura 32 Interfaz de selección del lenguaje.

Cuando pasa el caso que se le dé un dialecto, se abre la opción para indicar a que dialecto pertenece y poder asignarlo a la familia al que pertenece hasta que se remita a un único producto (código) en el lenguaje español al que pertenece (ver Figura 33).

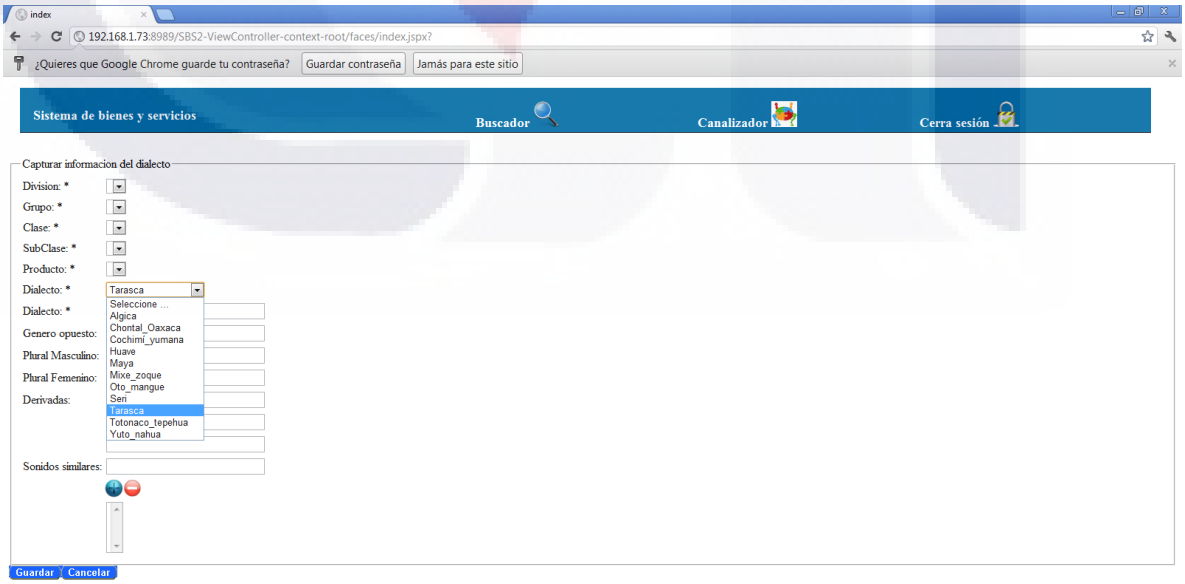


Figura 33 Interfaz de desarrollo lingüístico en dialectos

Si se selecciona la parte de Productos, Sinónimo o Dialecto se extenderán combos con las familias a la que quieren canalizarlo:

- División.
- Grupo.
- Clase.
- Subclase.
- Producto.

Si es en la parte de productos, lo que canaliza en los combos se extienden únicamente a la subclase que pertenece (ver Figura 34).

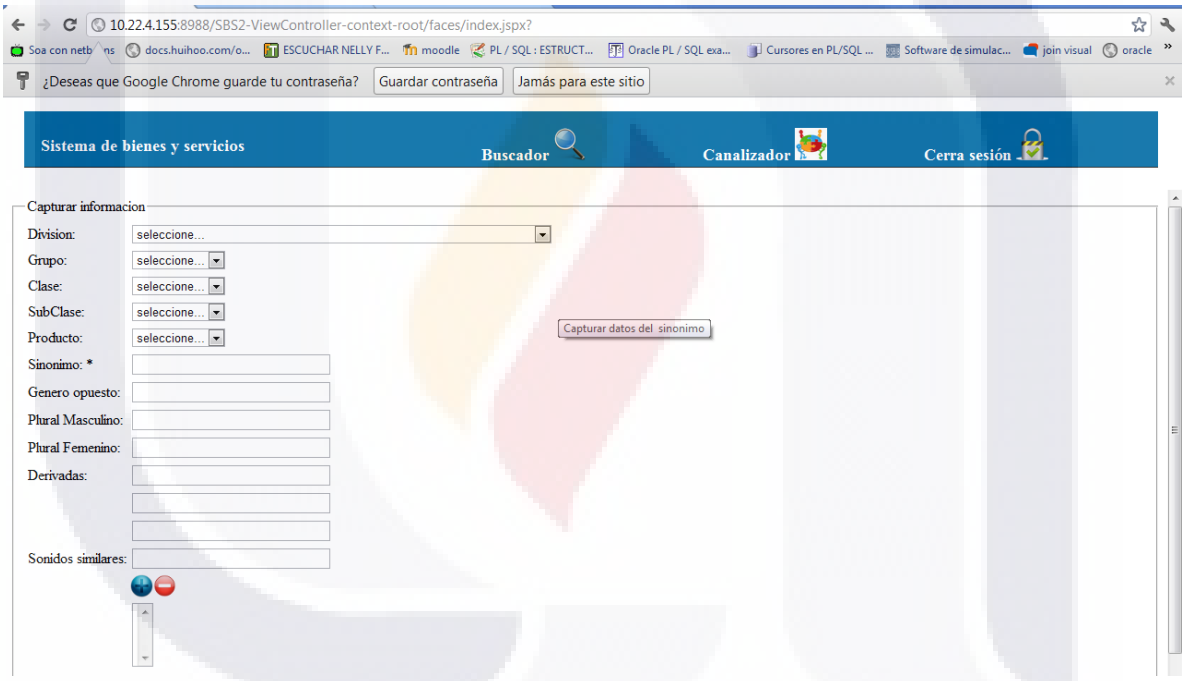


Figura 34 Interfaz del desarrollo del conocimiento.

En el módulo donde canaliza el producto, tiene la opción de dotar el conocimiento en lo más preciso que sea posible, ya que facilitará y usará todas las partes del algoritmo de búsqueda entregue un mejor resultado sirviendo a los usuarios al momento de hacer sus búsquedas. La precisión llega hasta lematizar (palabras derivadas) los productos.

Exclusión.

Al guardar el nuevo producto, sinónimo o dialecto, sigue la parte de exclusión. Exclusión sirve para dotar de un conocimiento preciso al sistema, que ubica los productos que se parecen pero son de familias diferentes (ver Figura 35).

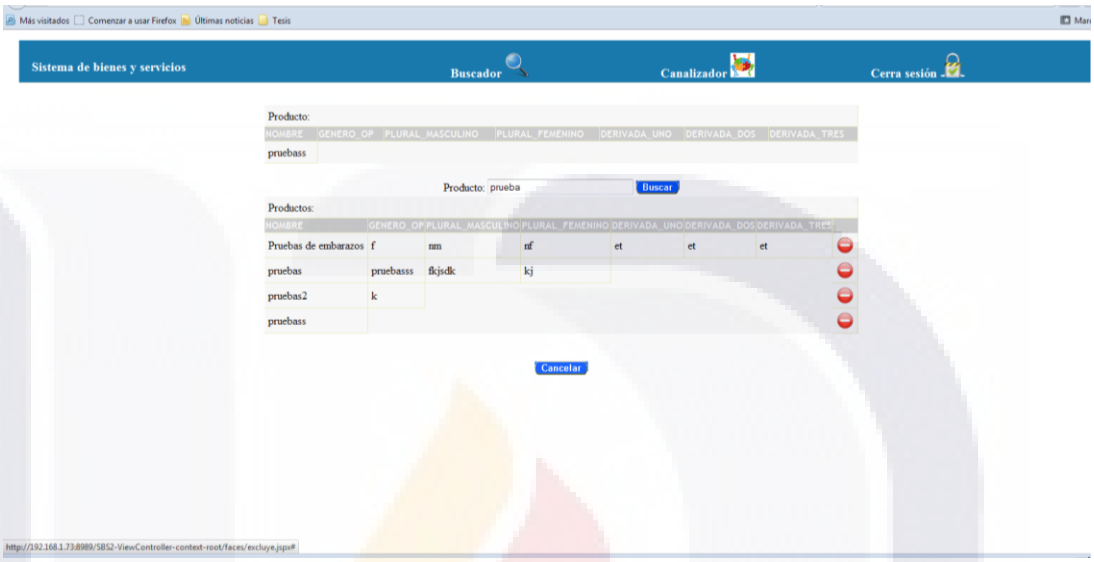


Figura 35 Interfaz de exclusión de productos.

Con esto logramos que cuando el usuario final haga una búsqueda, le dé información extra que permita visualizar un panorama más amplio de lo que va a codificar.

6.5.3.2 Corregir Conocimiento.

Obedece a las correcciones que se quieren hacer sobre un aprendizaje que se haya obtenido sea el caso de un producto, sinónimo o dialecto incluyendo sus fonéticas. (BD: TC_PRODUCTO, TR_SINONIMO, TR_*Tablas dialectos.*) (ver Figura 36).

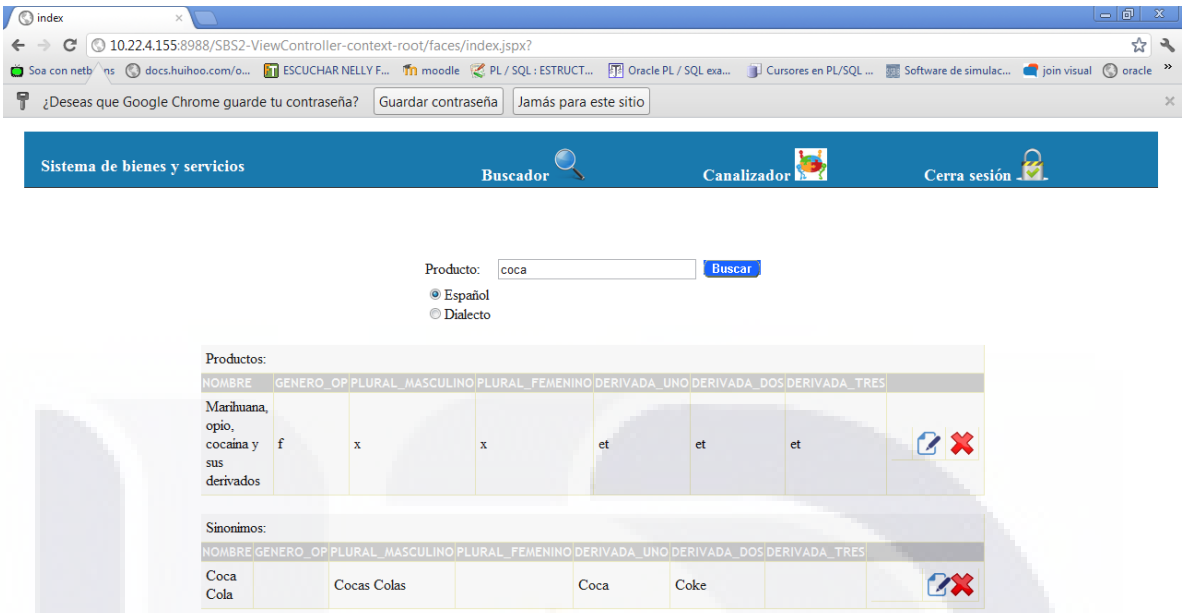


Figura 36. Interfaz de corrección del conocimiento.

La **lista de producto** son los resultados que se obtuvieron en la búsqueda y que esta alimenta a la lista de sinónimos que le pertenecen a ese producto.

La lista de fonética se alimenta conforme se selecciona el producto y el sinónimo, estas tablas se borran y editan a tiempo real.

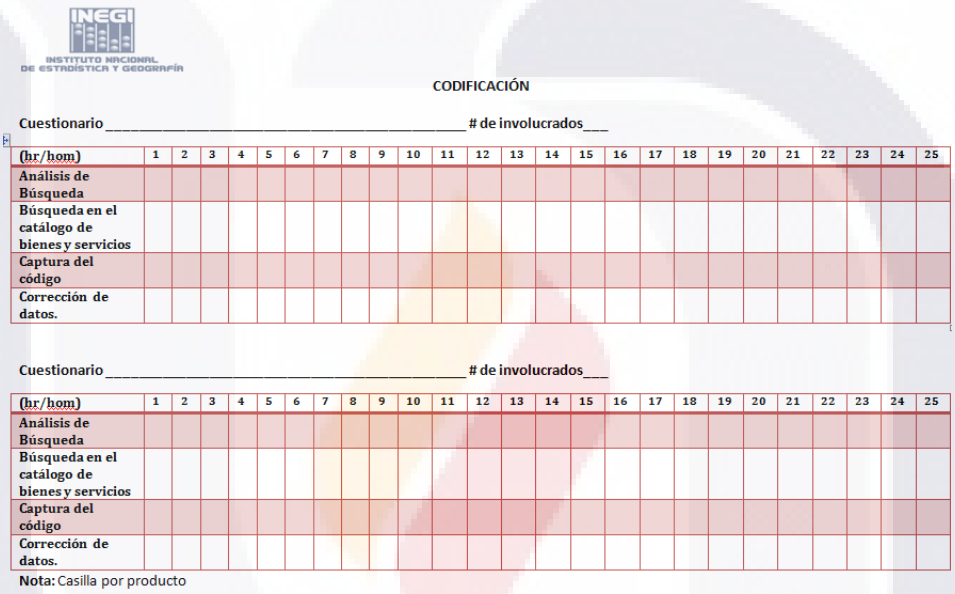
La edición de productos es la interfaz como en la que se da lenguaje al sistema teniendo los datos cargados para su corrección o aportación de lenguaje.

Basura: Borra de su memoria un producto, sinónimo o dialecto (BD: TC_PRODUCTO, TR_SINONIMO, TR_*Tablas dialectos.*)

7. RESULTADOS

Para medir los tiempos se viajó nuevamente a la ciudad de Mérida, ya que ahí se realizó la primera toma de tiempos que se hacía de forma Manual que se platica en la problemática.

Se tomó al azar una muestra de cuestionarios de todos los entrevistadores que están encuestando en ese momento y que tuvieran cuestionarios disponibles para codificar. Se usó un formato que permitió reflejar el acontecimiento (ver Figura 37).



INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

CODIFICACIÓN

Cuestionario _____ # de involucrados___

(hr/hom)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Análisis de Búsqueda																									
Búsqueda en el catálogo de bienes y servicios																									
Captura del código																									
Corrección de datos.																									

Cuestionario _____ # de involucrados___

(hr/hom)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Análisis de Búsqueda																									
Búsqueda en el catálogo de bienes y servicios																									
Captura del código																									
Corrección de datos.																									

Nota: Casilla por producto

Figura 37 Formato de captación de tiempos de codificación.

Una vez que se tuvo la muestra de cuestionarios, se tomó un cuestionario para medir los tiempos que hace el entrevistador en las 3 actividades que realiza para codificar (Análisis de Búsquedas, Búsqueda en el catálogo (Sistema SEBS), tiempo promedio de captura en el cuestionario).

El tiempo se midió de uno en uno de productos tomando los primeros 25 productos que se encuentren en el cuestionario (en algunos casos la muestra de 25 productos es el total que se tienen en el cuestionario). Después de la muestra, se sigue tomando el tiempo hasta que concluya de codificar todo el cuestionario para poder generar el tiempo promedio de codificación por cuestionario. Este proceso se repite en cada uno de los diferentes tipos de cuestionarios que se tomaron al azar (ver Anexo No. 2).

Los resultados fueron favorables en cuestión de los tiempos de codificación, en los cuestionarios que se tomaron como muestra, se afirma que ayudan a la codificación de productos.

Hay situaciones en que alguna de las actividades no se ve reflejado un resultado favorable si no al contrario, un número negativo, esto es debido a que pasaron situaciones externas como “platica entre ellos”, “cansancio de que venían de su zona de encuesta”, “dudas al usar el sistema” pero aun así al conjuntar las actividades para hacer el proceso de codificación, se observan cifras positivas que permiten afirmar que los tiempos disminuyeron.

7.1 Cuestionario Trimestral

Comenzando por el cuestionario trimestral el promedio neto de la codificación de tiempo por cuestionarios anteriormente estaba en los 7.8 min; en el reproceso se aprecia una baja de tiempo a 5.58 min por cuestionario (ver tabla 4), estamos hablando que se logró bajar un 28.54% del tiempo que se realizaba anteriormente. En las actividades de “Búsqueda en el catálogo” y “Captura de código” tuvo un descenso considerable del 2.84% y 4.87% respectivamente por producto de codificación (ver Figura 38).

TIEMPO MIN/HOMBRE PROCESO MANUAL		TIEMPO MIN/HOMBRE RE-PROCESO (AUTOMATIZACIÓN)		% IMPACTO
Tiempos Promedios (min/hombre) por producto		**Tiempos Promedios (min/hombre) por producto**		
Tiempo Promedio Análisis de Búsqueda	0.83	Tiempo Promedio Análisis de Búsqueda	1.05	-27.13
Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	4.11	Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	3.99	2.84
Tiempo promedio de captura del código	0.56	Tiempo promedio de captura del código	0.53	4.87
Tiempo promedio por cuestionario	**7.80**	**Tiempo promedio por cuestionario**	**5.58**	28.54

Tabla 4 Análisis de tiempos del cuestionario trimestral

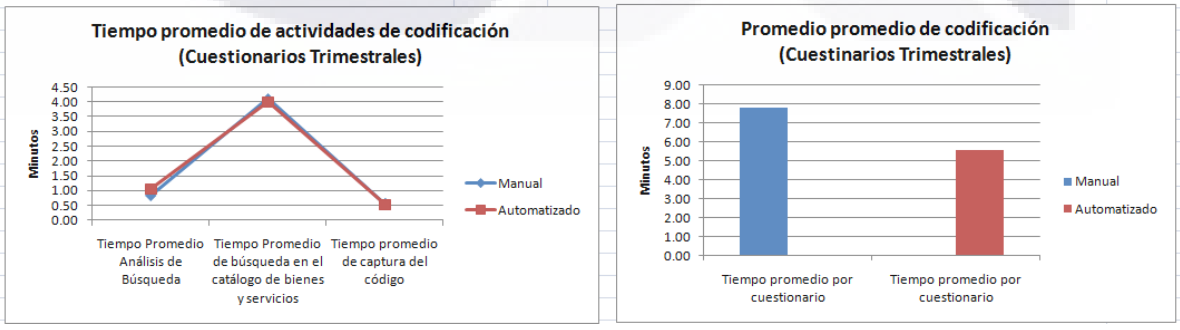


Figura 38 Gráficas comparativas del cuestionario trimestral

7.2 Cuestionario de gastos en el hogar

El cuestionario de gastos del hogar el promedio neto del tiempo por codificar los cuestionarios anteriormente estaba en los 43.78 min; en el reproceso se aprecia una baja de tiempo a 28.93 min por cuestionario (ver tabla 5), estamos hablando que se logró bajar un 33.92% del tiempo que se realizaba anteriormente. En las actividades de “Análisis de Búsqueda” y “Búsqueda en el catálogo” tuvo un descenso considerable del 7.66 % y 30.95% respectivamente por producto de codificación (ver Figura 39).

TIEMPO MIN/HOMBRE PROCESO MANUAL		TIEMPO MIN/HOMBRE RE-PROCESO (AUTOMATIZACIÓN)		% IMPACTO
Tiempos Promedios (min/hombre) por producto		**Tiempos Promedios (min/hombre) por producto**		
Tiempo Promedio Análisis de Búsqueda	1.30	Tiempo Promedio Análisis de Búsqueda	1.20	7.66
Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	5.83	Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	4.02	30.95
Tiempo promedio de captura del código	1.41	Tiempo promedio de captura del código	1.46	-3.64
Tiempo promedio por cuestionario	**43.78**	**Tiempo promedio por cuestionario**	**28.93**	33.92

Tabla 5 Análisis de tiempos del Cuaderno de Gastos del Hogar

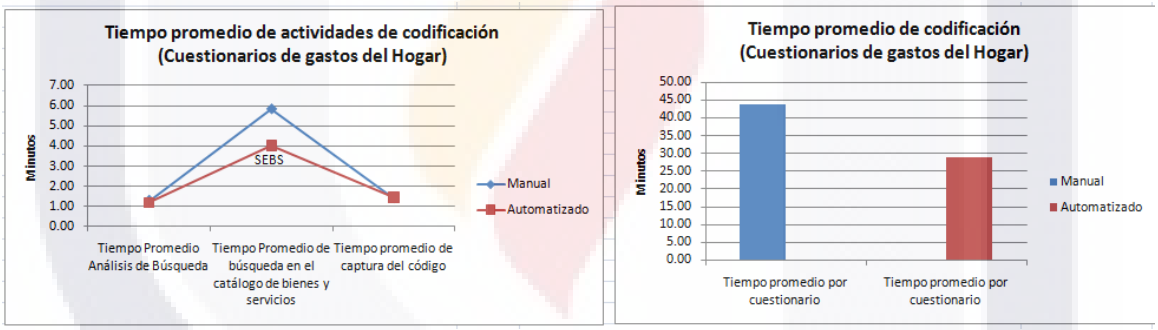


Figura 39 Gráficas comparativas del Cuaderno de Gastos del Hogar

7.3 Cuestionario de Gastos individuales

En el cuestionario de gastos individuales se aprecia un impacto importante en el reproceso, donde, el promedio neto del tiempo por codificar los cuestionarios anteriormente estaba en los 13.45 min; en el reproceso se aprecia una baja de tiempo a 5.09 min por cuestionario (ver Tabla 6), estamos hablando que se logró bajar un 62.18% del tiempo que se realizaba anteriormente. En las actividades de “Búsqueda en el catálogo” y “Captura de código” tuvo un descenso considerable del 78.19 % y 55.05% respectivamente por producto de codificación (ver Figura 40).

TIEMPO MIN/HOMBRE PROCESO MANUAL		TIEMPO MIN/HOMBRE RE-PROCESO (AUTOMATIZACIÓN)		% IMPACTO
Tiempos Promedios (min/hombre) por producto		**Tiempos Promedios (min/hombre) por producto**		
Tiempo Promedio Análisis de Búsqueda	0.94	Tiempo Promedio Análisis de Búsqueda	1.15	-21.41
Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	9.51	Tiempo Promedio de búsqueda en el catálogo de bienes y servicios	2.07	78.19
Tiempo promedio de captura del código	1.82	Tiempo promedio de captura del código	0.82	55.05
Tiempo promedio por cuestionario	**13.45**	**Tiempo promedio por cuestionario**	**5.09**	62.18

Tabla 6 Análisis de tiempos de la libreta de gastos individuales

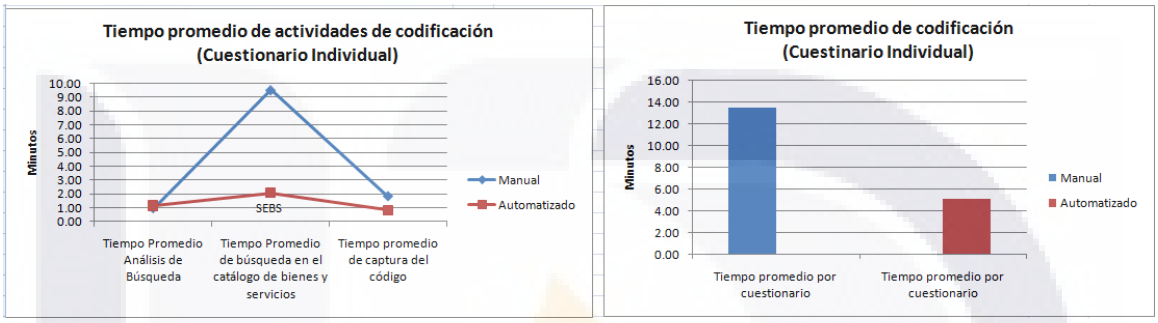


Figura 40 Gráficas comparativas de la libreta de gastos individuales

En cuanto a los errores al momento de codificar, también tuvo un impacto importante, se logró reducir el porcentaje. En los cuestionarios de Gastos del Hogar anteriormente estaba en 1.68% de error (muestra de 954 productos) en donde reduce a un 0.74 %, en la libreta de gastos individuales se logra un descenso a 2.40% de 3.37% y en el cuestionario de gastos trimestrales que anteriormente tenía el 4.4 % de error se reduce a 2.19% (ver Tabla 7).

TIEMPO MIN/HOMBRE PROCESO MANUAL				
	Cuaderno de Gastos del Hogar	Libreta de Gastos individuales	Cuestionario de Gastos Mensuales	Cuestionario de Gastos Trimestrales
Códigos Incorrectos	9/954	3/208	1/83	7/250
Concepto fuera de Catálogo	2/954	0	0/83	4/250
Concepto con modismo	5/954	4/208	0/83	0/250
Concepto en dialecto	0	0	0/83	0/250
Total codificación no acertada	**16/954**	**7/208**	**1/83**	**11/250**
Porcentaje de error	**1.68**	**3.37**	**1.20**	**4.4**
TIEMPO MIN/HOMBRE RE-PROCESO (AUTOMATIZACIÓN)				
	Cuaderno de Gastos del Hogar	Libreta de Gastos individuales	Cuestionario de Gastos Mensuales	Cuestionario de Gastos Trimestrales
Códigos Incorrectos	4/950	5/208	0/35	3/137
Concepto fuera de Catálogo	0	0	0/35	0/137
Concepto con modismo	3/950	0	0/35	0/137
Concepto en dialecto	0	0	0/35	0/137
Total codificación no acertada	**7/950**	**5/208**	**0/35**	**3/137**
Porcentaje de error	**0.74**	**2.40**	**0**	**2.19**

Tabla 7 Análisis de tiempos del proceso manual VS re-proceso

Como resultado podemos observar que el re-proceso (automatización) está siempre por debajo del manual (ver Figura 41) esto quiere decir que en todos los cuestionarios es más eficiente y eficaz el SEBS basado en Lingüística computacional.

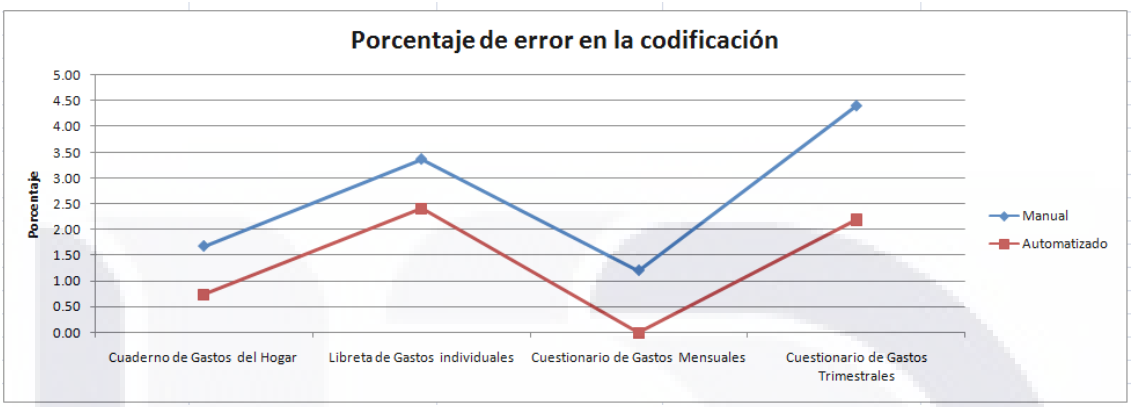


Figura 41 Gráfica del proceso manual VS re-proceso

Los usuarios finales también evaluaron el SEBS logrando medir así la satisfacción por parte de ellos (ver Anexo No. 3). Hacen mención el 100% que aprecian la reducción de tiempo para las labores cotidianas de codificación, el 75% aprecia que el sistema proporciona los códigos con mayor facilidad que el catálogo impreso y que ayuda al margen de error (ver Figura 42).

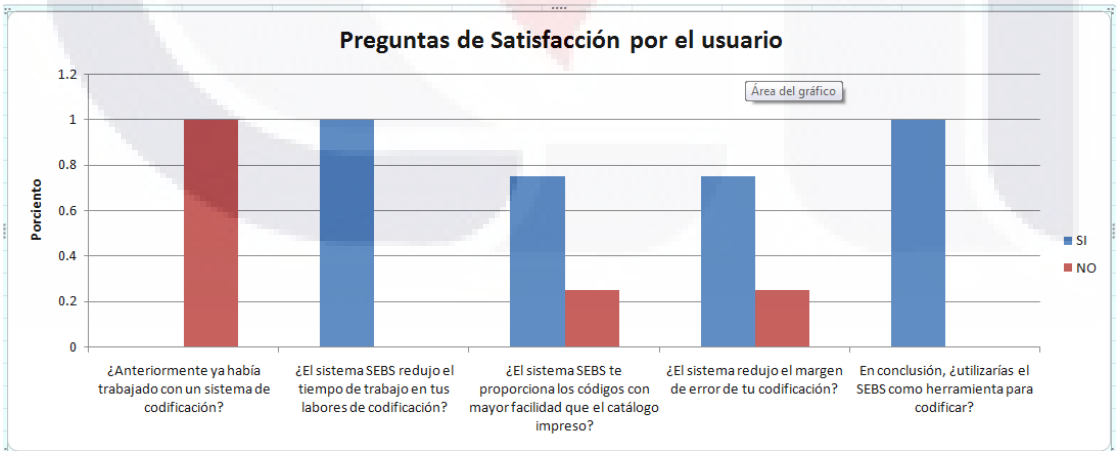


Figura 42 Gráficas de satisfacción por parte del usuario

Al mencionar al usuario que seleccione una de las herramientas para codificar, entre el catálogo impreso y el SEBS, el 100% prefiere el sistema. También el 100% está

convencido de que el SEBS es un sistema “bueno” para las actividades cotidianas de codificación (ver Figura 43).

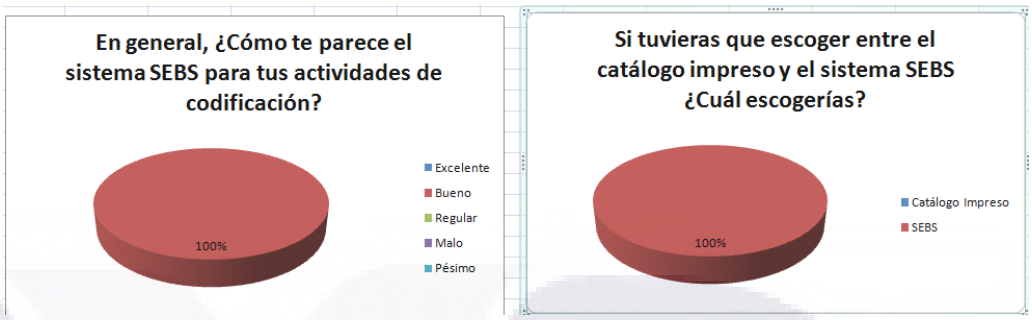


Figura 43 Gráfica de apreciación del SEBS

El 100% de los usuarios están dispuestos a cambiar el catálogo impreso que tienen al SEBS (ver Figura 43).

8. CONCLUSIONES.

Esta investigación sirvió para diseñar y aplicar un servicio informático especializado para la codificación de los productos de la encuesta ENGASTO, de tal manera que sea escalable y crezca tanto como lo es la lingüística Nacional, teniendo la capacidad de codificar el mayor número de productos con un menor margen de error, donde el margen de error está siempre por debajo del proceso manual.

Las actividades de codificación de los entrevistadores se vio facilitada con la reducción de tiempos y errores, pero aun así, un cuestionario se les aplico para saber cómo apreciaban el sistema, que a pesar de los números nos indican que si cumplió el objetivo general, esto nos sirvió para confirmar que los entrevistadores perciben y aprecian dicho re-proceso.

Nuestro objetivo general es que se diseñara y probara un sistema que pudiera mejorar los indicadores de operación, tal es el caso que el tiempo de ejecución promedio por cuestionario, siempre está por debajo del proceso manual, observándose en el cuestionario que más consume tiempo (cuaderno de gastos del hogar) que se logró bajar un 33.92% del tiempo utilizado anteriormente. En cuanto a nuestro indicador de número de errores, en todos los cuestionarios están por debajo de los errores que se cometían por el entrevistador (desempeño humano), debido a que este sistema centra todos los productos a una división y en específico a un producto no importando de qué manera lo llamen.

La información está siendo actualizada y cargada en el SEBS de tal manera que la etapa de explotación de la encuesta ENGASTO pretende hacer uso de esta, con el fin de indicar un nombre oficial a los códigos que se encuentran en la Base de Datos de la etapa de validación, logrando estadísticos publicado con nombres únicos de productos.

Nos ha quedado claro que aplicar nuevas tecnologías lingüísticas nos proporcionará oportunidades de mejorar la codificación facilitando la ejecución de los procesos que conforma la encuesta, logrando mayor precisión de las salidas de cada uno de estos, hasta un cambio radical de ejecutar un proceso manual a uno automatizado.

Las tecnologías del Lenguaje Humano es la teoría fundamental que nos permitió desarrollar el sistema, donde la parte del análisis morfológico, la lematización, una ligera normalización de los datos y la parte de la fonética que está solamente simulada, nos

brindan resultados muy satisfactorios en la búsquedas de códigos para los productos de la encuesta. Logrando con ello la reducción considerable del tiempo que se llevaba al codificar.

De la implementación de esta investigación y desarrollo, podemos decir que el sistema SEBS (Sistema de Entendimiento de Bienes y Servicios) tendrá un grado mayor de precisión en la búsqueda entre más tiempo pase e interactúe con los entrevistadores, quiere decir, que este sistema va ir almacenando día con día nuevas palabras, nuevas formas de decir a un producto, que provenga de ellos, de tal manera que, la rica lingüística Nacional que existe, es centralizada a un código que nos define a un producto que nos estamos refiriendo, afirmando que entre mayor conocimiento tenga el sistema, menor el margen de error en la codificación.

Con esto se plantea la posibilidad de que la entrada de datos ya venga codificada con cierto grado de calidad y facilite a los procesos consiguientes, tanto en la interpretación, generación de información, etc.

El SEBS nos permitió hacer una aportación en la reingeniería de procesos, ya que como se vio durante este trabajo, el proceso contaba con roles repetitivos. En el proceso rediseñado el usuario final (entrevistador) será directamente quien enseñe al sistema haciendo el llamado a Oficinas centrales (INEGI Aguascalientes) y ellos canalizando los productos que van siendo enviados por todos los usuarios, entonces, los roles que antes participaban en este proceso, quedan totalmente libre de ese tiempo y codificación no es una causa de saturación de trabajo para los roles que ayudan al entrevistador.

Además, se descubrió al momento de que estuvo en marcha el sistema que nuestros usuarios finales (entrevistadores) tuvieran una reacción positiva, teniendo al 100% dispuesto a cambiar del catálogo impreso al SEBS, de tal manera que se apreció que la parte de “resistencia al cambio” está prácticamente nula, facilitando la implementación a nivel Nacional.

La encuesta es continua y permanente de tal manera que con este avance en la codificación de productos en la encuesta ENGASTO, será implementado en los siguientes operativos estableciéndose como permanente y se use para todos los estados de la República Mexicana.

8.1 Trabajos a Futuro.

El mercado de la telefonía móvil sigue creciendo de manera imparable especialmente los dispositivos con tecnologías integradas con WiFi y VoIP. El sector que más rápidamente está creciendo en el mercado es el teléfono dual con WiFi y VoIP o voz sobre protocolo de Internet, es decir, los que permiten conversaciones a través de Internet otra red basada en IP (protocolo de Internet). Todos estos avances nos permiten oportunidades para desarrollar al SEBS en diferentes tecnologías que mencionamos a continuación:

1. Aplicación web para móvil: la oportunidad de desarrollar el sistema de tal manera que al ingresar desde un teléfono móvil a la URL pueda ser usado como si estuviera frente un ordenador, logrando con ello que sea más portable y los entrevistadores puedan codificar desde cualquier tecnología de conexión a internet (3G, wifi, etc.).

2. Aplicación para Smartphone (offline): se desarrollara una aplicación de tal manera que sea portable con los teléfonos con SO, tal sea el caso Android, iOS, Blackberry, etc. de tal manera que sea tan portable que aún estando en lugares sin conectividad a internet, puedan hacer uso del sistema SEBS para realizar su actividades de codificación en cualquier lugar y la información se actualice cada vez que el Smartphone tenga alguna conectividad con internet.

3. Desarrollo del nivel fonético real: las tecnologías del lenguaje mencionan de la comunicación hablada entre ordenador -- humano, brindando la oportunidad de agregar este nivel al SEBS generando un sistema de dialogo para que el entrevistador busque oralmente un producto, la computadora le dicte el código de dicho producto, de esta manera, se ahorrara tiempo de consulta escrita hacia el ordenador y eliminar errores de vista cuando se copia el código de la pantalla.

GLOSARIO

Bienes y Servicios: Se refiere a todos los productos que son consumidos por los integrantes de un hogar (por ejemplo, 2 refrescos de cola).

Libro Catálogo*:* Documento impreso que sirve como insumo para hacer la búsqueda de los códigos de los productos de forma manual.

ISE: Instructor Supervisor Estatal, cuyo finalidad es la coordinación y supervisión de los entrevistadores que cumplan con sus funciones.

REP: Responsable Estatal de Proyecto, cuya finalidad es la coordinación de todo el proyecto en el estado correspondiente.

INEGI: Instituto Nacional de Estadística y Geografía (Dependencia Federal).

Etapa de validación: Es la etapa encargada de procesar que los datos tengan congruencia con el contenido a lo largo de toda la encuesta del hogar (por ejemplo: un hombre no puede tener hijos) y que los datos estén limpios de cualquier dato que no aporte valor (por ejemplo un "#" en el registro del nombre de una persona).

Etapa de explotación: Es la etapa encarga de generar estadísticas oficiales de los gastos, ingresos, condiciones de vida, etc. generando información para los diferentes sectores federales.

BD: Abreviación del término "Base de Datos"

JavaServerPage: Tecnología de java que permite hacer páginas web dinámicas

Hilo: Un hilo es una secuencia de código en ejecución dentro del contexto de un proceso. Los hilos no pueden ejecutarse ellos solos; requieren la supervisión de un proceso padre para correr. Dentro de cada proceso hay varios hilos ejecutándose.

BIBLIOGRAFÍA

Ariza García, A., & Tapía Poyato, A. M. (s. f.). El corrector ortográfico y la presentación del texto escrito. *Universidad de Sevilla*.

Ballard, C., Herreman, D., Schau, D., Bell, R., & Kim, E. (1998). *Data Modeling Techniques for Data Warehousing* (1° ed.). San Jose, California: IBM Corporation.

Bodomo, A. (2006). Human Language Technology in Multilingual Perspectives: Institutions and Applications. *International Journal of Technology and Human Interection*.

Centros Europeos de Empresas Innovadoras de la Comunidad Valenciana (CEEI CV). (2008). Manual de Reingeniería de Procesos. Centros Europeos de Empresas Innovadoras de la Comunidad Valenciana (CEEI CV).

Domínguez Burgos, A. (2002). Lingüística computacional: un esbozo. *Boletín de Lingüística*, *18*.

Ferrero, P., & Alda, J. (2007). Aplicación de herramientas de lingüística computacional en foros virtuales. *III Jornada Campus Virtual UCM*.

Gelbukh, A., & Sidorov, G. (2006). *PROCESAMIENTO AUTOMÁTICO DEL ESPAÑOL CON ENFOQUE EN RECURSOS LÉXICOS GRANDES* (1.ª ed.). México: INSTITUTO POLITÉCNICO NACIONAL.

Instituto Nacional de Lenguas Indígenas. (2008). CATÁLOGO DE LAS LENGUAS INDÍGENAS NACIONALES: VARIANTES LINGÜÍSTICAS DE MÉXICO CON SUS AUTODENOMINACIONES Y REFERENCIAS GEOESTADÍSTICAS.

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing An Introduction to Natural Language Processing, computational Linguistics, and Speech Recognition* (2° ed.). New Jersey: Prentice Hall.

Lawler, J. M., & Dry, H. A. (1998). *Using computers in linguistics*. London: Routledge.

Llisterri, J. (2003). Lingüística y tecnologías del lenguaje. *Panorámica de Estudios Lingüísticos*.

Manganelli, R. L., & Klein, M. M. (2004). *Cómo hacer reingeniería*. Norma.

Naciones Unidas. (1999). Clasificaciones de gastos por finalidades. *Naciones Unidas*, *84*.

Read, T., & Bárcena, E. (2000). La ingeniería Lingüística con Java. *Revista Iberoamericana de educación a distancia*, *3*.

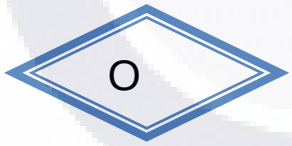
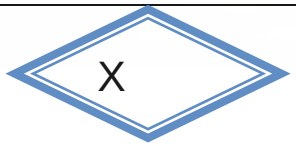
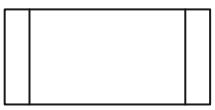
Tordera Illescas, J. C. (2009). Lingüística computacional y anáfora (Tésis Doctoral). *Universidad de Valencia*, *1*.

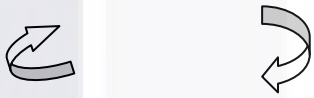
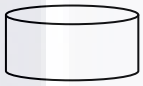
Vallez, M. (2009). *Web Semántica y Sistemas de Información Documental*. Trea, Gijón.

Villayandre Llamazares, M. (2010). APROXIMACIÓN A LA LINGÜÍSTICA COMPUTACIONAL (Tesis Doctoral). *Universidad de León*.

ANEXOS.

Anexo No. 1 Lenguaje de procesos de negocio a nivel detallado.

NOMBRE	DESCRIPCIÓN
	**FUNCIÓN/TAREA**
	**PROCESO**
	**EVENTO**
	**Calle del Rol**
	**OR**
	**XOR**
	**MÓDULO**

	**INICIO DE PROCESOS/ACTIVIDAD O FUNCIÓN**
	**TÉRMINO DE ACTIVIDA O FUNCIÓN**
	**MENSAJE**
	RELACIÓN FUNCIÓN/TAREA – MÓDULO
	FLUJO
	**Ciclo de Función(es)**
	**BASE DE DATOS**
	**MOTOR AUTOMÁTICO DE ACTIVIDAD.**

Anexo No. 2. Registro de tiempos de los cuestionarios Muestra.

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

CODIFICACIÓN

Cuestionario Gastos del Hogar 311210015 # de involucrados___ 23

(hr/hom) 23 min	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Análisis de Búsqueda	08	06	04	01	01	03	02	03	04	02	02	01	12	01	01	01	02	01	04	01	01	01	01	01	01
Búsqueda en el catálogo de bienes y servicios	08	08	04	01	01	03	53	04	04	05	06	03	03	02	04	02	08	05	11	02	09	05	03	06	02
Captura del código	06	02	02	02	02	03	02	03	11	02	04	02	03	03	06	02	03	01	04	02	02	02	01	02	03
Corrección de datos.																									

Cuestionario Gastos del Hogar 311210003 # de involucrados___

(hr/hom) 25 min	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Análisis de Búsqueda	04	01	02	02	02	01	02	01	02	01	03	02	01	05	01	01	04	01	02	02	01	05	01	07	01
Búsqueda en el catálogo de bienes y servicios	05	14	14	04	51	05	01	01	08	02	02	03	04	02	03	08	04	03	03	07	05	06	06	07	02
Captura del código	06	06	01	01	01	02	02	01	01	01	01	01	01	01	02	04	03	01	04	03	03	03	03	02	03
Corrección de datos.																									

Nota: Casilla por producto

02

Responsable de Proyecto

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

CODIFICACIÓN

Cuestionario Gasto del Hogar 3112180128 # de involucrados 2

(hr/hom) 21:37 min	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Análisis de Búsqueda	1'	4'	3'	6'	2'	8'	1'	3'	6'	2'	4'	6'	2'	2'	1'	5'	1'	1'	2'	11'	1'	3'	5'	2	4'
Búsqueda en el catálogo de bienes y servicios	21'	20'	1'	8'	8'	14'	2'	12'	14'	10'	12	1'	7'	1'	9'	4'	1'	3'	16'	4'	7'	1'11"	10'	1'	9'
Captura del código	6'	5'	1'	5'	4'	9'	2'	4'	5'	4'	1'	2'	3'	1'	2'	2	1'	1'	6'	2'	4'	4'	8'	1'	3'
Corrección de datos.																									

Cuestionario Libreta de Gastos Individuales 31121000 23 # de involucrados 2

(hr/hom) 5:07 min	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Análisis de Búsqueda	2'	2'	2'	5'	1'	5'	12'	1'	4'	6'	5'	3'	2	3'	2'	4'	6'	3'	1'						
Búsqueda en el catálogo de bienes y servicios	3'	1'	10'	2'	8'	1h 34'	10'	17'	5'	2'	1'	1'	4'	2'	1'	2'	2'	2'	1'						
Captura del código	2'	1'	3'	1'	2'	12'	2'	3'	4'	1'	1'	2'	2'	1'	1'	1'	1'	1'	1'						
Corrección de datos.																									

Nota: Casilla por producto

Entrevistador 11

Responsable Proyecto

Anexo No. 3 Cuestionarios de satisfacción del “SEBS”.

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Cuestionario de Satisfacción del SEBS (Sistema de Entendimiento de Bienes y Servicios)

1. ¿Anteriormente ya había trabajado con un sistema de codificación?
   SI NO X

2. ¿El sistema SEBS redujo el tiempo de trabajo en tus labores de codificación?
   SI X NO

3. ¿El sistema SEBS te proporciona los códigos con mayor facilidad que el catálogo impreso?
   SI X NO

4. ¿El sistema redujo el margen de error de tu codificación?
   SI X NO

5. Si tuvieras que escoger entre el catálogo impreso y el sistema SEBS ¿Cuál escogerías?
   Catálogo Impreso SEBS X

6. En general, ¿Cómo te parece el sistema SEBS para tus actividades de codificación?
   Excelente Bueno X Regular Malo Pésimo

7. En conclusión, ¿utilizarías el SEBS como herramienta para codificar?
   SI X NO

02

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Cuestionario de Satisfacción del SEBS (Sistema de Entendimiento de Bienes y Servicios)

1. ¿Anteriormente ya había trabajado con un sistema de codificación?
   SI NO

2. ¿El sistema SEBS redujo el tiempo de trabajo en tus labores de codificación?
   SI NO

3. ¿El sistema SEBS te proporciona los códigos con mayor facilidad que el catálogo impreso?
   SI NO

4. ¿El sistema redujo el margen de error de tu codificación?
   SI NO

5. Si tuvieras que escoger entre el catálogo impreso y el sistema SEBS ¿Cuál escogerías?
   Catálogo Impreso SEBS

6. En general, ¿Cómo te parece el sistema SEBS para tus actividades de codificación?
   Excelente Bueno Regular Malo Pésimo

7. En conclusión, ¿utilizarías el SEBS como herramienta para codificar?
   SI NO

E = 11

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Cuestionario de Satisfacción del SEBS (Sistema de Entendimiento de Bienes y Servicios)

1. ¿Anteriormente ya había trabajado con un sistema de codificación?
   SI NO

2. ¿El sistema SEBS redujo el tiempo de trabajo en tus labores de codificación?
   SI NO

3. ¿El sistema SEBS te proporciona los códigos con mayor facilidad que el catálogo impreso?
   SI NO

4. ¿El sistema redujo el margen de error de tu codificación?
   SI NO

5. Si tuvieras que escoger entre el catálogo impreso y el sistema SEBS ¿Cuál escogerías?
   Catálogo Impreso SEBS

6. En general, ¿Cómo te parece el sistema SEBS para tus actividades de codificación?
   Excelente Bueno Regular Malo Pésimo

7. En conclusión, ¿utilizarías el SEBS como herramienta para codificar?
   SI NO

E10

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Cuestionario de Satisfacción del SEBS (Sistema de Entendimiento de Bienes y Servicios)

1. ¿Anteriormente ya había trabajado con un sistema de codificación?
   SI (NO)

2. ¿El sistema SEBS redujo el tiempo de trabajo en tus labores de codificación?
   (SI) NO

3. ¿El sistema SEBS te proporciona los códigos con mayor facilidad que el catálogo impreso?
   SI (NO)

4. ¿El sistema redujo el margen de error de tu codificación?
   SI (NO)

5. Si tuvieras que escoger entre el catálogo impreso y el sistema SEBS ¿Cuál escogerías?
   Catálogo Impreso (SEBS)

6. En general, ¿Cómo te parece el sistema SEBS para tus actividades de codificación?
   Excelente (Bueno) Regular Malo Pésimo

7. En conclusión, ¿utilizarías el SEBS como herramienta para codificar?
   (SI) NO

#5