

Universidad de las Ciencias Informáticas
Facultad #3



Título:

**Propuesta de un Sistema de Información
para el Control Interno.**

**Trabajo de Diploma para optar por el título de
Ingeniero en Ciencias Informáticas.**

Autores:

Yoandy Lichilín Ríos
Yunior Luis Rodríguez Betancourt

Tutores:

Lic. Águeda García Martín
Ing. Eikel Inda Ramos

Ciudad de La Habana
Junio 2008

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Yoandy Lichilín Ríos.

Autor

Yunior Luis Rodríguez Betancourt

Autor

Lic. Águeda García Martín.

Tutora

Ing. Eikel Inda Ramos.

Tutor

DATOS DE CONTACTO

Lic. Águeda L. García Martín: Licenciada en Física, Universidad de La Habana, 1975. Especialista en Óptica y Láser, Instituto Superior Politécnico “José Antonio Echeverría”, 1980. Profesora Auxiliar del Instituto Superior de Tecnologías y Ciencias Aplicadas (InSTEC) del Ministerio de Ciencias, Tecnología y Medio Ambiente – CITMA. Profesora a Tiempo Parcial y Fundadora de la Universidad de las Ciencias Informáticas. arcoiris@infomed.sld.cu; agueda@uci.cu

Ing. Eikel Inda Ramos: Ingeniero en Ciencias Informáticas, Universidad de las Ciencias Informáticas, 2007. Profesor de la Universidad de las Ciencias Informáticas. einda@uci.cu

AGRADECIMIENTOS

En primer lugar, a la Revolución y al compañero Fidel Castro Ruz por hacer realidad esta universidad de excelencia y guiarnos en cada momento hacia un mundo mejor.

A la Universidad de las Ciencias Informáticas (UCI) por formarnos como mejores personas durante cinco largos años.

A nuestros queridos tutores que dedicaron su valioso tiempo para prepararnos y guiarnos en el camino de este conocimiento, sin los cuales no hubiera sido posible la realización de este trabajo.

A nuestros padres, tíos y abuelos por brindarnos un hogar cálido y enseñarnos que la perseverancia y el esfuerzo son el camino para lograr lo que nos proponíamos.

A todas aquellas personas que hicieron su valioso aporte para el desarrollo de esta investigación. En especial a los profesores M.Sc Raúl de la Cruz Cordovés, M.Sc. José Raúl Rodríguez Galera, M.Sc. José Ramón Ruiz de Zarate y Lic. Mirelis Rasua López, por su dedicación y paciencia.

DEDICATORIA

- *A mis padres Angela y Eladio, que siempre han confiado en mí y me han dedicado infinito amor.*
- *Al recuerdo de dos personas que fueron decisivas en mi formación y educación, mi tía Ana Lichilín Álvarez y a mi abuela Eugenia Ampara Piedad Lichilín Benítez.*
- *A todos mis seres queridos, que me han apoyado de una u otra forma a ser quien soy.*
- *A mis padres Zaida, Mary, Carlos y Fernando por brindarme ese amor inmenso, incondicional y sincero, por apoyarme en todo y por confiar en mí desde un principio.*
- *A mi abuelita Pella por encaminar mis primeros pasos hacia el conocimiento y por estar siempre a mi lado con su amor, dedicación y ternura.*
- *A abuelo Esmildo por ser ese hombre único en la tierra que no comete errores, capaz de ofrecer lo poco que tiene y de estar siempre ahí cuando lo necesite.*
- *A mis hermanitos Fernan y Esteban por quererme tanto y hacer esforzarme en la tarea de que me vieran como un ejemplo para seguir mis pasos.*
- *A cada uno de mis familiares y amigos, por estar siempre a mi lado, cuidándome, apoyándome y haciendo de mí un mejor hombre del mañana.*

A todos, gracias por existir.

Yunior L. Rodríguez Betancourt.

RESUMEN

Para llevar a cabo el Control Interno en las entidades se usa el Cuestionario Estándar de Control Interno (CECI), el cual aporta insuficiente información para la toma de decisiones a nivel empresarial o gubernamental.

A partir de esta situación, se ha decidido hacer una propuesta de un Sistema de Información para el Control Interno, que tribute a la confiabilidad de los datos y que permita al evaluador argumentar esa evaluación. Este nuevo cuestionario se ha denominado Cuestionario Ampliado para la Evaluación de Control Interno (CAECI).

Esta propuesta consiste en la adición de elementos adicionales que refrenden a la evaluación dada, utilizando técnicas de la Inteligencia Artificial, que permitan almacenar Datos y Metadatos para su procesamiento, indexación y recuperación de la información mediante búsqueda semántica.

Para llevar a cabo todos estos procesos se utilizan diferentes modelos y estándares, que puedan tributar a la eficacia de la propuesta para la evaluación del Control Interno.

PALABRAS CLAVES

Sistema de Información, Control Interno, Cuestionario, Metadatos, Lógica Borrosa, Indexación, Búsqueda Semántica, Modelos.

TABLA DE CONTENIDOS

AGRADECIMIENTOS	I
DEDICATORIA	II
RESUMEN	II
INTRODUCCIÓN	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA	8
1.1. Introducción	8
1.2. El Control Interno	8
1.2.1. Ventajas del Control Interno	9
1.2.2. Control Interno Informático	10
1.2.3. Componentes del Informe COSO	12
1.3. Evaluación de los Sistemas de Control Interno	13
1.3.1. Estados de madurez del Control Interno	14
1.3.2. Necesidad de un manejo inteligente de datos	15
1.3.3. Análisis de la Información, Toma de Decisiones y Plan de Prevención	17
1.4. Procesamiento de Datos y Metadatos	19
1.4.1. Lógica Borrosa	19
1.4.2. Árboles de Decisión	22
1.4.3. Los Metadatos	24
1.4.3.1. Generalidades de los Metadatos	24
1.4.3.2. Estándares para Metadatos	25
1.4.3.2.1. Dublin Core	26
1.4.3.2.2. RDF/XML	27
1.4.4. Recuperación de la Información	29
1.4.4.1. La Indexación Semántica	29
1.4.4.2. La Búsqueda Semántica	31
1.5. Sistemas de Información	32
1.6. Conclusiones	33

CAPÍTULO 2: PROPUESTA DE SOLUCIÓN DEL PROBLEMA.	34
2.1. Introducción.	34
2.2. Cuestionario Ampliado para la Evaluación del Control Interno.	34
2.3. El Sistema de Ayuda para la Toma de Decisiones.	35
2.3.1. Sistema Gestor de Base de Datos (SGBD).	36
2.3.1.1. Estructura básica de la Base de Datos.	37
2.3.1.2. Cálculo de pesos.	38
2.3.1.3. La frecuencia de términos.	38
2.3.1.4. Pesos de las consultas.	39
2.3.1.5. Resolución de consultas.	39
2.3.2. Sistema Gestor de Base de Modelos (SGBM).	40
2.3.2.1. Modelo Árbol de Decisión.	41
2.3.2.2. Modelo del Indicador.	47
2.3.2.3. Modelo del Espacio Vectorial.	47
2.3.2.3.1. Asignación de pesos a los términos.	49
2.3.2.3.2. Normalización de la frecuencia de términos.	50
2.3.2.3.3. Operaciones sobre los términos.	51
2.3.2.3.4. Fichero invertido.	53
2.3.3. Sistema de Generación y Gestión de Diálogo (SGGD).	55
2.3.4. Integración del DSS.	56
2.4. Conclusiones.	58
CAPÍTULO 3: ESTUDIO DE FACTIBILIDAD DE LA PROPUESTA.	59
3.1. Introducción.	59
3.2. Visualización de la Evaluación del cuestionario.	59
3.2.1. Herramienta “Gráfica evaluativa”.	60
3.2.2. Patrón Pentágono.	63
3.3. Aplicación de la Herramienta Tropes	64
3.3.1. Experimentos utilizando Tropes.	65
3.4. Conclusiones.	67
CONCLUSIONES	68
RECOMENDACIONES	69
BIBLIOGRAFÍA	70
ANEXOS	72
GLOSARIO	82

INTRODUCCIÓN

Los Sistemas de Información, principalmente los de soporte a la Toma de Decisiones, se están convirtiendo hoy en día, en piezas claves para muchas organizaciones. Según aumenta la potencia de cálculo y se abaratan los costos del hardware necesario para este tipo de herramientas, se incrementan día a día las empresas que se deciden a implantar sistemas de este tipo en diferentes niveles de la cadena de la Toma de Decisiones. Pero para adoptar un Sistema de Información es necesario comprender hasta sus últimas consecuencias la importancia de la información y la tecnología que las soportan y que son ellas, además de los Recursos Humanos, sus más valiosos activos. [Grandío, 1996; Blanco, 2004]

Las empresas exitosas reconocen los beneficios de la tecnología de información y la utilizan para impulsar el valor de sus interesados (*stakeholders*). Estas empresas también entienden y administran los riesgos asociados, así como la dependencia crítica de muchos procesos de negocio en Tecnologías de la Información (TI), la administración de los riesgos asociados a dichas TI, así como el incremento de requerimientos para controlar la información, se entienden ahora como elementos clave del gobierno de la empresa. El *valor*, el *riesgo* y el *control* constituyen la esencia del Gobierno de TI, lo cual recuerda que, teniendo en cuenta los cuatro pilares básicos de la actividad empresarial: planificación, organización, ejecución y control, requieran atender de manera especial este último proceso. [Blanco, 2002]

Entre los mecanismos reconocidos del *control*, se encuentra la llamada Auditoría, que solo ejerce una función de intervención, es externa y que evalúa y juzga. La importancia de institucionalizar el control desde adentro de la entidad, hizo necesario reconocer acciones internas de tipo correctivo y preventivo, por lo que apareció la necesidad de separarlas: la Auditoría Externa y la Auditoría Interna, aunque ambas responden a una necesidad de la Administración al nivel correspondiente. Desde el tristemente famoso Escándalo Watergate ocurrido en los Estados Unidos en la década del 70 del pasado siglo, el interés creciente por el tema del *control* permitió la adopción de diversos marcos regulatorios complementarios para nuevos formatos del *control*, entre ellas el Control Interno. Éste constituye una herramienta única y preventiva en su más amplia concepción, uno de los más importantes procesos actuales que se desarrollan no solo a nivel empresarial sino a nivel gubernamental. [Toledo, 2003; Durand, 2004]

El marco conceptual del Control Interno se ha incorporado al desarrollo de las Buenas Prácticas de la Gobernabilidad, se destacan diversos eventos de importancia en el escenario empresarial y una serie de instituciones reconocidas y estándares adoptados convergen en identificar la actualidad y relevancia de contar con un Sistema de Control Interno (SCI):

- Instituciones como la entidad no lucrativa **Committee of Sponsoring Organizations (COSO)** que concibió el Informe COSO, desde 1992, como se le conoce al cuestionario que se utiliza actualmente y que involucró las siguientes organizaciones: *American Accounting Association (AAA)*, *American Institute of Certified Public Accountants (AICPA)*, *Financial Executive Institute (FEI)*, *Institute of Internal Auditors (IIA)*, *Institute of Management Accountants (IMA)* [Treadway, 1987].
- La creación en 1998 del **IT Governance Institute - ITGI**- establecido para evolucionar el pensamiento y los estándares internacionales respecto a la dirección y control de la Tecnología de Información (TI) de una empresa, y el diseño riguroso de la bien conocida **Information Systems Audit and Control Association (ISACA)** ha emitido desde el año 2005 un conjunto de estándares, directrices y procedimientos, en plena concordancia con los estándares de COSO y que enfoca ese proceso como sistema ó Sistema de Información (SI). [Peza, 2007; Ibáñez, 2008].
- La necesidad de extender la concepción del Gobierno Corporativo a tenor de las TI que insiste en el apropiado control de los procesos con el objetivo de un monitoreo efectivo del desempeño que mantenga el registro y evite resultados inesperados, surge entonces COBIT acrónimo de **Control Objectives for Information and related Technology** como marco de referencia para la adecuada administración de los recursos de las TI que han evolucionado siendo también coherentes con el estándar COSO. La propia evolución se muestra en sus diferentes versiones: COBIT1.0 (1996) centrado en la Auditoría, COBIT2.0 (1998) centrado en el Control, COBIT3.0 (2000) centrado en la Administración, COBIT4.0 (2005) centrado en el Gobierno. [CobiT 4.0, 2005].

A raíz de la Convención de Usuarios de COBIT en Cancún, México (desarrollada entre el 14-16 de mayo del 2008) se anuncia la futura versión COBIT4.1, en el cual el procedimiento **ME2 Monitorear y evaluar el Control Interno**, identifica los diferentes estadios de madurez del Control

Interno, lo cual hace re-pensar en la realidad de la empresa cubana actual a la luz de estas tendencias.

En estos momentos avanza un nuevo formato de gobierno empresarial: el denominado Gobierno de la Gestión orientada a Riesgos (*Enterprise Risk Management- ERM*) en cuyos formatos sigue vigente la importancia del Control Interno y de su cuestionario asociado ampliándose las componentes a ser evaluadas, en el caso especial la asociada al Riesgo.

En Cuba desde 2003, se desarrollan ingentes esfuerzos a nivel gubernamental, en la implantación del Control Interno. Refrendado por la *Resolución No. 297/2003* del Ministerio de Finanzas y Precios y ratificada su importancia a través de la adopción, siguiendo los estándares internacionales, en un nuevo formato de Auditoría en el 2006 por respaldo legal de la *Resolución No. 26/2006* del Ministerio de Auditoría y Control, se estableció en una de sus tres partes, la Auditoría al Sistema del Control Interno.

Pero si bien desde el punto de vista de las instancias gubernamentales se reconoce la importancia del Control Interno, aún en la base no se reconoce de manera convincente, no está clara su utilidad. Siguen siendo los resultados de las Auditorías Externas la que en definitiva determinan las acciones ulteriores de los directivos y/o especialistas y no se aprovecha la rica experiencia cotidiana que está en la propia empresa, en sus acciones y en sus resultados, en el conocimiento que su Capital Humano debe incorporar a las acciones concretas por el mejoramiento de la entidad, tal y como prevee la concepción moderna y actualizada del Control Interno. Entonces ¿por qué no se ha avanzado más en el Control Interno en la empresa cubana? [García et. al, 2006].

Es que no basta con contar con el respaldo legal, no basta con acciones de Capacitación si éstas se limitan a la explicación de las normas, de las guías y de la documentación que regula el mismo. Se requiere de una profundización de su sentido y de los diferentes estándares que en su evolución siguen apuntando a el perfeccionamiento de los métodos para responder al cuestionario asociado al Informe COSO, el cual será referido como el Cuestionario Estándar del Control Interno (CECI). El cuestionario consta de 116 preguntas asociadas en 27 temáticas y a su vez en las conocidas Cinco Componentes del Control Interno. En su interacción con directivos y especialistas de diversas instituciones, de manera preliminar se han identificado un conjunto de insuficiencias en

relación a la recogida de datos, su procesamiento y la posibilidad de extraer de ellos información útil y significativa. [García et. al., 2007].

Para lograr incorporar el Control Interno en el eje de las Buenas Prácticas de la Gobernabilidad se requiere trascender lo legal, llegar a la esencia del proceso, comprender su especificidad y reflexionar y actuar en consecuencia con creatividad. Se requiere por tanto, profundizar en lo axiológico, emprendiendo acciones en torno al Manejo Inteligente de Datos y en un adecuado procesamiento de los mismos que le permita al directivo y/o a los especialistas de la empresa visualizar adecuadamente información significativa que le permita poder, de una manera sencilla y en tiempo real, valorar el estado de la entidad y su tendencia. Todo lo anterior difícilmente pudiera realizarse sin la incorporación de las Tecnologías de la Informática y las Comunicaciones (TICs).

Con este trabajo se pretende contribuir al desarrollo de un SI para llevar a cabo de la mejor manera el Control Interno en las instituciones. Actualmente se cuenta con el Cuestionario Estándar del Control Interno (CECI) aplicado en Cuba y en América Latina, éste solo genera datos de tipo cualitativo en una escala valorativa en cuatro niveles, con imposibilidad de brindar información sintética que se pueda procesar y utilizar en la Toma de Decisiones de las entidades. Partiendo del CECI, se propone un sistema capaz de permitir al usuario realizar una valoración más fiel y personalizada de sus elementos, basándose en elementos adicionales a dicho cuestionario como comentarios y/o evidencias que respalden cualquier evaluación dada. A este cuestionario modificado se le llamará Cuestionario Ampliado para la Evaluación del Control Interno (CAECI).

Parecería lógico, por tanto, que dentro del mundo empresarial y/o gubernamental se plantee la utilización de estos SI para el estudio de diversos problemas, la búsqueda de soluciones alternativas, lo cual dependerá del adecuado procesamiento de los datos y la extracción de la información requerida. De ahí la importancia de modificar el CECI, generar otros tipos de datos, y procesarlos según modelos que permitan la utilización de técnicas de Inteligencia Artificial. Un proyecto de diseño de un Sistema de Información también estará sujeto a las especificidades de la entidad en que se aplique y de la creatividad de los especialistas por lo que siempre estará sujeto a peticiones de cambio y mejora por parte de sus usuarios.

SITUACIÓN PROBLEMÁTICA

- Inadecuadas vías de almacenamiento de los reportes asociados al Control Interno.
- Insuficiencia de información útil, significativa y oportuna que permita identificar desviaciones.
- Desaprovechamiento de la información implícita en los reportes y en sus posibles comparaciones.
- Insuficientes análisis comparativos del historial de Controles Internos desde el punto de vista empresarial.
- Inexistentes formas de realización de análisis comparativos de los Controles Internos entre entidades de un mismo sector o una misma jurisdicción desde el punto de vista gubernamental.
- Insuficientes formas de visualización adecuadas.

El estudio de factibilidad y la manera de incorporar la Minería de Datos, contribuirá a una valoración fiel y personalizada del estado del Control Interno que antes no se realizaba.

PROBLEMA

El Cuestionario Estándar del Control Interno que gestiona datos calificadores de forma cualitativa no garantiza la seguridad razonable que promueve el propio Control Interno.

OBJETO DE ESTUDIO

Los Sistemas de Información en los mecanismos asociados al Control Interno.

CAMPO DE ACCIÓN

El proceso de creación, análisis y gestión de los Metadatos asociados a la propuesta del Cuestionario Ampliado del Control Interno (CAECI).

OBJETIVO GENERAL

Desarrollar una propuesta para el procesamiento de información asociada al Control Interno y que pueda ser utilizada tanto por la empresa como por las entidades gubernamentales.

OBJETIVOS ESPECÍFICOS

- Realizar un estudio detallado sobre el Control Interno y su integración con los Sistemas de Información como apoyo a las Decisiones.
- Estudiar las técnicas de procesamiento de información (Minería de Datos, Minería de Texto, Reconocimiento de Patrones, etc.) asociadas a cuestionarios abiertos.
- Realizar un estudio sobre la factibilidad de la propuesta obtenida.

Como posibles resultados se pretende obtener toda la información necesaria para la implementación del CAECI, donde las técnicas de obtención de la información sobrepasen las expectativas del Cuestionario Estándar.

HIPÓTESIS

Si se elabora y diseña el CAECI será factible la Minería de Datos y/o el Reconocimiento de Patrones elevando la eficacia del Control Interno y las soluciones esperadas.

TAREAS

Para el logro del objetivo propuesto se determinaron las siguientes tareas de investigación:

- Realización de un estudio detallado sobre el Control Interno, sus estándares y su evolución tanto en empresas como en instituciones gubernamentales, así como las técnicas de procesamiento de cuestionarios, la Búsqueda Semántica y sus aplicaciones en el campo de la Informática, la utilización de estándares en la estructuración y recuperación de Metadatos.
- Análisis del actual Cuestionario Estándar del Control Interno (CECI) utilizado en Cuba y en América Latina, y de las posibilidades de ampliar el mismo en lo que se ha llamado Cuestionario Ampliado para la Evaluación del Control Interno (CAECI).
- Valoración de resultados del CECI y CAECI en algunas entidades cubanas, a través de la utilización de diversas herramientas, representaciones, cálculos y análisis correspondientes con el objetivo de analizar la factibilidad de la propuesta.
- Validación del estudio de factibilidad de la propuesta del Sistema de Información a partir del análisis comparativo de los resultados obtenidos en el estudio.

Este trabajo consta de tres capítulos que incluyen todo lo relacionado con temas actuales de investigación y una propuesta para la realización de un cuestionario ampliado que contribuye a la solución del problema planteado.

Capítulo 1

En este capítulo se hace un estudio del estado del arte tanto de los aspectos de las Ciencias Empresariales involucradas, como de algunas técnicas asociadas al Manejo Inteligente de Datos. Se analizan teóricamente los antecedentes y la evolución del Control Interno. También trata sobre el tratamiento de datos asociados a cuestionarios y de cómo la utilización de Metadatos puede aportar a la extracción de información útil en este campo del Control Interno. Se profundiza en la Indexación de datos, y en los requerimientos para el tema de interés con el objetivo de lograr la adecuada recuperación de la información significativa y la visualización dirigida a los usuarios.

Capítulo 2

En este capítulo se muestra la aplicación de los conocimientos adquiridos durante la revisión bibliográfica y el estudio teórico de los diversos temas tratados. Se fundamenta la propuesta que cumple con el objetivo trazado y se hace un análisis teórico de algunos modelos básicos que puedan ser útiles para dar solución al problema.

Capítulo 3

Después de realizar el modelado de la propuesta, en este capítulo se muestran solamente algunos resultados de los variados experimentos que fueron realizados. Se utilizan diferentes herramientas para el procesamiento de los datos que ayudan a la visualización de los resultados, se contrastan estos resultados y se fundamenta la viabilidad del diseño propuesto.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.

1.1. Introducción.

Se realiza un estudio del estado del arte en las disciplinas de Inteligencia Artificial y Ciencias Empresariales vinculadas al tema tratado, donde se efectúa la sistematización del estudio de la teoría que sustenta la fundamentación del problema. Se profundiza en temas que constituyen consulta obligatoria como: Control Interno, Técnicas para el Procesamiento de Datos asociados a los cuestionarios (Minería de Datos, Minería de Textos, entre otros), así como la estructuración de Metadatos, estándares a utilizar, así como su imbricación en los Sistemas de Información.

1.2. El Control Interno.

En primer lugar, el Control Interno es un concepto internacional. No es una especificidad cubana ni mucho menos. Debe entenderse como el conjunto de actividades de planeación y ejecución realizado por la empresa para lograr que los objetivos se cumplan. Se trata del sistema conformado por las actividades y procesos orientados a satisfacer los objetivos indicados. Su concepto va mucho más allá de la percepción inicial que se tiene con frecuencia sobre este instrumento. [Argadoña, 2002].

La palabra *control* se refiere a los procesos de análisis, planeación, ejecución, monitoreo y mejora continua, mientras que la palabra *interno* hace alusión a que es una responsabilidad que deben asumir las empresas no solo por requerimiento legal sino por conveniencia gerencial y que complementa al control de verificación o externo que realizan entidades ajenas a la propia empresa. [Toledo, 2003].

Actualmente, la modernización de la empresa a la luz de los nuevos enfoques de las Ciencias Económicas considera en enfoque más abarcador y diferenciador. Si la visión anterior del control estaba limitado a: controles contables, controles tácticos, supervisión, revisión, comprobación, verificación, ahora se convoca a establecer:

- Un Sistema de Control Interno acorde a las características de cada entidad
- Potenciar la función de Control como parte del proceso de Dirección.

- Filosofía de acción estratégica compatible con el perfeccionamiento empresarial como modelo de gestión.

1.2.1. Ventajas del Control Interno.

Según las tendencias mundiales se llega a considerar que el Control Interno debe ser un instrumento eficaz para lograr la eficiencia y eficacia en el trabajo de las entidades. El desarrollo de nuevos términos relacionados con el Control Interno, a diferencia de lo que hasta ahora se identificaba, es decir sólo contable, y la introducción de otros mecanismos de gestión en las entidades (Perfeccionamiento Empresarial, Sistemas de Gestión de Calidad) aportan elementos generalizadores para la elaboración de los Sistemas de Control Interno en cada entidad. [Escorza, 2000].

Puede interpretarse que el Control Interno es un proceso integrado a las operaciones efectuadas por la dirección y el resto del personal de una entidad para proporcionar una *seguridad razonable* para el logro de los objetivos siguientes:

- Confiabilidad de la información.
- Eficiencia y eficacia de las operaciones.
- Cumplimiento de las leyes, reglamentos y políticas establecidas.
- Control de los recursos - de todo tipo - a disposición de la entidad

Lo que persigue el Control Interno es amplio y profundo, incluye toda la actividad de la entidad y debe indicar el camino de su perfeccionamiento.

En el propio texto del título de dicha Resolución 297/2003 se identifican varias palabras o frases claves: *realización sistemática de los controles y de su aplicación, toma de decisiones, tareas priorizadas de cada dirigente y del colectivo de trabajadores, fiabilidad de la información, garantizar la preservación de los recursos materiales, financieros y humanos*. Estas palabras claves invitan a nuevos cuestionamientos.

Lo expuesto, sustenta el origen del Control Interno que se entiende como una permanente actitud positiva de autocontrol, autogestión, autoevaluación y automejoramiento, que le permite a la entidad mantener su vista puesta en sus propios objetivos y lograr la calidad, con eficiencia y eficacia. [Peza 2007]

A casi dos décadas de ser utilizado el Control Interno, en su evolución e integración con otros estándares, y en un balance actual de sus resultados, es ineludible la realidad de que este proceso interno de autocontrol puede aportar al perfeccionamiento de la empresa en dependencia de cómo se enfrente. Realmente el trabajador que realice el Control Interno en cierta entidad, deberá conocer la esfera que controla, ser ético e imparcial, y finalmente lo que se exigirá es el llenado del Informe COSO que se traduce en un cuestionario. Esto es un aspecto a atender, por cuanto se está reduciendo a un marcaje un conjunto de actividades y/o procesos de diversa índole. [Lloyd, 2002]

También es de destacar la necesidad de contar con un *historial* de los diferentes momentos en los que se realizaron Controles Internos en la empresa, lo cual permitirá conocer si se avanza o no.

1.2.2. Control Interno Informático.

En la actualidad, es difícil encontrar una empresa que no utilice las TICs en su gestión. Es por ello que existe el Control Interno Informático. Éste debe controlar sistemáticamente que todas las actividades de sus sistemas de información sean realizadas cumpliendo los procedimientos, estándares y normas fijadas por la dirección de la organización y/o la dirección informática, así como los requerimientos legales.

La función del Control Interno Informático, es asegurarse de que las medidas que se obtienen de los mecanismos implantados por cada responsable sean correctas y válidas. Ha de definir los diferentes controles periódicos a realizar en cada una de las funciones informáticas, de acuerdo al nivel de riesgo de cada una de ellas, y ser diseñados conforme a los objetivos de negocio y dentro del marco legal aplicable. Éstos se plasmarán en los oportunos procedimientos de Control Interno y podrán ser preventivos o de detección. Realizará periódicamente la revisión de los controles establecidos de Control Interno Informático informando de las desviaciones a la Dirección de Informática y sugiriendo cuantos cambios crea convenientes en los controles, así como transmitirá constantemente a toda la organización de Informática la cultura y políticas del riesgo informático. [Ibañez, 2008].



Figura 1. Pirámide del Control Interno Informático.

Esta es una herramienta enfocada a la adecuada Gestión de los Sistemas de la Información. En una entidad pueden ocurrir una serie de problemas informáticos, algunos de estos factores pueden ser el indiscriminado uso de la Web dentro de la entidad, el uso de Internet en horario laboral con fines personales, el cual es muy común en estos tiempos en la empresa. Esto trae graves consecuencias ya que los trabajadores no emplean el horario de trabajo correctamente, ocasionando bajo rendimiento y atraso en el trabajo, sin contar los riesgos informáticos que implica la navegación por sitios poco seguros.

Las malas decisiones, los usuarios insatisfechos, los ataques externos, la dependencia de las tecnologías de la información, sus costos y mantenimientos, la integridad de la información, recaen sobre la empresa, y la mejor vía para aliviar dicho peso es implementando un sistema de Control Interno Informático. Como principales objetivos se pueden indicar los siguientes:

- Controlar que todas las actividades se realicen cumpliendo los procedimientos y normas fijados, evaluar su bondad y asegurarse del cumplimiento de las normas legales.
- Asesorar sobre el conocimiento de las normas.
- Colaborar y apoyar el trabajo de Auditoría informática, así como de las auditorías externas al grupo.
- Definir, implantar y ejecutar mecanismos y controles para comprobar el logro de la creación de adecuadas vías para llevar a cabo el servicio informático.

En resumen, el Control Interno Informático es cualquiera y cada una de las tareas o actividades que se llevan a cabo automática o manualmente para evitar, prevenir y corregir errores o malas sorpresas que pueden traer problemas al adecuado funcionamiento de un sistema, empresa o institución para alcanzar sus metas y objetivos propuestos.

Analizada la real importancia de un sistema de Control Interno Informático es necesario crear la conciencia de también implantarlo en cada empresa o institución. Para llegar a conocer la configuración del sistema es necesario documentar los detalles de la red, así como los distintos niveles de control y otros elementos relacionados.

Para la implantación del sistema de Control Interno Informático, se tienen que definir algunos elementos, como son la gestión de sistema de información: políticas, pautas y normas técnicas que sirvan de base para el diseño y la implantación de los sistemas de información y de los controles correspondientes; administración de sistemas: controles sobre la actividad de los centros de datos y otras funciones de apoyo al sistema, incluyendo la administración de las redes, seguridad: incluye las tres clases de controles fundamentales implantados en el software del sistema, integridad del sistema, confidencialidad (control de acceso) y disponibilidad; gestión del cambio: separación de las pruebas y la producción a nivel del software y controles de procedimientos para la migración de programas software aprobados y probados.

1.2.3. Componentes del Informe COSO.

El denominado Informe COSO sobre Control Interno, publicado en EE.UU. en 1992, surgió como una respuesta a las inquietudes que planteaban la diversidad de conceptos, definiciones e interpretaciones relativas de un mejor Control Interno a llevar a cabo en las instituciones. Éste plasma los resultados de la tarea realizada durante más de cinco años por el grupo de trabajo que la *Treadway Commission, National Commission On Fraudulent Financial Reporting* creó en Estados Unidos en 1985 bajo su acrónimo COSO. Este grupo cuenta con alrededor de cincuenta mil miembros y opera en aproximadamente cincuenta países y su primera contribución: el Informe COSO fue encargada a la empresa de auditoría Coopers & Lybrand.

Existen infinidad de manuales y guías sobre el Control Interno, e independientemente de las orientaciones, sugerencias, argumentaciones, todas terminan con el mismo cuestionario. [Comptroller's Handbook, 2001]. Numerosos organismos reguladores de diferentes países hayan

modificado sus normas de auditoría relacionadas con el Control Interno para adaptarlas al marco conceptual desarrollado por el mismo, lo cual evidencia la fortaleza de este estándar [(Ibáñez 2008). Los Cinco Componentes del Control Interno son: Ambiente de Control, Evaluación de Riesgos, Control de Actividades, Información y Comunicación, Supervisión y Monitoreo, que se estructuran en el conocido Cubo COSO. (Figura 2)



Figura 2: Cubo COSO.

El Control Interno, no consiste en un proceso secuencial, donde alguno de los componentes afecta sólo al siguiente, sino en un proceso multidireccional repetitivo y permanente, en el cual más de un componente influye en los otros. Las cinco componentes forman un sistema integrado que reacciona dinámicamente a las condiciones cambiantes.

1.3. Evaluación de los Sistemas de Control Interno.

La evaluación de los sistemas de Control Interno se puede hacer de formas muy variadas. Sería razonable contar con que el grado deseado de documentación esté en proporción con el tamaño y las actividades de la organización. Los métodos generalmente adoptados son: Valoración por flujos de trabajo, Valoración por deberes, Valoración por cuestionarios. La adopción de alguno de los métodos anteriores, no imposibilita el uso de otros de ellos para utilizarlos de forma conjunta.

En Europa, se elaboró una idea inicial sobre los cuestionarios para el Control Interno de las entidades y empresas, que posteriormente EEUU la adoptó también. Las preguntas están enmarcadas de forma tal que las respuestas se den por “sí”, “no” o “no aplicable”. Las respuestas afirmativas indican generalmente buen Control Interno en la entidad, mientras que las respuestas

negativas indican debilidades de la misma, pero todo esto, no abarca ampliamente la información real que se desea. En tales casos, sería recomendable agregar la nota explicativa.

Ya en América Latina, se modificó el método de evaluación, queriendo lograrse uno aparentemente más completo. Surge así, el denominado Cuestionario Estándar de Control Interno, donde cada elemento sería evaluado y enmarcado en un solo y único nivel, ya sea de excelente (E), bien (B), regular (R), mal (M) o No Aplicado (N/A). Al final de los cuestionarios podría aparecer un espacio en el cual se puedan agregar los comentarios y las sugerencias con respecto a las calificaciones otorgadas. [Lloyd, 2000]

Pero: ¿Qué significa para la entidad este resultado cualitativo? ¿Basta este indicador para identificar acciones correctivas, decidir alternativas de solución de las deficiencias? ¿Utiliza este resultado la empresa para poder analizar evolutivamente la mejoría en su gestión? ¿El número o calificación total ó los cambios en la ponderación de los diferentes aspectos? ¿Cómo identifica la entidad la marcha del proceso de perfeccionamiento?

Existen numerosos tipos de escalas valorativas, pero una evaluación externa aunque sea justa, o una valoración superficial de cualquier fenómeno o proceso que se desee evaluar como una simple asignación de valores dados a cualidades o características, no es suficiente. Es necesario llegar a la esencia del mismo. Aquí es donde entra a jugar el papel la posible aplicación de la Lógica Borrosa o “Fuzzy” para determinar a qué conjunto pertenece realmente el elemento del subconjunto analizado. [Espín et. al., 2007]

1.3.1. Estados de madurez del Control Interno.

No basta con una completa y sofisticada informatización del Control Interno en la entidad, ni una amplia y efectiva capacitación de la gerencia de la misma, las dificultades sobrepasan las fronteras del medio tecnológico y de la utilización inmediata de los datos. Es algo más complejo que está en la propia esencia del denominado Control Interno y en la valoración de su eficacia.

Recientemente, ha sido una tendencia la clasificación de los diferentes estadios de madurez del Control Interno porque incluso la Auditoría lo intenta valorar. Hasta el momento han sido detectadas dos clasificaciones: la de una reconocida Consultora [PricewaterhouseCoopers, 2006] que se sintetiza en la Figura 3 y la del IT Governance [COBIT 4.0, 2005] mostrada en la Figura 4.

Escalas clasificatorias del Control Interno	Descripción del estado de madurez del Control Interno.
NO REALIZABLE	Entorno no predecible donde las Actividades de Control no se encuentran diseñadas o no están instrumentadas.
INFORMAL	Actividades de Control diseñadas e instrumentadas pero no adecuadamente documentadas y no se comunican.
ESTANDARIZADO	Actividades de Control diseñadas e instrumentadas y comunicadas a los empleados pero no se detectan desviaciones.
MONITOREADO	Controles estandarizados y periódicos enfocados a la Gerencia aunque la automatización y herramientas se utiliza de manera limitada para apoyar las Actividades de Control.
OPTIMIZADO	Marco de Control Integrado con monitoreo en tiempo real y mejoramiento continuo con automatización y herramientas que apoyan a las Actividades de Control para efectuar cambios rápidos en las mismas.

Figura 3. Estados de madurez del Control Interno según PricewaterhouseCoopers.

1.3.2. Necesidad de un manejo inteligente de datos.

Para poder tener acceso a los datos, es necesario primeramente almacenarlos y organizarlos, para acceder a los mismos y aplicar los modelos que permitan extraer la información asociada. De alguna forma el usuario o cliente debe dar a conocer su interés en una búsqueda específica y esto lo hará mediante una consulta que será procesada por los motores de búsqueda o por el Sistema de Información. Dentro de éste se encuentra el Sistema de Recuperación de la Información (SRI), que devuelve al usuario la información útil y significativa.

Estos sistemas deben permitir ordenar los elementos de la colección por el orden de importancia o relevancia para la consulta de cualquier usuario. La dificultad no está sólo en conocer cómo extraer esta información sino también cómo utilizarla para decidir la relevancia de cada documento. En consecuencia, la noción de relevancia es un concepto clave en la recuperación de información. De hecho, el objetivo primordial de un SRI es recuperar todos los documentos que son relevantes para la consulta. En un sistema de gestión de bases de datos se puede incorporar documentos en formato nativo, así como indexarlos automáticamente a fin de que cualquiera de las palabras presentes en los mismos puedan ser utilizadas como puntos de acceso. [Pinto, 2007]

Si el tipo de dato almacenado fueran textos, es entonces mayor la complejidad de los modelos a utilizar. Aparece entonces la Minería de Textos, tarea de gran complejidad.

MODELO DE MADUREZ
ME2 Monitorear y evaluar el Control Interno

La administración del proceso de *Monitorear y evaluar el control interno* que satisfaga el requisito de negocio de TI de *proteger el logro de los objetivos de TI y cumplir con las leyes y regulaciones relacionadas con TI* es:

0 No existente cuando

La organización carece de procedimientos para monitorear la efectividad de los controles internos. Los métodos de reporte de control interno gerenciales no existen. Existe una falta generalizada de conciencia sobre la seguridad operativa y el aseguramiento del control interno de TI. La gerencia y los empleados no tienen conciencia general sobre el control interno.

1 Inicial/Ad Hoc cuando

La gerencia reconoce la necesidad de administrar y asegurar el control de TI de forma regular. La experiencia individual para evaluar la suficiencia del control interno se aplica de forma ad hoc. La gerencia de TI no ha asignado de manera formal las responsabilidades para monitorear la efectividad de los controles internos. Las evaluaciones de control interno de TI se realizan como parte de las auditorías financieras tradicionales, con metodologías y habilidades que no reflejan las necesidades de la función de los servicios de información.

2 Repetible pero intuitiva cuando

La organización utiliza reportes de control informales para comenzar iniciativas de acción correctiva. La evaluación del control interno depende de las habilidades de individuos clave. La organización tiene una mayor conciencia sobre el monitoreo de los controles internos. La gerencia de servicios de información realiza monitoreo periódico sobre la efectividad de lo que considera controles internos críticos. Se están empezando a usar metodologías y herramientas para monitorear los controles internos, aunque no se basan en un plan. Los factores de riesgo específicos del ambiente de TI se identifican con base en las habilidades de individuos.

3 Proceso definido cuando

La gerencia apoya y ha institucionalizado el monitoreo del control interno. Se han desarrollado políticas y procedimientos para evaluar y reportar las actividades de monitoreo del control interno. Se ha definido un programa de educación y entrenamiento para el monitoreo del control interno. Se ha definido también un proceso para auto-evaluaciones y revisiones de aseguramiento del control interno, con roles definidos para los responsables de la administración del negocio y de TI. Se usan herramientas, aunque no necesariamente están integradas en todos los procesos. Las políticas de evaluación de riesgos de los procesos de TI se utilizan dentro de los marcos de trabajo desarrollados de manera específica para la función de TI. Se han definido políticas para el manejo y mitigación de riesgos específicos de procesos.

4 Administrado y medible cuando

La gerencia tiene implantado un marco de trabajo para el monitoreo del control interno de TI. La organización ha establecido niveles de tolerancia para el proceso de monitoreo del control interno. Se han implantado herramientas para estandarizar evaluaciones y para detectar de forma automática las excepciones de control. Se ha establecido una función formal para el control interno de TI, con profesionales especializados y certificados que utilizan un marco de trabajo de control formal avalado por la alta dirección. Un equipo calificado de TI participa de forma rutinaria en las evaluaciones de control interno. Se ha establecido una base de datos de métricas para información histórica sobre el monitoreo del control interno. Se realizan revisiones entre pares para verificar el monitoreo del control interno.

5 Optimizado cuando

La gerencia ha implantado un programa de mejora continua en toda la organización que toma en cuenta las lecciones aprendidas y las mejores prácticas de la industria para monitorear el control interno. La organización utiliza herramientas integradas y actualizadas, donde es apropiado, que permiten una evaluación efectiva de los controles críticos de TI y una detección rápida de incidentes de control de TI. La compartición del conocimiento, específico de la función de servicios de información, se encuentra implantada de manera formal. El benchmarking con los estándares de la industria y las mejores prácticas está formalizado.

Figura 4: Estados de Madurez del Control Interno según CoBIT 4.1

El manejo inteligente de datos consiste en el procesamiento automático principalmente de documentos, donde los datos van a estar separados por aspectos de clasificación y distribución del flujo de documentos, búsqueda de documentos relevantes y parecidos. Todos los documentos que describen casos parecidos pueden agruparse y de esta forma la búsqueda sería óptima y completa.

Un ejemplo de herramienta eficiente y muy utilizada por la comunidad para la recuperación de la información es Lemur; incluye tecnologías como recuperación distribuida, con preguntas estructuradas, resúmenes automáticos, filtrado y clasificación; funciona sobre múltiples sistemas operativos y es capaz de indexar grandes colecciones de documentos. [Pinto 2007]

1.3.3. Análisis de la Información, Toma de Decisiones y Plan de Prevención.

De los autores consultados, la mayor parte hablan en forma general sobre el análisis de la información y en ocasiones, no llegan a proponer su concepto. Mientras tanto se afirma que *"El análisis de información puede ser de oportunidad -para decidir el mejor momento; de valor agregado -para potenciar el valor de significados aparentemente inconexos; de objetivos -para identificar un blanco y abatirlo en el mejor momento posible"* Para su realización, se requiere del dominio de las fuentes y las técnicas de búsqueda y de procesamiento, así como para el análisis cuantitativo y cualitativo de la información. *"Puede considerarse como el proceso básico que permite la existencia de la inteligencia empresarial y que posibilita la toma de decisiones"*. [Leiva, 2003]

En este sentido, puede considerarse el Análisis de Información como el límite entre los sistemas de inteligencia y los sistemas de gestión de información y el escalón donde la información adquiere su máximo valor. Es indiscutible, la relación existente entre el Análisis de Información y la Toma de Decisiones. La primera es un proceso de carácter continuo y sistémico de transformación de la información en conocimiento y de este último, en decisiones estratégicas, en otras palabras, buscan extraer conocimiento por medio de métodos y técnicas, tanto cuantitativas como cualitativas para proporcionar información útil y valiosa al primer nivel de dirección de la organización, para la oportuna toma de decisiones y por consiguiente, la obtención de ventajas competitivas. [Faúndez 2003]

La situación económica y social que caracteriza a la sociedad moderna genera profundos cambios en las organizaciones las cuales se preparan para ser más flexibles y establecen estrategias con el objetivo de adaptarse al contexto altamente turbulento en el que desarrollan sus

acciones. En presencia de ambientes tan poco estables y la imposibilidad de actuar a ciegas, los miembros de la organización y, en particular, su alta gerencia, necesitan manipular grandes volúmenes de información para cumplir con sus funciones esenciales. Deben implementarse entonces, prácticas administrativas dirigidas a garantizar el éxito organizacional, entre ellas, la toma de decisiones es vital, soportada en el análisis de información. [Blanco, 2002]

Al tomar una decisión, se debe identificar todas las alternativas disponibles, pronosticar sus consecuencias y evaluarlas según los objetivos y metas trazadas. Para ello, se requiere en primer lugar, tener la información actualizada sobre las alternativas que se encuentran disponibles en el presente o las que se deben considerar. También se necesita información sobre el futuro, conocer las consecuencias de actuar según cada una de las diversas opciones. Por último y no menos importante, es indispensable obtener la información sobre cómo pasar del presente al futuro, saber cuáles son los valores y las preferencias que se deben utilizar para seleccionar, entre las alternativas que se tengan, las que conducen del mejor modo a los resultados que se desean.

Puede definirse entonces, la Toma de Decisiones como una actividad indispensable en las organizaciones, con un significado especial para todos sus niveles, porque es parte fundamental inherente a todas las demás actividades de la empresa. La misma se ha convertido en una función imprescindible para la vida de cualquier organización. Interviene en la racionalidad de quienes toman las decisiones, cuando la incertidumbre impera. La triada: dato-información-conocimiento lleva a hacer de las entidades un fuerte, lejos de sorpresas no deseadas, pues se tiene más conocimiento del contexto en que está enmarcada la misma.

El Plan de Prevención es una herramienta contra la corrupción, permite prever las ilegalidades, las indisciplinas. Éste no es más, que un conjunto de acciones político-ideológicas, técnico-organizativas y de control, para reducir al mínimo las acciones negativas. Constituye un instrumento de trabajo para los cuadros, dirigentes y funcionarios a fin de concentrar y dirigir el trabajo individual con cada trabajador, con el firme propósito de elevar los valores, el sentido de pertenencia, la responsabilidad y el ahorro de los recursos. Los planes de prevención proponen un conjunto de medidas las cuales son muy efectivas. Entre los objetivos de este plan, se pueden mencionar:

- Crear un clima de máxima honradez y eficiencia en los colectivos laborales.
- Lograr que el control y la prevención constituya una cultura de la administración en forma de pensar y actuar.

- Ayudar a prevenir los puntos vulnerables y a determinar acciones, así como, posibilitar la interrelación entre los colectivos laborales y direcciones administrativas.

Puede considerarse que sería una aspiración de un Sistema de Control Interno que, a partir del Procesamiento de Datos y Análisis de la Información, puedan inferirse las principales tareas del Plan de Prevención.

1.4. Procesamiento de Datos y Metadatos.

1.4.1. Lógica Borrosa.

La Lógica Borrosa o Difusa es una teoría que permite manejar y procesar ciertos tipos de información en los cuales se manejen términos inexactos, imprecisos o subjetivos. Esta teoría permite manejar y procesar información de una manera similar a como lo hace el cerebro humano, es posible ordenar un razonamiento basado en reglas imprecisas y en datos incompletos. Estos sistemas son generalmente robustos y tolerantes a imprecisiones y “ruidos” en los datos de entrada. Para esto se debe ampliar la Teoría de Conjuntos y la Lógica Booleana, de manera que un individuo pueda pertenecer parcialmente a un conjunto y que las operaciones lógicas además de unos y ceros, también puedan tomar valores reales en dicho rango. [Winston, 1994]

Con reglas difusas es posible procesar las relaciones entre las variables borrosas y producir una salida difusa. Y lo mejor de todo, a partir de esas salidas nebulosas se pueden proporcionar cantidades binarias y cantidades continuas, como el estado de un interruptor o una cantidad de dinero.

En Inteligencia Artificial, la Lógica Difusa o Lógica Borrosa, se utiliza para la resolución de una variedad de problemas, principalmente los relacionados con control de procesos industriales complejos y sistemas de decisión, en general, la resolución, la compresión de datos. Consiste en la aplicación de la Lógica Difusa con la intención de imitar el razonamiento humano en la programación de computadoras. Con la lógica convencional, las computadoras pueden manipular valores estrictamente duales, como verdadero/falso o sí/no. En la Lógica Difusa, se usan modelos matemáticos para representar nociones subjetivas, como *joven/adulto/anciano*, para valores concretos que puedan ser manipuladas por los ordenadores.

En este paradigma, también tiene un especial valor la variable del tiempo, ya que los sistemas de control pueden necesitar retroalimentarse en un espacio concreto de tiempo, pueden necesitar datos anteriores para hacer una evaluación media de la situación en un periodo de tiempo anterior.

Para mostrar ejemplos de su posible aplicación en situaciones de la vida real surge una sencilla pregunta: ¿cómo se podrá llamar al color que posee un grado de verdor de 0,80 y un grado de enrojecimiento de 0,20? Se podría denominar verde amarillento o amarillo verdoso, pero esto también pudiera ocurrir con el color de un grado de verdor de 0,70 y un grado de enrojecimiento de 0,30. Aquí es donde se puede aplicar la Lógica Borrosa o "Fuzzy" para brindar posibles soluciones, pues ésta no define la cantidad de variación de colores entre el verde y el rojo de una forma finita como lo haría un humano, sino que dentro de los conjuntos difusos del enrojecimiento y el verdor asigna grados de pertenencia. Esto es semejante a cuando se dice: "éste verde es más amarillento que aquél otro".

Este es un ejemplo de la aplicación de la "fusificación". Es muy frecuente encontrarse ante objetos que no tienen muy bien definidos sus límites, como es el caso de los colores. Las imprecisiones en las mediciones no se deben a defectos en los aparatos de medida empleados, sino a la vaguedad intrínseca de los objetos medidos.

La Lógica Borrosa se caracteriza por querer cuantificar o llevar a números las incertidumbres que surgen. Si X es una proposición conocida, se le puede asociar un número $v(X)$ en el intervalo $[0,1]$ tal que:

- Si $v(X) = 0$, X es falso.
- Si $v(X) = 1$, X es verdadero.
- La veracidad de X aumenta con $v(X)$.

A primera vista hace tener la idea de una Teoría de la Probabilidad, aunque persigue fines distintos, más relativos a la Teoría de la Posibilidad. Los operadores lógicos que se utilizarán en Lógica Difusa (AND, OR, etc.) se definen también usando tablas de verdad, pero mediante un "principio de extensión" por el cual gran parte del aparato matemático clásico existente puede ser adaptado a la manipulación de los conjuntos difusos y, por tanto, a la de las variables lingüísticas. Se fundamenta en los denominados conjuntos borrosos y un sistema de inferencia borroso basado en

reglas heurísticas de la forma "SI (antecedente)... entonces (consecuente)..... ", donde los valores lingüísticos de la premisa y el consecuente están definidos por conjuntos borrosos, es así como las reglas heurísticas siempre convierten un conjunto borroso en otro. [Martínez, 2002; Vallez, 2007]

La última característica de los sistemas lógicos es el procedimiento de razonamiento, que permite inferir resultados lógicos a partir de una serie de antecedentes. Generalmente, el razonamiento lógico se basa en silogismos, en los que los antecedentes son por un lado las proposiciones condicionales (las reglas), y las observaciones presentes por otro (serán las premisas de cada regla). [Espín, 2007]

La función característica proporciona una medida del grado de similaridad de un elemento de U con el conjunto difuso. La forma de la función característica utilizada, depende del criterio aplicado en la resolución de cada problema y variará en función de la cultura, geografía, época o punto de vista del usuario. La única condición que debe cumplir una función característica es que tome valores entre 0 y 1, con continuidad. Las funciones características más comúnmente utilizadas por su simplicidad matemática y su manejabilidad son: triangular, trapezoidal, gaussiana, sigmoideal, gamma, pi, campana, etc.

Las operaciones básicas entre conjuntos difusos son las siguientes:

- El conjunto complementario A' de un conjunto difuso A es aquel cuya función característica viene dada por: $\mu_{A'}(x) = 1 - \mu_A(x)$.
- La unión de dos conjuntos difusos A y B es un conjunto difuso $A \cup B$ en U cuya función de pertenencia es: $\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)]$
- La intersección de dos conjuntos difusos A y B es un conjunto difuso $A \cap B$ en U con función característica: $\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)]$.

Como principal ventaja, hay que destacar los excelentes resultados que brinda un sistema de control basado en lógica difusa ya que ofrece salidas de una forma veloz y precisa, disminuyendo así las transiciones de estados fundamentales en el entorno físico que controle, además posee la capacidad de adelantarse en el tiempo a los acontecimientos, estabilizando siempre el medio físico que controla.

Como su nombre lo indica, refiriéndose a lo difuso, es muy complicado escoger la verdadera y correcta función que caracterice a un conjunto específico, ya que se debe mencionar la dificultad de introducir o delimitar el alcance del efecto de los cuantificadores de nuestro lenguaje en dicha función y este es uno de sus principales inconvenientes. El hecho de que cualquier función de pertenencia del sistema estuviese mal especificada, haría fallar, muy probablemente, el sistema completo.

1.4.2. Árboles de Decisión.

En Inteligencia Artificial existen diferentes clasificaciones de árboles, uno de ellos es el conocido como Árbol de Regresión o CART (Classification And Regresion Trees), es en él donde se centra este estudio. [Winston, 1994]

Un árbol es un grafo conexo acíclico dirigido y éstos tienen dos tipos de nodos. Los nodos de decisión que indican que se debe tomar una decisión y los nodos de eventos que indican que se debe realizar un experimento aleatorio. El objetivo y función de los mismos es determinar las acciones óptimas, que maximicen la utilidad esperada, en todos los nodos de decisión. Éste daría de forma cualitativa la verdadera evaluación o característica de algún elemento del que se desee analizar su pertenencia. [Winston, 1994; Vallez, 2007]

Se cuenta con un conjunto de objetos de cualquier tipo: O_1, \dots, O_N . Cada objeto O_i ($i=1,2,\dots,N$) lleva asociado un vector (x_{i1}, \dots, x_{ir}) , siendo cada x_{ij} ($j=1,2,\dots,r$) una característica cuantitativa o cualitativa de dicho elemento. La población $\{O_1, \dots, O_N\}$ está dividida en p grupos, Π_1, \dots, Π_p . con lo que se puede conformar la Tabla 1.

	C1	C2	...	Cj	...	Cr	Grupo
O1	X11	X21	...	X1j	...	X1r	—
O2	X21	X22	...	X2j	...	X2r	—
...
Oi	Xi1	Xi2	...	Xij	...	Xir	—
...
On	Xn1	Xn2	...	Xnj	...	Xnr	—

Tabla 1. Ubicación de los elementos del conjunto de objetos.

Lo que en realidad se desea es identificar de forma secuencial, por orden de relevancia, todas las características C_1, \dots, C_r que logren de la mejor manera ayudar a identificar cada objeto de la

población como miembro de alguno de los grupos Π_j al que realmente pertenece. Sin llegar a utilizar ninguna variable, se asignan los individuos O_i al grupo Π_j más probable o frecuente. Si por ejemplo se tiene una característica para la que Π_1 aparece el 90% de la veces y Π_2 el 10% restante, y tenemos otra característica para la que Π_1 aparece el 60% de las veces y Π_2 el 40 % restante, se elegiría la primera característica como regla discriminante, ya que nos ofrece una mayor fiabilidad que la segunda a la hora de clasificar los individuos y asignarlos a un grupo. Al tenerse una variable cualitativa con r categorías, y frecuencias f_1, \dots, f_p . Se define su *entropía* como:

$$- \sum_{j=1}^p (f_j \cdot \log_2 f_j) .$$

Si se tiene una población S de individuos de los grupos Π_1, \dots, Π_p con frecuencias f_1, \dots, f_p , se define la entropía de la población S como:

$$e(S) = - \sum_{j=1}^p (f_j \cdot \log_2 f_j)$$

La población S se llega a dividir según los valores de una cierta característica C_k en las siguientes subpoblaciones S_1, \dots, S_t . En cada subpoblación S_i ($i=1, 2, \dots, t$) se tendrá individuos de los grupos Π_1, \dots, Π_p con frecuencias $f_1^{(i)}, \dots, f_p^{(i)}$, con lo que se puede llegar a calcular la entropía para cada subpoblación como:

$$S_i \rightarrow e(S) = - \sum_{j=1}^p (f_j^{(i)} \cdot \log_2 f_j^{(i)})$$

La entropía de la población S tras realizar la división queda definida como:

$$e^*(S) = \sum_{i=1}^t \frac{|S_i|}{|S|} \cdot e(S_i) = \sum_{i=1}^t \frac{|S_i|}{|S|} \cdot \sum_{j=1}^p f_j^{(i)} \cdot \log_2 f_j^{(i)}$$

Al elegir una característica C_k por la que se quiera o se vaya a ramificar, se toma aquella característica C_k tal que la entropía tras la división de la población S sea mínima. Esto se conoce como el método ID3 (Iterative Dichotomiser 3).

Los árboles de decisión se pueden generar a partir de distintos tipos de atributos, ya sean atributos discretos como atributos numéricos. Cuando se trabaja con atributos discretos, la partición

del conjunto según el valor de un atributo es simple. Por ejemplo, se agrupan todos los software de gran tamaño, siendo “gran_tam” un atributo y sus posibles valores si y no. En el caso de los atributos numéricos esta división no es tan simple. Por ejemplo, si se desea dividir la colección de software en función de sus tamaños, es muy difícil encontrar varios de ellos con el mismo tamaño o capacidad. Para solucionar este problema, puede recurrirse a la “binarización”, éste consiste en formar dos rangos de valores de acuerdo al valor de un atributo, que pueden tomarse como simbólicos. Por ejemplo, si el tamaño de uno de ellos es de 4000kb, se pueden crear los intervalos siguientes: $[0,4000)$ y $[4000, +\infty)$ y el cálculo de la entropía se realiza como si los dos intervalos fueran los dos valores simbólicos que puede tomar el atributo. [Martínez 2002]

1.4.3. Los Metadatos.

Son datos que describen otros datos, es decir, información relativa a los propios datos que facilitan su catalogación y además proporcionan información semántica asociada. En general, un grupo de metadatos se refiere a un grupo de datos, llamado *recurso*. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

Para varios campos de la informática como la recuperación de información o la Web semántica, los metadatos son un enfoque importante para construir un puente sobre el intervalo semántico. Debido a que los mismos son datos en sí mismos, es posible crear metadatos sobre metadatos. Aunque, a primera vista, parece absurdo, estos últimos pueden ser muy útiles. Por ejemplo, fusionando dos imágenes y sus metadatos distintos, puede ser muy importante deducir cuál es el origen de un grupo de ellos. [Mesarina et.al., 2007; Michard, 2005]

1.4.3.1. Generalidades de los Metadatos.

Los metadatos facilitan la descripción de recursos no textuales e información binaria, como: audio, software, imágenes, videos, etc., lo cual, supone una ventaja con respecto a la indización automática dirigida sólo a texto. Deben proporcionar instrumentos para describir el contenido semántico de un recurso y están mejor preparados para soportar la recuperación de información que

el propio documento. En muchos casos, los recursos de información no son capaces de facilitar por sí mismos sus propias relaciones semánticas.

Según la función que proporcionan, se pueden clasificar en descriptivos, estructurales y administrativos. (Figura 5)

Tipo	Objetivo	Ejemplos
Descriptivos	Describen e identifican recursos de información. Permite a los usuarios la búsqueda y recuperación de la información.	Dublin Core; Etiquetas META de HTML
Estructurales	Facilitan la navegación y la presentación de los recursos. Proporcionan información sobre la estructura interna de los documentos, así como la relación entre ellos.	XML y RDF o SGML
Administrativos	Facilitan la gestión de conjuntos de recursos. Incluye la gestión de derechos y sobre control de acceso y uso.	MOA2

Figura 5. Clasificación de los Metadatos.

Una categoría especial de metadatos que hace que los sistemas que los utilizan presenten una ventaja comparativa en la recuperación frente a los que buscan sobre el texto completo, son los metadatos sobre recursos que no son fácilmente accesibles, porque están ocultos (protegidos por contraseñas) o que son accesibles sólo por medio de protocolos específicos (Internet invisible). [Mesarina et.al., 2007]

1.4.3.2. Estándares para Metadatos.

En 1998 Sanders elaboró una definición general sobre los estándares refiriéndose a los mismos como principios para orientar o como guías para evaluar los informes de evaluación; no los definió como reglas fijas que se pueden o deben aplicar mecánicamente. Los estándares, por lo general, identifican prácticas sobre las cuales existen acuerdos o conocimientos generalizados de aceptabilidad, proponen pautas que reflejan la mejor práctica vigente o tendencia y contienen precauciones y recomendaciones o alertas contra errores potenciales.

Un estándar es tanto una meta que se refiere a lo que debiera hacerse o lograrse como una medida de progreso hacia esa meta. Pueden ser de obligatorio cumplimiento exigidos por la ley, pero

también pueden ser voluntarios establecidos por consenso o aceptados por costumbre o por acuerdo de la dirección de la entidad.

1.4.3.2.1. Dublin Core.

Creado en 1995 por iniciativas de las asociaciones de bibliotecarios americanos, y patrocinado por la *On Line Computer Library Center (OCLC)*, tiene su origen en un círculo intelectual de Dublin, en el estado de Ohio en Estados Unidos. El Dublin Core (DC) tiene como objetivo, definir un conjunto básico de atributos que sirvan para describir todos los recursos existentes en la red; este formato ha constituido un estándar de inestimable valor que ya no solo se utiliza en las Bibliotecas Digitales, donde tuvo su origen.

Para cumplir con su objetivo, el Dublin Core define un conjunto de quince elementos, que pueden modificarse y ampliarse debido a su flexibilidad, esto permite que los autores de las páginas Web puedan codificar sus documentos en el momento de generarlos. [Founier 2006]

El total de 15 elementos que conforma el formato DC (Ver Anexo 1) pueden clasificarse en tres grupos que indican la clase o el ámbito de la información que se guarda en ellos:

- Elementos relacionados principalmente con el contenido del recurso: título, tema, descripción, fuente, lenguaje, relación y cobertura.
- Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual: autor, editor, otras colaboraciones y derechos.
- Elementos relacionados principalmente con la temporalidad y formato del documento así como su identificación: fecha, tipo de recurso, formato, identificador del recurso.

El desarrollo y evolución que ha alcanzado el formato Dublin Core se debe al arduo y constante trabajo desarrollado por los diferentes grupos de trabajos vinculados a este formato en los diferentes talleres, seminarios y otras reuniones realizadas en función de lograr un estándar cuyas características garanticen la interoperabilidad en la recuperación e intercambio de información en la red.

La utilización de Dublin Core puede hacerse, incluyendo los metadatos dentro del propio recurso (a menudo se hace mediante las etiquetas META de la cabecera de las páginas HTML), o

bien, formando un registro separado con los metadatos. Los metadatos pueden redactarse manualmente (“picando a mano” el código) u obtenerse automáticamente a través de un software. Existen distintas herramientas para la asignación de metadatos y otras tareas conexas. [Founier 2006].

Existen gran número de proyectos, productores de contenidos y sitios que han tenido en cuenta el Dublin Core. También existen implementaciones donde los metadatos Dublin Core se utilizan como formato de intercambio común, donde se almacenen colecciones de metadatos.

Para concluir con Dublin Core, un ejemplo de metadatos sencillo:

```
<dc:title>Sistema de Información</dc:title>
<dc:description>Recuperación de Información</dc:description>
<dc:date>2008-05-07</dc:date>
<dc:creator>Jorge Leiva</dc:creator>
<dc:contributor>Maria Lopez</dc:contributor >
<dc:format>text/html</dc:format>
<dc:language>spa</dc:language>
<dc:publisher>UCI</dc:publisher>
```

1.4.3.2.2. RDF/XML.

Los *Resource Description Framework* (RDF) son una estructura flexible para la recuperación de información basada en Metadatos (describe, procesa e intercambia Metadatos). Este estándar facilita el intercambio de recursos en Internet. Es una tecnología XML (*eXtensible Markup Language*) creada para la publicación de titulares en la Web. El objetivo de los RDF es definir un mecanismo para describir recursos que no cree ninguna suposición sobre un dominio de aplicación particular, ni defina la semántica en dicho dominio.

Éstos fundamentalmente funcionan enviando consultas inteligentes a las bases de datos o “metabases”. Cuando se hace referencia a una consulta inteligente se trata de una consulta enriquecida con sinónimos y analizada por el motor de búsqueda solo por las raíces de sus palabras. Estas consultas inteligentes tienen mucho menor tiempo de respuesta y mucho más calidad en la información encontrada.

La interfaz con el usuario del prototipo RDF tiene cuatro componentes, la primera es una búsqueda inteligente, que permite la entrada de varias palabras llaves, para realizar una consulta; la segunda, permite la búsqueda en cada uno de los elementos de vocabulario de metadatos; la tercera es una búsqueda orientada a la navegación, desde una palabra llave inicial, permite recorrer las palabras relacionadas y a su vez muestra los artículos relacionados con esas palabras, los resultados aparecen en la cuarta componente de la interfaz. Los RDF están contruidos sobre las siguientes reglas:

- Un recurso es cualquier cosa que puede tener un URL, esto incluye todas las páginas Web, todos los elementos individuales de cada documento XML y mucho más.
- Una propiedad es un recurso que tiene un nombre y que puede usarse como tal, por ejemplo: autor o título. En muchos casos todo lo que nos importa en realidad es el nombre, pero una propiedad necesita ser un recurso de forma tal que pueda tener sus propias características.
- Una sentencia, consiste en la combinación de un recurso, una propiedad y un valor. Estas partes son conocidas como el sujeto, predicado y el objeto de la sentencia. [Baptista 2004]

También existe una sintaxis RDF básica para su implementación. La sentencia RDF extraordinariamente aparece en forma aislada, comúnmente varias propiedades de un recurso son indicadas al mismo tiempo. La sintaxis RDF/XML ha sido diseñada para acomodar esto con facilidad, agrupando varias sentencias sobre un mismo recurso en un elemento "Description". El elemento "Description" menciona en un atributo "about" el recurso para el cual las sentencias se aplican. Si el recurso no existe aún, el elemento "Description" puede asignarle un identificador en el momento usando un atributo ID.

Ejemplo:

Pedro Álvarez es el creador del recurso <http://www.cubasi.cu/home/bienvenido>.

```
<rdf:RDF>
  <rdf:Description about="http://www.cubasi.cu/home/bienvenido">
    <s:creator> Pedro Álvarez </s:creator>
  </rdf:Description>
</rdf:RDF>
```

El XML es un lenguaje que está diseñado con la intención de reemplazar al estándar actual HTML. Es un metalenguaje extensible de etiquetas desarrollado. En conclusión es un sencillo formato de texto, también derivado del SGML, y diseñado especialmente para documentos Web, un lenguaje universal de marcado para documentos estructurados y datos en la Web, más amplio, rico y dinámico que HTML. Permite la creación de etiquetas propias. El XML es considerado como un metalenguaje de definición de documentos estructurados mediante marcas o etiquetas. Se puede decir además, que es una norma recomendada para definir nuevos tipos de documentos, permitiendo al usuario definir sus propias etiquetas de marcado para extender las capacidades del HTML.

1.4.4. Recuperación de la Información.

Existen numerosas formas de lograr recuperar la información más importante o deseada de una pequeña o gran biblioteca de datos, una de las más factibles y utilizadas es la conocida como Indexación. [Pinto, 2007; Vallez, 2007]

La indexación, es la operación destinada a representar los resultados del análisis del contenido de un documento o de una parte del mismo, mediante elementos (denominados genéricamente términos de indexación) de un lenguaje documental o natural, generalmente para facilitar la recuperación. Los sistemas informáticos ya disponen de criterios de ordenación y búsquedas de palabras para muchos de los documentos que deben ser almacenados. Se recogen los datos y se añaden autónomamente al archivador como términos de índice. Se han propuesto tres nuevas formas de indexar los contenidos: indexación morfológica, sintáctica y la semántica, esta última también conocida como indexación basada en el sentido de las palabras.

1.4.4.1. La Indexación Semántica.

Aunque existen varias formas de realizar una indexación, la indexación basada en el significado de las palabras o semántica, ya sea en documentos o consultas, también tiene varios métodos. La referida a diccionarios tiene la ventaja de una buena organización y permite el tratamiento diferenciado a diferentes significados de una misma palabra, así se puede lograr el tratamiento de la *polisemia* y *homonimia*. Mucho más completa es la que se utiliza haciendo uso de una red semántica como WordNet organizada en conjuntos de sinónimos, conocida la técnica como

sysnsets, esta trata los fenómenos anteriores además de la *sinonimia*, *meronímia* e *hiponimia*, ya que en la base de datos se recogen también las relaciones entre dichos conjuntos de sinónimos.

Las relaciones de los términos se obtienen a partir de un análisis jerárquico del origen del mismo, estableciendo su significado, palabras subordinadas y relacionadas, proceso dado mediante la utilización de una herramienta terminológica denominada *tesauro* que permite la conversión del lenguaje natural (vocabulario manejado en la cotidianidad), al lenguaje normalizado (vocabulario estandarizado usado por los bibliotecólogos). La estructura básica de un tesauro es un vocabulario dinámico de términos relacionados semánticamente, aplicados a un dominio específico.

Una vez que se ha establecido un listado de palabras, se vinculan con un término dado (pregunta del usuario), se establece una matriz de 300 dimensiones, que es el listado de las mismas y se les asigna un valor singular, con el fin de extraer aquellas más cercanas, e indexarlas con los documentos que las contienen mediante la obtención de un espacio semántico. En el gráfico que se encuentra a continuación se describe con mayor detalle y paso a paso el proceso de indexación semántica realizado por la máquina. Ver Anexo 2.

Para entender mejor este proceso, se simulará una búsqueda: Por ejemplo, cuando se busca la palabra “software”:

- Primero se establece el significado de “software” (utilizando un diccionario de términos)
- Analiza por medio del tesauro cuáles son las palabras relacionadas, subordinadas y a qué familia terminológica pertenece (obteniendo palabras como aplicación, herramienta, informatizar, etc.).
- Con esta información se dirige a sus bases de datos e indexa los documentos que manejan el término y las palabras relacionadas.

La Indexación Semántica depende de un poderoso análisis matemático que es capaz de inferir correctamente relaciones muy estrechas a partir de un análisis estadístico de las palabras en uso, estableciendo una representación semántica. En la Tabla 2, se pueden observar las relaciones que se encuentran frente a la palabra “software”, recordando que entre más cerca se encuentre de 1 más acercamiento en cuanto a similitud, y por tanto más cerca de la respuesta a la pregunta formulada por el usuario (vectores que se encuentran cerca). En la columna 1 se localiza el número del ítem o palabra, en la columna dos el resultado del coseno de los dos vectores (palabras o textos) que

permiten dar origen a la columna tres que es listado de las palabras o vecindario semántico, que proporcionan gran información de la palabra.

#	Similitud	Término
1	0.99	Software
2	0.90	Programa
3	0.74	Aplicación
4	0.74	Aplicar
5	0.68	Herramienta
6	0.52	Informatizar
7	0.49	Proyecto
8	0.47	Código
9	0.42	Interfaz
10	0.38	Facilidad

Tabla 2: Relación de similitud de términos

A partir del cálculo del coseno de cada vector, se puede establecer sobre qué documentos conoce más indexación semántica, gracias a una medida denominada longitud del vector, donde entre más larga sea la longitud del mismo más conocimiento existe sobre la temática, esto genera que palabras como “el”, “de”, “los” entre otras se encuentren con vectores bajos de los que casi no se conoce. El modelo indexación semántica es de gran utilidad en la recuperación de información, ya que permite aprovechar de una mejor manera, las relaciones entre los términos de modo que se logre coherencia y comprensión textual.

1.4.4.2. La Búsqueda Semántica.

El conocimiento semántico es un proveedor líder del Procesamiento del Lenguaje Natural de software, incluidos los motores de búsqueda semántica, análisis de texto, búsqueda inteligente en el escritorio, Minería de Datos y de los sistemas de clasificación automática. La búsqueda semántica es una exploración en el lenguaje natural y de análisis documental basado en la comprensión de los contenidos a tratar. Permite clasificar rápidamente un conjunto de documentos textuales o una base de conocimiento de naturaleza tecnológica, de recursos humanos, etc.

Un ejemplo es cuando es posible extraer todos los documentos que hacen referencia al vocabulario informático, a la industria informática, etc. Otro de éstos, sería que a partir de "software" encontrara textos que contienen "aplicación" y "programa", incluso si la palabra "software " no se encuentra explícitamente en el texto.

1.5. Sistemas de Información.

El uso de la información puede tener poco o nulo sentido si no tiene un objetivo específico, y en términos empresariales, si no está conectado al sistema administrativo. Los tomadores de decisiones necesitan de la información para poder tener resultados. Cada vez se utiliza mejor la información, se recolecta mejor información y se procesa la misma de diversas fuentes de forma sistemática, lo que se le ha llamado inteligencia. Las empresas, sobre todo las grandes, no han adoptado esta práctica y continúan tomando decisiones sin procesar la información.

La facilidad de capturar y almacenar información en medios magnéticos, el crecimiento acelerado de la información almacenada y la necesidad de descubrimiento de nuevos conocimientos (deducidos o inducidos) constituyen los incentivos fundamentales en el desarrollo de la Minería de Datos. Aunque los formatos de almacenamiento de información son disímiles, es vital hurgar dentro de ellos y encontrar el conocimiento necesario.

Un Sistema de Información es un conjunto de componentes organizados, los cuales interactúan entre sí para procesar los datos y la información, la cual es distribuida de la manera más adecuada posible en una determinada organización en función de sus objetivos. También permiten capturar, procesar, almacenar y distribuir la información. Los Sistemas de Información realizan cuatro actividades fundamentales:

- **Entrada de información:** Es el proceso mediante el cual el Sistema de Información toma los datos que requiere para procesar la información. Las entradas pueden ser manuales o automáticas.
- **Almacenamiento de información:** El almacenamiento es una de las actividades o capacidades más importantes que tiene una computadora, ya que a través de esta propiedad el sistema puede recordar la información guardada en la sección o proceso anterior. Esta información suele ser almacenada en estructuras de información denominadas archivos.

- **Procesamiento de Información:** Es la capacidad del Sistema de Información para efectuar cálculos de acuerdo con una secuencia de operaciones preestablecidas. Esta característica de los sistemas permite la transformación de datos fuente en información que puede ser utilizada para la toma de decisiones.
- **Salida de Información:** La salida es la capacidad de un Sistema de Información para sacar la información procesada o bien datos de entrada al exterior. Es importante aclarar que la salida de un Sistema de Información puede constituir la entrada a otro sistema o módulo

1.6. Conclusiones.

Luego del análisis y estudio de la búsqueda semántica para agilizar dicho proceso, el discernimiento de la triada *dato-información-conocimiento*, el estudio de cuestionarios empleados para el Control Interno, los componentes de cada uno de éstos y la aplicación de la Lógica Borrosa en los cuestionarios se arriba a la conclusión que mediante el estudio y desarrollo de estas técnicas se podrá ofrecer solución al problema planteado.

CAPÍTULO 2: PROPUESTA DE SOLUCIÓN DEL PROBLEMA.

2.1. Introducción.

La propuesta contiene el cómo se puede llevar a cabo el mejoramiento del Control Interno y del autocontrol en cualquier entidad o empresa. Lo que permite procesar la información recogida mediante cualquier forma de recolección de datos, para crear indicadores que permitan no solo "cuantificar" el estado o el cambio, sino identificar aquellos aspectos más necesarios para mejorar el desempeño de la entidad.

La utilización de la confluencia de la Lógica Borrosa al tratamiento del Cuestionario Estándar del Control Interno, que responde a un estándar internacional pero que debe ser personalizado para discernir "más allá del cuestionario...", hace que surja la necesidad de crear una manera de tratar los datos y metadatos que se guardarán, o sea, del análisis de las posibles operaciones a desarrollar con esta información. Una vez almacenada la información es necesario buscar una vía para lograr recuperarla, mediante la búsqueda semántica utilizando la técnica de fichero invertido.

2.2. Cuestionario Ampliado para la Evaluación del Control Interno.

Llegando a la base del problema planteado referente a la evaluación de los elementos del CAECI, es necesario obtener clara y suficiente información de ella. Un elemento evaluado de Bien (B) básicamente no contiene mucha información que se pueda procesar, pues cabe analizar diferentes cuestiones: esa evaluación de B puede llegar a ser casi regular o casi excelente, esto no permite mostrar la medida exacta para conocer a qué conjunto de evaluación pertenece realmente el elemento. Un ejemplo parecido resulta de un vaso que contenga un poco de agua, se puede decir que el vaso se encuentra o medio vacío o medio lleno, dependiendo del punto de vista que se mire. Esto también puede depender de la preparación que tenga el evaluador o del estado anímico del mismo. Por eso es necesario poder contar con más elementos que ayuden a definir realmente la pertenencia a un conjunto específico, aquí es donde se necesita aplicar la Lógica Difusa o los árboles de decisión para lograr "jugar" con los datos y determinar la posibilidad, no la probabilidad, de pertenencia a un conjunto dado.

Al llevar el COSO al medio informático se obtiene el CECI donde cada elemento sería evaluado con un solo y único nivel, ya fuera de excelente (E), bien (B), regular (R), mal (M) o No Aplicado (N/A). La evaluación dada a cada elemento del Cuestionario Estándar del Control Interno se tratará como E_{ijk} , donde la i hace referencia a uno de los 5 componentes del control interno, la j a la cantidad de temáticas que componen a cada uno de ellos, y la k a los elementos de cada temática. En el CAECI se mantiene la misma política de evaluación para los elementos, solo que se modifica el esquema de evaluación E_{ijk} agregándole un nuevo valor, designado por δ , obtenido de la Lógica Difusa aplicada a los nuevos elementos, quedando finalmente como $(E_{ijk} + \delta)$.

Esta modificación ocurre porque el objetivo y aporte del CAECI es agregarle a cada uno de sus elementos nuevas vías para hacer más fiel la evaluación de los mismos, así como lograr realizar una búsqueda de consultas deseadas por los usuarios para poder procesar la información realmente útil. Básicamente serían nuevos elementos que permitirían marcar nuevas opciones, escribir comentarios y hasta adjuntar evidencia que corroboren la veracidad de la evaluación. Ahora bien, ésta se tiene que llevar a descripciones cuantitativas para el uso futuro más útil de esta información.

Uno de los grados de complicación más alto de este estudio está basado en el interés y la necesidad de la aplicación de la Lógica Borrosa al Cuestionario Ampliado para la Evaluación del Control Interno, el otro se refiere a la indexación semántica. Desde que se hizo este aporte a la Inteligencia Artificial su objetivo es dar mayor funcionalidad y veracidad a situaciones reflejantes de la realidad, pero esto no quiere decir que sea sencilla su aplicación.

Por ser el CECI un estándar internacional no se puede cambiar sus elementos, pero eso no quita que se puedan agregar nuevos elementos a cada uno de los anteriores del estándar, simplemente para recoger mayor información y al procesarla mediante Lógica Borrosa nos muestre el grado real de un bien, un mal o un excelente.

2.3. El Sistema de Ayuda para la Toma de Decisiones.

El resultado más importante del trabajo con un Sistema de Ayuda a la Toma de Decisiones o DSS es poder profundizar en el problema de decisión. Cuando se añade un sistema de normativas éste puede usarse como guía, tributando a la mejoría en la forma de pensar y forma de razonar en el tiempo por el usuario.

La información importante y necesaria se debe identificar, capturar y comunicar de una forma correcta y en un tiempo que permita a las personas realizar sus responsabilidades. Los sistemas de información producen informes sobre la información operacional y financiera que permiten hacer funcionar y controlar el negocio. Se ocupan no sólo de datos generados internamente, sino también información sobre los acontecimientos externos, de actividades y condiciones necesarias para la toma de decisiones del negocio y del reporte externo.

Los sistemas de ayuda a la toma de decisiones, se pueden dividir en tres componentes principales:

- **Sistema de Gestión de las Bases de Datos (SGBD):** Es la fuente o banco de datos para el DSS. Este componente proporciona al sistema una abstracción respecto a la forma de almacenamiento y procesamiento de los datos, separando a los usuarios de los aspectos físicos de la estructura, contenido y del análisis de la base de datos.
- **Sistema de Gestión de la Base del Modelo (SGBM):** El funcionamiento SGBM similar al SGBD. Su principal objetivo es dar independencia a los modelos definidos usados en el DSS y las aplicaciones que lo usan. Éste transforma los datos en información útil.
- **Sistema de Generación y Gestión del Diálogo (SGGD).** Es el componente que permite al usuario interactuar con el sistema. Como los usuarios son a menudo gestores sin formación informática, se necesitan interfaces intuitivos y fáciles de usar. Estos interfaces ayudan en la construcción del modelo y en la interacción con el mismo.

Existen múltiples tipos de DSS, los que generalmente coinciden en los tres componentes anteriores formando parte de sus arquitecturas y jugando un rol importante también en sus estructuras.

2.3.1. Sistema Gestor de Base de Datos (SGBD).

Se conoce como Sistema Gestor de Base de Datos a un conjunto de programas o software que permiten al usuario crear una base de datos, para poder mantener y asegurar la integridad total de sus datos, así como su seguridad. Éstos deben permitir en primer lugar definir la Base de Datos, lo que incluye especificar estructura, tipo y restricciones de datos. También permiten construirla que no

es más que lograr guardar los datos en algún medio o recurso controlado por el mismo SGBD. Y por último se debe poder manipular los datos guardados, es decir, realizar consultas, generar informes, actualizarla, etc. Ejemplos de gestores son Oracle y SQL Server de Microsoft.

Los SGBD deben contar con características deseables como:

- Controlar las redundancias: éstas tienen muchos efectos negativos como duplicación de trabajos al actualizarse, puede provocar datos inconsistentes, además de desperdiciar espacio en disco.
- Restringir los accesos no autorizados: se deben asignar permisos de accesos y autorización.
- Cumplir las restricciones de seguridad: el gestor debe ofrecer recursos para definir y garantizar el total cumplimiento de las restricciones de integridad.

El sistema de búsqueda que se desea crear parte de la idea de que es posible almacenar un fichero invertido procedente de una colección de documentos en una tabla, de manera que, a partir de ahí es posible calcular pesos de términos así como similitudes entre documentos y consultas.

Un fichero invertido, en su forma básica, no es más que una serie de entradas, una para cada término de la colección de documentos; para cada uno de estos términos se almacena una lista de los documentos en que aparece. Esta estructura puede mapearse simplemente a una tabla con dos campos: término y clave de documento. A partir de aquí es sencillo obtener información adicional para calcular pesos:

(select count(documento) from tabla;): el número de documentos en la colección.

(select termino, count (documento) from tabla group by termino;): el número de documentos en que aparece cada término.

2.3.1.1. Estructura básica de la Base de Datos.

La base de datos consta de varias tablas. Algunas podrían ser temporales y desaparecer una vez calculados los pesos, y dependen del esquema de cálculo concreto adoptado en cada ocasión; dado que no son necesarias para la resolución de consultas. Las tablas temporales podrían ser sustituidas por vistas, solución conceptualmente más elegante, pero que puede dar problemas de rendimiento e impide, en todo caso, la observación y manipulación de resultados intermedios. Las tablas básicas podrían ser algo como lo siguiente:

terminos(termino char(50), documento char(50), veces double)

pesos def(termino char(50), documento char(50), peso double)

El tamaño de los campos de tipo char podría ser más pequeño, obviamente: términos de 50 caracteres de longitud pueden considerarse directamente errores tipográficos, y las claves de los documentos pueden diseñarse para ocupar bastante menos espacio; incluso los términos podrían ser sustituidos por punteros o claves numéricas, pero eso restaría facilidades de observación.

De otro lado, el número de veces que un término aparece en un documento, que debería ser un entero, se establece como doble. La razón es que esto permite que el parser u otro proceso previo aplique, si se desea, algún tipo de coeficiente que prime de distinta forma los términos en función de distintos criterios (lugar del documento donde aparece, tipografía, función sintáctica, etc.) Para la resolución de consultas, en realidad, solo es precisa la tabla pesos def, pero la tabla términos puede ser interesante conservarla para posibles recalculados de pesos posteriores.

2.3.1.2. Cálculo de pesos.

El cálculo de pesos puede efectuarse en tres fases, una para cada componente del peso. Cada una de estas fases termina con una tabla temporal que recoge el componente calculado y que es usada en la fase siguiente; la tercera y última fase finaliza con la consecución de la tabla pesos def, con lo que esas tablas temporales dejan de ser necesarias. El sistema aquí empleado en tres fases puede resultar, en ocasiones, innecesariamente costoso. Por ejemplo, para operar con un esquema de pesos ntc (el más frecuente), no es preciso calcular frecuencia del término en el documento, ni almacenarla en ninguna tabla intermedia. Pero como se puede encontrar con cualquier esquema, y las combinaciones posibles son numerosas, parece más sensato y más manejable, operar en esas tres fases.

2.3.1.3. La frecuencia de términos.

Hay diversas maneras de estimar este componente del peso, y que no tiene por qué coincidir con el número de veces que cada término aparece en cada documento; en cualquier caso, el objetivo, en esta fase, es doble: por un lado, obtener una tabla temporal con estas frecuencias, calculadas en la forma en que se desee, que pueda ser utilizada en las sucesivas fases del cálculo de pesos, por otro, mantener los datos de entrada originales en su propia tabla, de forma que

puedan ser reutilizados (por ejemplo, para recalcular pesos con otro esquema distinto). Así, la esencia de esta fase es una simple sentencia SQL:

Create table frecuencias as select...

El contenido concreto depende del esquema de cálculo aplicado, pero, en general, se tratará de un select sobre la tabla “términos” que contiene los datos originales de entrada. Con algunos esquemas, que utilizan, por ejemplo, parámetros como la frecuencia máxima en el documento, es preciso algún paso intermedio que calcule tales elementos. Para los esquemas más conocidos, la solución requiere básicamente sentencias SQL y obviamente, se deberían construir índices sobre los campos término y documento para agilizar las operaciones.

2.3.1.4. Pesos de las consultas.

Los pesos de los términos de las consultas pueden estimarse de la misma manera que los de los documentos, aunque, como ya se ha dicho, aplicando esquemas que no tienen por qué ser iguales. El hecho de que, al procesarse una sola consulta de cada vez, el volumen de información involucrado sea mucho menor, puede aconsejar buscar métodos más ágiles para calcular los pesos de los términos de las consultas. De hecho, en sistemas interactivos las consultas suelen ser muy cortas (2 ó 3 términos); en muchos sistemas ni siquiera se calculan pesos para las consultas. Una posibilidad, dependiente del rendimiento del sistema SQL que se utilice, es el uso de vistas; esto permite, al tiempo que se indexa la colección de documentos, dejar construidas las vistas necesarias para calcular los pesos de los términos de las consultas. En el momento de la consulta, estas vistas se ejecutan, obteniendo los pesos correspondientes. De una forma u otra, el producto de este proceso es una tabla o vista con dos campos: término y peso.

2.3.1.5. Resolución de consultas.

La resolución de consultas es muy sencilla, una vez que se cuenta con los pesos de los términos. Básicamente, se trata de localizar los documentos con algún término en común con la consulta, ya una vez localizados, calcular un coeficiente de similitud entre cada uno de esos documentos y la consulta. Posteriormente, se deben ordenar esos documentos de forma decreciente en función de ese coeficiente de similitud.

El coeficiente de similitud más habitual consiste en el producto de los vectores, analizado anteriormente. Dado que los pesos de esos vectores han sido normalizados previamente, se puede resolver una consulta cualquiera mediante una sentencia como la que sigue:

```
select pesos_def.documento,  
sum(pesos_def.peso*pesos_c.peso) as simil  
from pesos_def, pesos_c  
where pesos_def.termino=pesos_c.termino  
group by pesos_def.documento order by simil DESC ;
```

2.3.2. Sistema Gestor de Base de Modelos (SGBM).

El papel del SGBM es análogo al del SGBD. Su objetivo es transformar datos en información útil para la toma de decisiones, proporcionando independencia entre los modelos específicos utilizados en el DSS y las aplicaciones que presentan los resultados. También puede asistir en la construcción de los modelos, especialmente cuando éstos son muy complejos.

La superioridad de los modelos, incluso lineales simples, sobre el juicio intuitivo humano sugiere que un modo de mejorar la calidad de las decisiones es descomponer el problema de la decisión en componentes simples que sean bien entendidos y definidos. El estudio de un sistema complejo construido con tales componentes puede ser ayudado por una técnica formal teóricamente sólida. Esto permite aplicar conocimiento científico que puede ser transferido a otros problemas (o incluso dominios) permitiendo analizar, explicar y argumentar acerca del problema de decisión.

En las diferentes herramientas de modelado desarrolladas, el conocimiento acerca del sistema es representado por ecuaciones o reglas lógicas, mejoradas con una representación explícita de la incertidumbre. Una vez que un modelo ha sido formulado, pueden usarse una variedad de métodos matemáticos para analizarlo. La toma de decisiones bajo incertidumbre, introduce mejoras en base a aproximaciones estadísticas, tales como análisis de la confianza, simulación, y toma de decisiones basada en estadísticas.

2.3.2.1. Modelo Árbol de Decisión.

Una de las dificultades de la actual investigación está en complementar toda la información recogida en una sola, y una de las vías para lograrlo puede ser una de las técnicas de Inteligencia Artificial conocida como Árboles de Decisión. Mediante este modelo se permitirá determinar la evaluación real de alguno de los elementos de forma cualitativa.

Para hacer más sencillo el ejemplo solo se toman 10 elementos de los 116 que contiene el CAECI, cuatro condiciones o aspectos a evaluar en cada uno de los nuevos elementos y dos evaluaciones (E y B) de las 5 posibles.

N = 10 elementos

r = 4 → C₁ = ¿Cumple la condición 1?

C₂ = ¿Cumple la condición 2?

C₃ = ¿Cumple la condición 3?

C₄ = ¿Cumple la condición 4?

Grupos → Π₁ = "E"

Π₂ = "B"

La población S está formada por 10 elementos, S = {O₁,...,O₁₀} y se toma la subpoblación S₁ como aquella donde sus elementos cumplan con la condición C_n y S₂, las que no las cumplan.

	C1	C2	C3	C4	Grupo
1	Si	Si	Si	No	Π ₁
2	No	Si	No	No	Π ₁
3	No	No	Si	Si	Π ₁
4	Si	No	Si	No	Π ₁
5	Si	No	No	Si	Π ₁
6	Si	Si	Si	Si	Π ₁
7	No	Si	Si	Si	Π ₂
8	No	No	Si	No	Π ₂
9	Si	No	No	No	Π ₂
10	No	No	No	No	Π ₂

Tabla 3. Clasificación de los casos asociados a la Tabla 1.

Ramificando por C_1 :

$$S_1 = \left\{ \underbrace{0_1, 0_4, 0_5, 0_6, 0_9}_{\Pi_1} \right\} \rightarrow e(S_1) = -\left(\frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) = -\frac{1}{\log_{10} 2} \left(\frac{4}{5} \log_{10} \frac{4}{5} + \frac{1}{5} \log_{10} \frac{1}{5} \right) = 0,72$$

$$S_2 = \left\{ \underbrace{0_2, 0_3, 0_7, 0_8, 0_{10}}_{\Pi_1} \right\} \rightarrow e(S_2) = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5} \right) = -\frac{1}{\log_{10} 2} \left(\frac{2}{5} \log_{10} \frac{2}{5} + \frac{3}{5} \log_{10} \frac{3}{5} \right) = 0,97$$

$$e^*(S) = \frac{5}{10} \cdot e(S_1) + \frac{5}{10} \cdot e(S_2) = 0,5 \cdot 0,72 + 0,5 \cdot 0,97 = 0,85$$

Ramificando por C_2 :

$$S_1 = \left\{ \underbrace{0_1, 0_2, 0_6, 0_7}_{\Pi_1} \right\} \rightarrow e(S_1) = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) = -\frac{1}{\log_{10} 2} \left(\frac{3}{4} \log_{10} \frac{3}{4} + \frac{1}{4} \log_{10} \frac{1}{4} \right) = 0,81$$

$$S_2 = \left\{ \underbrace{0_3, 0_4, 0_5, 0_8, 0_9, 0_{10}}_{\Pi_1} \right\} \rightarrow e(S_2) = -\left(\frac{3}{6} \log_2 \frac{3}{6} + \frac{3}{6} \log_2 \frac{3}{6} \right) = -\frac{1}{\log_{10} 2} \left(\frac{3}{6} \log_{10} \frac{3}{6} + \frac{3}{6} \log_{10} \frac{3}{6} \right) = 1$$

$$e^*(S) = \frac{4}{10} \cdot e(S_1) + \frac{6}{10} \cdot e(S_2) = 0,4 \cdot 0,81 + 0,6 \cdot 1 = 0,92$$

Ramificando por C_3 :

$$S_1 = \left\{ \underbrace{0_1, 0_3, 0_4, 0_6, 0_7, 0_8}_{\Pi_1} \right\} \rightarrow e(S_1) = -\left(\frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6} \right) = -\frac{1}{\log_{10} 2} \left(\frac{4}{6} \log_{10} \frac{4}{6} + \frac{2}{6} \log_{10} \frac{2}{6} \right) = 0,91$$

$$S_2 = \left\{ \underbrace{0_2, 0_5, 0_9, 0_{10}}_{\Pi_1} \right\} \rightarrow e(S_2) = -\left(\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right) = -\frac{1}{\log_{10} 2} \left(\frac{2}{4} \log_{10} \frac{2}{4} + \frac{2}{4} \log_{10} \frac{2}{4} \right) = 1$$

$$e^*(S) = \frac{6}{10} \cdot e(S_1) + \frac{4}{10} \cdot e(S_2) = 0,6 \cdot 0,91 + 0,4 \cdot 1 = 0,95$$

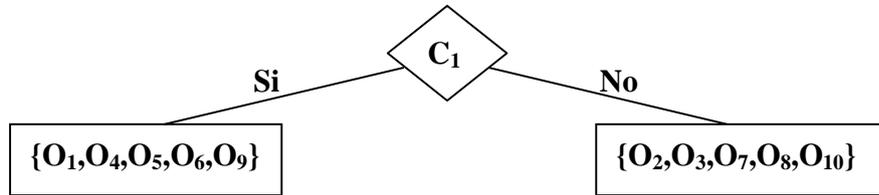
Ramificando por C_4 :

$$S_1 = \left\{ \underbrace{0_2, 0_3, 0_5, 0_6, 0_7}_{\Pi_1} \right\} \rightarrow e(S_1) = -\left(\frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) = -\frac{1}{\log_{10} 2} \left(\frac{4}{5} \log_{10} \frac{4}{5} + \frac{1}{5} \log_{10} \frac{1}{5} \right) = 0,72$$

$$S_2 = \left\{ \underbrace{0_1, 0_4, 0_8, 0_9, 0_{10}}_{\Pi_1} \right\} \rightarrow e(S_2) = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5} \right) = -\frac{1}{\log_{10} 2} \left(\frac{2}{5} \log_{10} \frac{2}{5} + \frac{3}{5} \log_{10} \frac{3}{5} \right) = 0,97$$

$$e^*(S) = \frac{5}{10} \cdot e(S_1) + \frac{5}{10} \cdot e(S_2) = 0,5 \cdot 0,72 + 0,5 \cdot 0,97 = 0,85$$

Se puede ramificar por las características C_1 o C_4 , ya que aquí se obtiene las entropías mínimas (0,85). Se escoge C_1 y se divide su población inicial según al grupo al que pertenece.



Se repite el estudio pero tomando la siguiente población $S = \{O_1, O_4, O_5, O_6, O_9\}$ obtenida de la rama positiva de C_1 , entonces se prueban a ramificar por las demás características buscando la mínima entropía.

Ramificando por C_2 :

$$S_1 = \left\{ \begin{matrix} o_1, o_6 \\ \Pi_1 \end{matrix} \right\} \rightarrow e(S_1) = -(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = -\frac{1}{\log_{10} 2} (1 \cdot \log_{10} 1 + 0 \cdot \log_{10} 0) = 0$$

$$S_2 = \left\{ \begin{matrix} o_4, o_5, o_9 \\ \Pi_1 \end{matrix} \right\} \rightarrow e(S_2) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) = -\frac{1}{\log_{10} 2} \left(\frac{2}{3} \log_{10} \frac{2}{3} + \frac{1}{3} \log_{10} \frac{1}{3} \right) = 0,91$$

$$e^*(S) = \frac{2}{5} \cdot e(S_1) + \frac{3}{5} \cdot e(S_2) = 0,4 \cdot 0 + 0,6 \cdot 0,91 = 0,55$$

Ramificando por C_3 :

$$S_1 = \left\{ \begin{matrix} o_1, o_4, o_6 \\ \Pi_1 \end{matrix} \right\} \rightarrow e(S_1) = -(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = -\frac{1}{\log_{10} 2} (1 \cdot \log_{10} 1 + 0 \cdot \log_{10} 0) = 0$$

$$S_2 = \left\{ \begin{matrix} o_5, o_9 \\ \Pi_1 \end{matrix} \right\} \rightarrow e(S_2) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = -\frac{1}{\log_{10} 2} \left(\frac{1}{2} \log_{10} \frac{1}{2} + \frac{1}{2} \log_{10} \frac{1}{2} \right) = 1$$

$$e^*(S) = \frac{3}{5} \cdot e(S_1) + \frac{2}{5} \cdot e(S_2) = 0,6 \cdot 0 + 0,4 \cdot 1 = 0,40$$

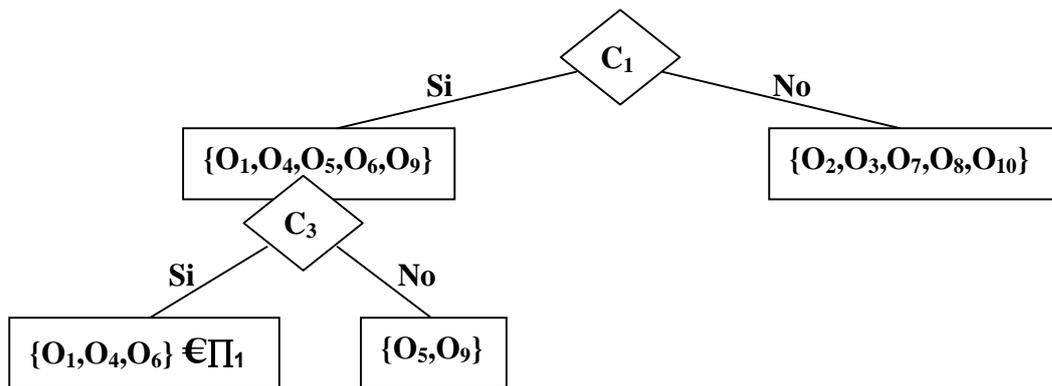
Ramificando por C_4 :

$$S_1 = \left\{ \begin{matrix} o_5, o_6 \\ \Pi_1 \end{matrix} \right\} \rightarrow e(S_1) = -(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = -\frac{1}{\log_{10} 2} (1 \cdot \log_{10} 1 + 0 \cdot \log_{10} 0) = 0$$

$$S_2 = \left\{ \begin{matrix} o_1, o_4, o_9 \\ \Pi_1 \end{matrix} \right\} \rightarrow e(S_2) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) = -\frac{1}{\log_{10} 2} \left(\frac{2}{3} \log_{10} \frac{2}{3} + \frac{1}{3} \log_{10} \frac{1}{3} \right) = 0,91$$

$$e^*(S) = \frac{2}{5} \cdot e(S_1) + \frac{3}{5} \cdot e(S_2) = 0,4 \cdot 0 + 0,6 \cdot 0,91 = 0,55$$

Se ramifica por la característica C₃, con la que se obtiene la entropía mínima (0,40) y se vuelve a dividir su población según al grupo al que pertenece cada uno de sus elementos.



Se repite el estudio pero tomando la siguiente población $S = \{O_2, O_3, O_7, O_8, O_{10}\}$ obtenida de la rama negativa de C₁.

Ramificando por C₂:

$$S_1 = \left\{ \underset{\Pi_1}{o_2, o_7} \right\} \rightarrow e(S_1) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = -\frac{1}{\log_{10} 2} \left(\frac{1}{2} \log_{10} \frac{1}{2} + \frac{1}{2} \log_{10} \frac{1}{2} \right) = 1$$

$$S_2 = \left\{ \underset{\Pi_1}{o_3, o_8, o_{10}} \right\} \rightarrow e(S_2) = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3} \right) = -\frac{1}{\log_{10} 2} \left(\frac{1}{3} \log_{10} \frac{1}{3} + \frac{2}{3} \log_{10} \frac{2}{3} \right) = 0,91$$

$$e^*(S) = \frac{2}{5} \cdot e(S_1) + \frac{3}{5} \cdot e(S_2) = 0,4 \cdot 1 + 0,6 \cdot 0,91 = 0,95$$

Ramificando por C₃:

$$S_1 = \left\{ \underset{\Pi_1}{o_3, o_7, o_8} \right\} \rightarrow e(S_1) = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3} \right) = -\frac{1}{\log_{10} 2} \left(\frac{1}{3} \log_{10} \frac{1}{3} + \frac{2}{3} \log_{10} \frac{2}{3} \right) = 0,91$$

$$S_2 = \left\{ \underset{\Pi_1}{o_2, o_{10}} \right\} \rightarrow e(S_2) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = -\frac{1}{\log_{10} 2} \left(\frac{1}{2} \log_{10} \frac{1}{2} + \frac{1}{2} \log_{10} \frac{1}{2} \right) = 1$$

$$e^*(S) = \frac{3}{5} \cdot e(S_1) + \frac{2}{5} \cdot e(S_2) = 0,6 \cdot 0,91 + 0,4 \cdot 1 = 0,95$$

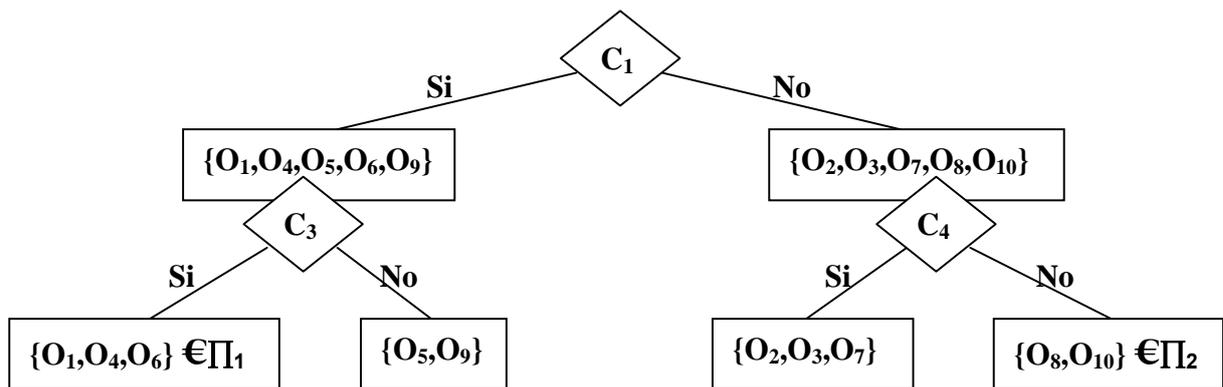
Ramificando por C_4 :

$$S_1 = \left\{ \underbrace{o_2, o_3, o_7}_{\Pi_1} \right\} \rightarrow e(S_2) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) = -\frac{1}{\log_{10} 2} \left(\frac{2}{3} \log_{10} \frac{2}{3} + \frac{1}{3} \log_{10} \frac{1}{3} \right) = 0,91$$

$$S_2 = \left\{ \underbrace{o_8, o_{10}}_{\Pi_1} \right\} \rightarrow e(S_1) = -(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = -\frac{1}{\log_{10} 2} (1 \cdot \log_{10} 1 + 0 \cdot \log_{10} 0) = 0$$

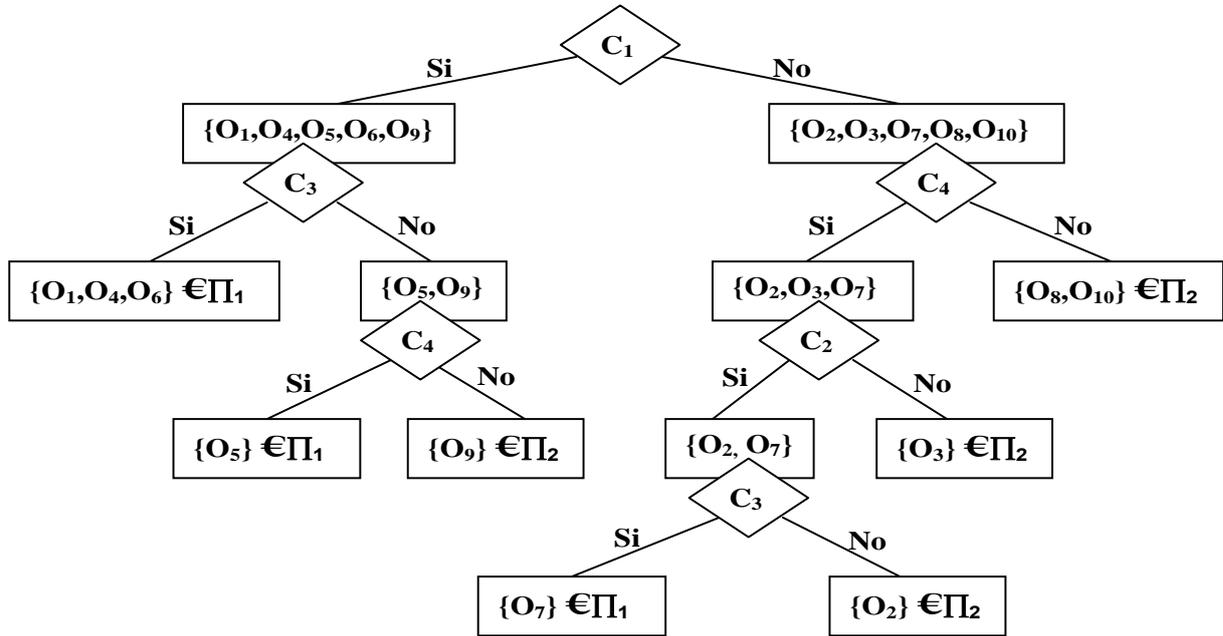
$$e^*(S) = \frac{3}{5} \cdot e(S_1) + \frac{2}{5} \cdot e(S_2) = 0,6 \cdot 0,91 + 0,4 \cdot 0 = 0,55$$

Se ramifica por la característica C_4 , con la que se obtiene la entropía mínima (0,55) y se vuelve a dividir su población según al grupo al que pertenece cada uno de sus elementos.



Por dos de las cuatro ramas que se tienen ya se ha terminado (por la positiva de C_3 y por la negativa de C_4), puesto que está claro a que grupo Π_i pertenecen los elementos. Por las otras dos ramas que quedan se sigue con la misma metodología.

Ahora se tienen la subpoblaciones $S_1 = \{O_5, O_9\}$ y $S_2 = \{O_2, O_3, O_7\}$ obtenidas de las demás ramas. Luego se continúa aplicando el mismo procedimiento anterior para calcular las entropías de las subpoblaciones hasta que todos los elementos de muestra tengan definida su pertenencia. Se construye finalmente el árbol como muestra la figura.



Resumiendo, se ha obtenido una nueva información:

- Si $C_1=Si \wedge C_3=Si \Rightarrow \Pi_1$
- Si $C_1=Si \wedge C_3=No \wedge C_4=Si \Rightarrow \Pi_1$
- Si $C_1=Si \wedge C_3=No \wedge C_4=No \Rightarrow \Pi_2$
- Si $C_1=No \wedge C_4=No \Rightarrow \Pi_2$
- Si $C_1=No \wedge C_4=Si \wedge C_2=No \Rightarrow \Pi_2$
- Si $C_1=No \wedge C_4=Si \wedge C_2=Si \wedge C_3=No \Rightarrow \Pi_2$
- Si $C_1=No \wedge C_4=Si \wedge C_2=Si \wedge C_3=Si \Rightarrow \Pi_1$

Esta información, proporcionado por el algoritmo ID3, se puede usar para clasificar a futuros individuos $O = \{X_1, \dots, X_r\}$ en los que se desconoce a qué grupo Π_j realmente pertenecen. Para ello una vez construido el árbol, desde cada uno de los nodos terminales del mismo, se construye un nodo de decisión con p ramas (“clasifica en Π_1 ”, ..., “clasifica en Π_p ”) de cada una de las cuales sale un nodo de eventos con p ramas (“es de Π_1 ”, ..., “es de Π_p ”).

2.3.2.2. Modelo del Indicador.

La evaluación cualitativa E_{ijk} de los elementos del CAECI para que aporten mayor cantidad de información deben ser llevadas a valores entre 0 y 1, sin olvidar que es necesario que exista un rango de esos valores que permita al indicador de la evaluación, dada inicialmente, “desplazarse” según el valor de δ dado por Lógica Difusa y defina si se mantiene en el mismo o cambia para otro nivel de evaluación. El valor igual a cero sería N/A, entre 0,01 y 0,25 es para una evaluación de M, entre 0,26 y 0,50 de R, entre 0,51 y 0,75 de B y entre 0,76 y 1,0 de E. El indicador cuantitativo de cada evaluación dada, sería el punto medio del rango correspondiente a dicha evaluación desde donde comenzaría a “desplazarse” para llegar a la verdadera evaluación final, después de sumársele o restársele el valor de δ .

Por ejemplo, al evaluar el elemento $E_{132} = B$ se está refiriendo a una evaluación de **B** dada al segundo elemento de la tercera temática del primer componente referido al Ambiente de Control. Aquí se tendría que el indicador de $E_{132} = 0,63$, pues sería el punto medio dentro del rango de (0,51-0,75). Al procesarse en el CAECI, los elementos extras de E_{132} , se tiene que $\delta = +0,15$, por lo tanto, $E_{ijk} + \delta = 0,63 + (+0,15) = 0,78$, ya el indicador sobrepasaría el rango de la evaluación de B y pertenecería a una evaluación de E. Si existiera el caso donde el indicador quedara por debajo de 0 o por encima de 1, se acota el valor a 0 ó 1, respectivamente.

2.3.2.3. Modelo del Espacio Vectorial.

El modelo vectorial se utiliza para asignar pesos a cada término de una consulta o documento, para más tarde realizar la búsqueda que se desee utilizando los mismos. Cada elemento está basado en medidas, que aunque sean diferentes formas de precisar sus características no dejan de ser mediciones. Para llegar a determinar el grado de similitud entre documentos o elementos con las consultas a realizar se construyen modelos para seleccionar cuáles de estos son relevantes y cuáles no. Una buena propuesta utilizado en la herramienta Lemur es el modelo vectorial formulado en los años 70 por G. Salton y muy aplicado desde entonces.

En este modelo se tiene un documento d_j y una consulta q que son representados como vectores en un espacio n -dimensional, siendo “ n ” el número total de términos que forman el vocabulario o número de términos indexables que existen en la colección documental. Si la palabra

no está presente en el documento, ese valor es igual a 0. En caso contrario, ese valor es calculado teniendo en cuenta diversos factores, dado que una palabra dada puede ser más o menos significativa (tanto en general como, sobre todo, en ese documento en concreto); este valor se conoce con el nombre de peso del término en el documento. Hay, pues, un vector para cada documento, y, en cada vector, un elemento o para cada término palabra susceptible de aparecer en el documento. Cada uno de esos elementos es cubierto u ocupado con un valor numérico. Siempre según el modelo del espacio vectorial, las consultas son representadas también mediante un vector de las mismas características que las de los documentos (variando los valores numéricos de cada elemento en función de las palabras que forman parte de la consulta). Esto permite calcular fácilmente una función de similaridad dada entre el vector de una consulta (q) y los de cada uno de los documentos (d_j). La relación que existe entre los vectores d_j y q permite determinar el grado de similitud que presenta un documento con la consulta, esta está dada por el coseno del ángulo que los relaciona:

$$Sim(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} = \text{Cos } \alpha$$

Un documento es, entonces, un vector $d = (w_1, w_2, \dots, w_n)$, donde “n” es el número de términos de la colección y w_i es el peso del i -ésimo término en el documento d . Un documento puede ser representado por una lista de palabras o por un vector, cada una de sus palabras tiene un peso dentro del documento, o lo que es igual, un coeficiente que expresa la medida en que esa palabra representa el contenido del documento analizado. Naturalmente, se asume que la consulta se formula en lenguaje natural (podría ser, incluso un documento de muestra, para recuperar los que fuesen parecidos a él), y, de lo dicho se deduce que el resultado de la consulta consiste en una lista de documentos ordenada en orden decreciente en función de su similaridad con la consulta. Diversos factores son modificables dentro de este modelo de recuperación. Así, entre otros, el sistema de cálculo de pesos. Los esquemas habituales se basan, de una u otra forma, en las frecuencias de aparición de la palabra cuyo peso se quiere calcular. Al lograr obtener este coeficiente que represente la similitud entre el vector consulta y cada uno de los de cada documento, los de más alto índice indicarían una mejor respuesta a la consulta del usuario. Del mismo modo, es también

fácilmente sustituible la ecuación de cálculo de similaridad. Las palabras indexables necesitan ser normalizadas, lo que se consigue a través de la lematización, aunque ésta puede llevarse a cabo también mediante un preprocesado externo de los documentos, o ceñirla sólo a las consultas. Luego se indexa el título seleccionado mediante un sistema de recuperación de la información y se ordenan en orden decreciente de similitud. [Pinto 2007]

Partiendo de que se puede representar los documentos como vectores de términos, entonces los documentos pueden situarse en un espacio vectorial de n dimensiones, es decir, con tantas dimensiones como elementos tenga el vector. Matemáticamente hace tiempo que se trabaja con espacios de n dimensiones, dando a “ n ” un valor superior a tres. Situado en ese espacio vectorial, cada documento cae entonces en un lugar determinado por sus coordenadas, al igual que en un espacio de tres dimensiones cada objeto queda bien ubicado si se especifica sus tres coordenadas espaciales (Ver Anexo 3). Se crean así grupos de documentos que quedan próximos entre ellos a causa de las características de sus vectores. Estos grupos o *clusters* están formados, en teoría, por documentos similares, es decir, por grupos de documentos que son relevantes para la misma clase de problemas de información.

2.3.2.3.1. Asignación de pesos a los términos.

Para poder realizar la indexación mediante el modelo vectorial es necesario asignar un peso determinado a cada término importante tanto de los documentos como de las consultas. Existen múltiples técnicas para lograr este objetivo.

- *Booleano*: Los pesos de los términos del documento solo son 0 ó 1, lo que indica la presencia o ausencia del término en el mismo.
- *Frecuencia de un término* o *TF (Term Frequency)*: Es la forma más simple de pesar un término. Cada uno de los términos tienen una importancia proporcional a la cantidad de veces que aparece en un documento $TF(d,t)$. El peso de un término t en un documento d es $w_t = TF(d,t)$.

Como lo que se necesita es una asignación de pesos y no solo conocer si existe el término o no, entonces la mejor técnica a utilizar sería la TF. Hay que señalar que es muy importante normalizar de alguna manera la frecuencia interna de un término en un documento, para moderar el efecto de las altas frecuencias (por ejemplo, el término ‘el’ que aparece 18 veces no es más importante que el término ‘informática’ que aparece 6 veces) y para compensar la longitud del

documento (en documentos más largos, previsiblemente más veces aparecerá cada término). Además, algunas colecciones tienen documentos muy cortos y no tienen sentido las frecuencias dentro de uno de ellos. El propósito de la normalización es lograr que el peso o importancia de un término no dependa de la frecuencia de su ocurrencia relativa con los otros términos. Pesar un término por la frecuencia absoluta obviamente tiende a favorecer los documentos más extensos sobre los menos extensos.

2.3.2.3.2. Normalización de la frecuencia de términos.

Existen dos formas de normalizar, una de ellas es la *Normalización por la longitud del vector*, y la otra es la *Normalización de la frecuencia de los términos*. En esta última el TF se divide por la *Frecuencia Máxima* de todos los términos, logrando que el peso esté entre 0 y 1.

$$\text{Normalización} = (TF / TF_{\max})$$

Se conoce que la frecuencia de un término mejora la relevancia en la recuperación de información, pero no siempre mejora la precisión. Como los términos frecuentes tienden a aparecer en muchos textos, éstos tienen poco poder discriminante. Para resolver esto, se eliminan los términos de alta frecuencia.

Mientras el *TF* tiene que ver con la frecuencia de un término en un documento, la Inverse Document Frequency (IDF) tiene que ver con la frecuencia de un término en un conjunto de documentos. La importancia del término es inversamente proporcional al número de documentos que contiene al término:

$$w_t = TF(d, t) * IDF(t) ;$$

$$IDF(t) = \log\left(\frac{m}{df(t)}\right)$$

De aquí se conoce que *m* es la cantidad de documentos de la colección y *df(t)* es el número de documentos que contienen al término *t* en dicha colección. Mientras menos documentos contengan al término *t* mayor es el *IDF(t)*. Por el contrario, si todos los documentos de la colección contienen al término *t* entonces *IDF(t)* es cero. El factor *TF(d,t)* contribuye a mejorar la relevancia y el factor *IDF(t)* contribuye a mejorar la precisión, pues él representa la especificidad del término, distinguiendo los documentos en los que éste aparece de aquellos en los que no aparece. El *IDF(t)* es útil como indicador de la bondad del término *t* como discriminador de documentos. Esto expresa la idea

intuitiva de que un término que ocurra en toda la colección no es útil para distinguir documentos relevantes de los no relevantes.

Por ejemplo, en una colección de documentos que tratan sobre la informática, el término Información probablemente aparezca en todos y, por tanto, no será bueno para discriminar los documentos relevantes a una consulta. Por otra parte, si la colección de documentos abarca una temática muy general, entonces sí será útil para distinguir a los documentos que hablen sobre Informática. La combinación del TF con el IDF da mayor importancia a los términos que ocurren frecuentemente en el documento e infrecuentemente en la colección.

2.3.2.3. Operaciones sobre los términos.

No todas las palabras o términos que componen un documento deben ser considerados términos de indexación. Los términos sufren una serie de transformaciones antes de incluirse en el índice. Si se toman como elementos de indexación las palabras que aparecen en el texto se analizan, entonces, las operaciones que pueden realizarse sobre los términos (Figura 6):

- Identificación de los signos de puntuación y espaciados, eliminación de los acentos (uno de los eternos problemas en el procesamiento del lenguaje natural en español), reducción de las mayúsculas y las minúsculas, reconocimiento del formato del documento (por ejemplo, si es una página HTML se eliminan las etiquetas), reconocimiento de las palabras, etc. Cuando termina esta etapa se tiene el texto plano y las palabras identificadas en él.
- Eliminación de palabras vacías (*stopwords*): La lista de palabras vacías o antidiccionario es una relación de términos considerados como valores no indexables, usados para eliminar potenciales términos de indexación. Los términos de una lista vacía están carentes de todo significado a la hora de recuperar información, como, por ejemplo, el artículo "la" no posee ninguna funcionalidad a la hora de recuperar documentos, ya que en todos los documentos de la base de datos aparecerá este término de forma casi segura y no resalta nada el contenido del documento almacenado. Esta lista vacía está formada por las preposiciones, conjunciones, artículos, pronombres, etc.

- Extracción de raíces o lemas (*stemming* o lematización): Los algoritmos de extracción de raíces se encuentran orientados a obtener un único término a partir de diferentes palabras que constituyen esencialmente variaciones morfológicas con un mismo significado. Por ejemplo, se puede considerar la obtención del término “*producto*” a partir de “*producción*” y “*producido*”. En el caso de los verbos, por ejemplo, se obtiene el infinitivo “*producir*” a partir de “*produzco*” y “*producirá*”.
- Utilización de un tesoro: Un tesoro proporciona una agrupación o clasificación de términos en un determinado dominio o área en categorías denominadas clases. Pueden utilizarse diversos tipos de 50 tesauros en la indexación automática o manual de los documentos. En una aproximación basada en el análisis automático de textos y consultas, la utilización del tesoro permite identificar términos lexicográficamente diferentes pero equivalentes semánticamente (sinónimos). De esta forma, se puede conseguir disminuir el problema que plantea la sinonimia en el proceso de recuperación: documentos que son adecuados a una consulta del usuario no se recuperan por no aparecer en ellos los términos de la consulta, aunque sí aparezcan sinónimos suyos.

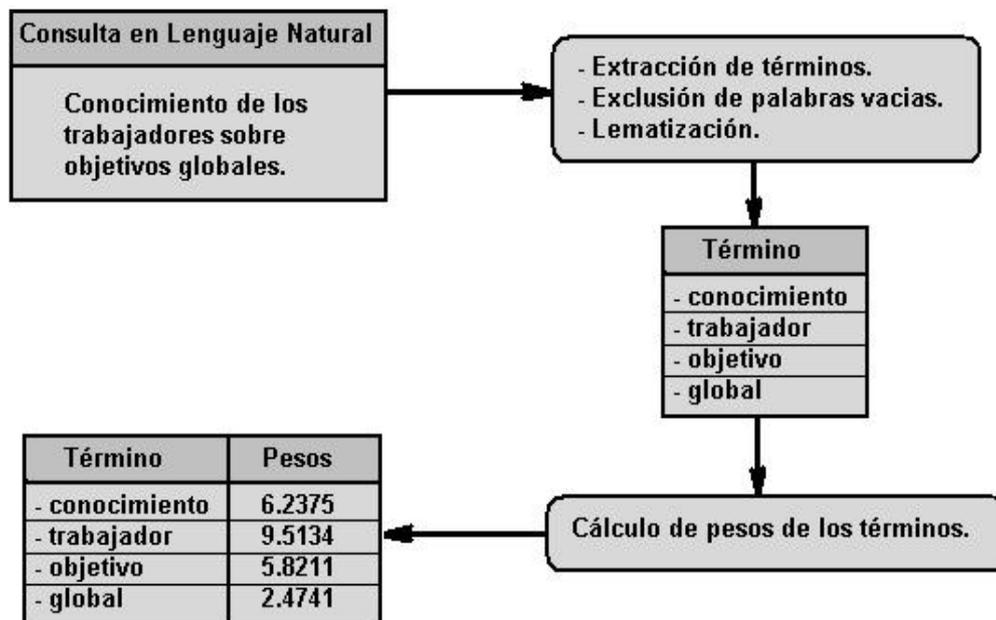


Figura 6. Preprocesado de consultas.

2.3.2.3.4. Fichero invertido.

Hoy en día la utilización de la técnica del fichero invertido es un elemento clásico en los sistemas de recuperación de información textual. Dada la gran cantidad de información contenida en los documentos textuales, ya sea un fichero Word o una página Web, los procedimientos clásicos de búsqueda secuencial o de ficheros indexados no son capaces de responder de manera adecuada a los requerimientos necesarios de velocidad y exactitud en la respuesta para satisfacer al usuario. Por esta razón, estos sistemas usan una especialización de los ficheros indexados, desarrollando un nuevo tipo de fichero, al que se conoce como fichero invertido.

Como se ha analizado anteriormente ya en muchos estudios y programas se han añadido funciones que permiten calcular el grado de relevancia de un documento para satisfacer una consulta. En otros se calcula la ocurrencia de términos, permitiendo recuperar documentos similares tomando como modelo uno de ellos, o una porción de texto. Para estos y otros cálculos, se suele utilizar la llamada función inversa del documento, que relaciona la frecuencia absoluta del término en cada documento con la frecuencia de aparición del término en el conjunto de la base de datos. En general, la gestión documental está marcada por un doble factor que los buenos programas de RI deben contemplar: primero, que el texto es uno de los medios más poderosos para representar y acceder a la información; segundo, que la información documental es probabilística, y no determinista, y toda estrategia de gestión documental que ignore estos dos extremos corre el serio peligro de verse abocada al fracaso.

La técnica de fichero invertido, o de índice invertido, es un mecanismo orientado a palabras para la indexación de una colección de documentos de texto para agilizar las tareas de búsqueda. La estructura de un fichero invertido se compone de dos elementos: el vocabulario y las ocurrencias. El vocabulario es el conjunto formado por todas las ocurrencias de distintas palabras que aparecen en el texto. Asociado a cada palabra se encuentra una lista de los documentos donde se puede encontrar esa palabra. El conjunto de todas esas listas es lo que constituye las ocurrencias. Esta lista de ocurrencias está organizada en sí misma como una base de datos, cuyos registros tienen sólo dos campos: uno con el término y otro que indica en qué registros de la base de datos y en qué posiciones aparece ese término.

Un fichero invertido es un tipo de fichero indexado, donde la estructura de cada ítem o entrada del fichero es generalmente:

Término índice	Identificadores de documentos	Identificador de campo
----------------	-------------------------------	------------------------

- Un término índice es una palabra clave, frase o término con significado semántico que describe a un documento, al menos, de la colección de documentos. El fichero invertido contiene a estos términos índices ordenados de forma alfabética.
- El identificador de documento es único para cada documento. Cada término índice tiene asociado una lista de los identificadores de documentos que contienen ese término.
- El identificador de campo se utiliza para almacenar información útil dependiendo de los requerimientos del sistema de recuperación de información.

Por ejemplo, el identificador de campo puede indicar dentro de qué campo del documento aparece el término índice. Algunos sistemas incluyen también información acerca de la localización del término en el documento, por ejemplo, el párrafo o frase en que el término aparece dentro del documento. Si se utiliza el modelo conceptual de espacio vectorial, para cada identificador de documento asociado a un término índice puede incluirse su peso en el documento.

La utilización de fichero invertido aumenta la eficiencia en las búsquedas varios órdenes de magnitud, por lo que es una estructura básica para grandes volúmenes de información. Por otra parte, la principal desventaja es la cantidad de espacio necesario para almacenar la estructura del índice que puede variar entre un 10% y 100% del texto original, o incluso más. Además, es necesario tener en cuenta las penalizaciones para las inserciones de nuevos documentos y las actualizaciones en los documentos ya existentes, que requieren accesos al índice para su modificación. Normalmente, se imponen una serie de restricciones a la hora de construir los índices, que consecuentemente afectarán directamente a las búsquedas realizadas a posteriori.

La construcción de un fichero invertido comprende, a grandes rasgos, los siguientes pasos:

- Formar una lista de los términos que aparecen en el texto, junto con su localización en el mismo.

- Invertir la lista anterior, es decir, formar una lista de los términos ordenados lexicográficamente y por orden de aparición con las localizaciones asociadas a los términos.
- Opcionalmente, procesar posteriormente el fichero invertido, añadiendo pesos a los términos, reorganizándolos o comprimiéndolos.

Algunos ejemplos de restricciones son los siguientes:

- En algunos casos, puede ser útil utilizar un vocabulario controlado, que constituirá el vocabulario que será indexado. Aquellas palabras del texto no pertenecientes al vocabulario no serán indexadas, y por lo tanto, posteriormente no podrán ser buscadas.
- Normalmente se dispone de una lista de palabras comunes (artículos, preposiciones, etc.) que por razones de volumen de información no se incluyen en el índice, y por lo tanto no serán localizadas en la búsqueda.
- Existe un conjunto de reglas que se encargan de definir el inicio y el final de una palabra para ser indexada. Estas reglas se encargan del tratamiento de espacios en blanco, signos de puntuación, etc. y pueden tener un impacto significativo en los términos indexados.

Es importante destacar que estas restricciones encargadas de determinar lo que será indexado son críticas para la efectividad de las búsquedas, y por lo tanto son un parámetro clave a la hora de la construcción del índice.

2.3.3. Sistema de Generación y Gestión de Diálogo (SGGD).

Debido a que a menudo los usuarios son gestores sin formación informática, necesitan una “interfaz de usuario” intuitiva y sencilla que facilite su uso, ya que las incómodas o poco claras, no tendrán aplicabilidad.

Una interfaz debe soportar:

- La construcción y el análisis del modelo.
- Razonamiento acerca de la estructura del problema por la adición de cálculos numéricos.
- La elección y optimización de variables de decisión.

2.3.4. Integración del DSS.

Por si solos, los 3 componentes del DDS no tienen funcionalidad específica, por lo que se necesita integrarlos de algún modo. Los usuarios tendrán acceso al Sistema de Información mediante la interfaz de usuario o SGGD, a través de la cual manipularán los elementos necesarios de los componentes SGBD y SGBM (Figura 7).

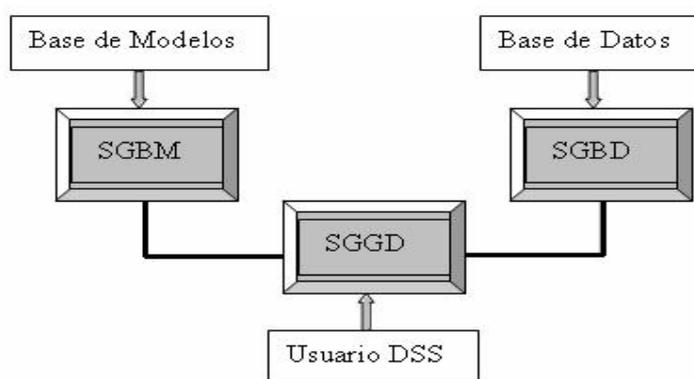


Figura 7: Integración del DSS.

CAECI es un cuestionario modificado que permite, además de adjuntar evidencias documentales de las calificaciones cualitativas, incluir la descripción evaluativa general de la temática mediante opiniones en forma de textos. Como parte del Sistema de Información que pretende desarrollarse, además de ser almacenada, también ésta podrá ser procesada por Minería de Textos. En este caso además de almacenar información, también podrá ser capaz de procesarla mediante el análisis de textos y finalmente recuperarla. Esto no es más, que un análisis semántico de la información, haciendo uso de una base de conocimiento (Ontología), la cual estará definida por un conjunto de toquen y reglas. Dicha recuperación se logra a través de los modelos de metadatos creando un repositorio con los mismos, pues los datos que se introducen en la interfaz, antes de llegar a la base de datos, se analizarán es dicho repositorio, para ser guardados en formato RDF/DC que muchas veces puede ser transparente para el usuario (Figura 8).

A los textos y contenidos que se deseen guardar se les tienen que añadir los metadatos y propiedades para cada información diferente. Dentro de éstos se incluirían los metadatos administrativos (relacionados con el creador del control, estado de aprobación, etc.) y los necesarios para facilitar su posterior recuperación (palabras clave, códigos de clasificación, etc.). Por lo tanto,

una aplicación de este tipo precisa de un repositorio de información y contenido con el que interactúan los distintos usuarios del sistema (autores, editores, programadores, diseñadores y consumidores y usuarios de la información) a través de las distintas herramientas e interfaces de usuario que ofrezca la aplicación.

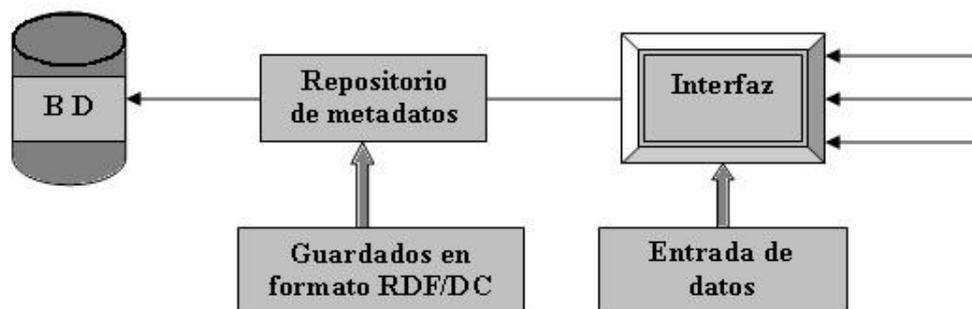


Figura 8. Guardado de Datos y Metadatos.

Este repositorio de metadatos puede adoptar distintas formas: una base de datos relacional, un sistema de ficheros, un repositorio de datos XML, o la combinación de ellos que sería la mejor propuesta. Al considerar el almacenamiento y la gestión del repositorio debe tenerse en cuenta diversos requerimientos, ya que todos los elementos deben estar perfectamente identificados, controlados y administrados desde un punto de acceso común y con una utilidad previa.

Como ya se había aclarado anteriormente la función primordial de XML es la de facilitar el intercambio y agregación de contenidos. XML no se trata, únicamente, de un formato para codificar textos y documentos, sino de una familia de especificaciones que establece la forma en la que se pueden procesar y presentar dichos textos. La posibilidad de obtener documentos XML y procesarlos con facilidad para cualquier fin (por ejemplo, para integrarlos en un repositorio o base de datos, o para visualizarlos como parte de un sistema como este), ofrece una flexibilidad extrema y abre las puertas a cualquier tipo de integración. Para representar esta información, se utilizarán diversas herramientas, que responden a la disciplina de la Inteligencia Artificial. Si se recupera en forma de texto se acudiría a la Minería de Datos, mientras que si se hace en forma de gráfico, al reconocimiento de patrones.

Las bases de datos se destinan para lograr contener todos los parámetros del sistema, almacenar los datos del dominio de aplicación y guardar las ontologías.

La base de modelos, formada por un conjunto de ontologías desarrolladas anteriormente debe facilitar primeramente la definición de ellas, su visualización, la incorporación de axiomas y por último, la generación de formularios de adquisición del conocimiento.

2.4. Conclusiones.

Se presenta la propuesta Audintec, la que permite desde la argumentación de los evaluadores obtener mayor información a partir de los datos procesados, donde en ésta los resultados del CAECI son más objetivos desde la seguridad y exactitud. La propuesta de CAECI intenta resolver el problema de la pérdida de datos importantes a tener en cuenta por el evaluador del Control Interno y que realmente tendrían tanto o más significación que los calificadores cualitativos.

CAPÍTULO 3: ESTUDIO DE FACTIBILIDAD DE LA PROPUESTA.

3.1. Introducción.

Se tienen en cuenta herramientas para probar algunos de los modelos de la propuesta, aunque las mismas no sean definitorias, proporcionan una representación del avance del trabajo y de la eficacia de la propuesta realizada. Se valora la posibilidad de una representación directa del resultado valorativo de CECI. Para el tratamiento de CAECI, se desarrolló una aplicación que compara los niveles de evaluación en una estructura geométrica abstracta de representación y además se hizo uso de una herramienta profesional llamada Tropes para el análisis semántico en los textos de los comentarios asociados a cada elemento.

Se simularon numerosos experimentos de representación y visualización, de los cuales se muestran solamente aquellos relacionados con la experiencia en el centro seleccionado para la validación.

3.2. Visualización de la Evaluación del cuestionario.

Una de las principales problemáticas existentes en Cuba para convencer de la utilidad del Control Interno, es la inexistencia de formas de visualización para poder caracterizar el estado general de la empresa en primer lugar, y posteriormente poder desarrollar comparaciones, tanto entre los diferentes componentes, en diferentes instantes de tiempo así como descripciones comparativas entre dos entidades en un mismo instante de tiempo.

En esta primera fase de la investigación, éste fue un aspecto relevante al cual se le dedicó tiempo y esfuerzo: buscar nuevas vías para representar los datos asociados al cuestionario CECI.

Se vio necesario obtener datos reales de los cuales se pudiera obtener información adicional, por lo tanto, se realizó un estudio y se aplicó el primer componente del informe COSO en el Centro de Neurociencias de Cuba, en Ciudad de La Habana. Las evaluaciones del mismo fueron dadas de acuerdo a la situación real de la entidad en cada aspecto, haciendo los comentarios correspondientes y respaldando muchas de las evaluaciones con documentos que las evidencian, estos elementos se resaltaron en letra *negrita* (Ver Anexo 8). No todos ellos pudieron ser evaluados,

estos se trataron como “No Proceden” (NP). Estos datos fueron también “entradas” para la creación de los patrones.

3.2.1. Herramienta “Gráfica evaluativa”.

Durante el desarrollo de esta propuesta del sistema se ha implementado una herramienta para ayudar a visualizar y comparar los resultados de las evaluaciones dadas por CECI y CAECI en un cubo referente al primero de los cinco componentes del cubo COSO, llamada “Gráfica evaluativa” (Ver Anexo 4). La información recogida durante el estudio realizado en el Instituto de Neurociencias se utiliza en la herramienta para visualizar los niveles de grises de las evaluaciones dadas y poder hacer comparaciones generales.

Esta nueva herramienta permite realizar las evaluaciones asignándoles un color: excelente (blanco), bien (gris claro), regular (gris oscuro), mal (negro) y no aplicado (rojo), con el fin de tener una visualización final que permita tener una idea general del estado de ese componente en la entidad, es decir, según el grado de oscuridad de los grises representará mayores dificultades en el mismo. Para un mejor procesamiento de los datos se llevaron estas evaluaciones cualitativas a cuantitativas, tomando como valor inicial el punto medio del rango asignado a cada evaluación como se explica en la temática 2.2 del capítulo anterior.

Una vez evaluados todos los elementos se pasa a modo CAECI, donde a cada una de las evaluaciones dadas anteriormente, se les podrá sumar o restar el valor δ dado “supuestamente” por el procesamiento de datos mediante la Lógica Borrosa, ya que esta técnica no es la base de este estudio no se llegó a realizar su implementación para obtener un valor real. Este valor de δ unido a la evaluación inicial de ese elemento, devuelve un número final que será llevado nuevamente a escala de grises para pintar el cubo.

Acabado de asignar la escala de colores a ambos cuestionarios, se puede proceder a comparar visualmente los dos cubos, el de CECI y el de CAECI. Aquella evaluación dada inicialmente a algún elemento pudo haber mejorado o bajado su nivel, lo que ahora permitirá detectar dónde pueden existir problemas en la entidad. (Figura 9)

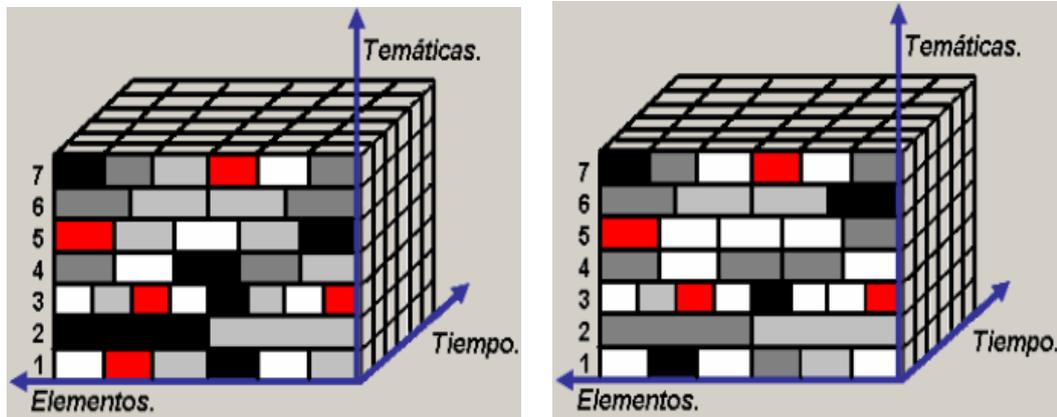


Figura 9. Comparación de evaluaciones entre CECI y CAECI.

Por ejemplo, en la temática 1 referida a “Integridad y valores éticos”, se da una evaluación de B al elemento 1.3, el indicador se posiciona en el centro del rango para esa evaluación a la vez que el voxel correspondiente en el cubo, se pinta de gris claro. Pasando a modo CAECI, se selecciona el número del elemento y se le suma el valor del delta dado por Lógica Borrosa, suponiendo que fuera +0,20, esto llevaría a mejorar la evaluación y a cambiar el color de gris claro a blanco, demostrando que en ese elemento no existe problema alguno.

Con la utilización de esta nueva herramienta se puede llegar a representar visualmente las evaluaciones dadas a cada elemento, para arribar a una conclusión general de la entidad mediante la escala de grises. El valor del delta no solo permite mejorar la evaluación dada, este valor ayuda a dar una confiabilidad mayor a dicha evaluación, con posibilidad de cambiar el nivel de la misma. El cubo final permitirá definir e identificar los elementos que afectan al mejor desempeño y rendimiento de la entidad, enfocándose en las tonalidades más oscuras de los grises.

Este estudio sería útil para una sola empresa evaluando cada uno de sus elementos, pero eso no prohíbe que se realice a nivel gubernamental cambiando los ejes guías y representando empresas en vez de temáticas o componentes en un prisma de trabajo. (Figuras 10 y 11)

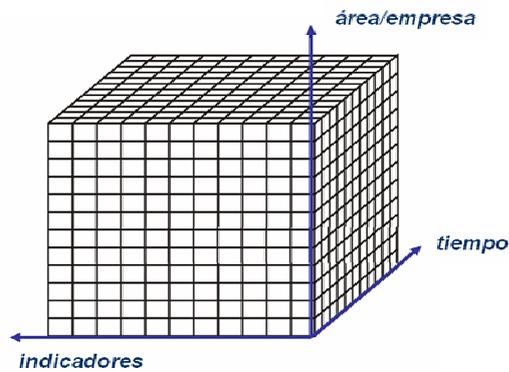


Figura 10. Prisma de trabajo a nivel gubernamental.

<p>11(a) Imagen asociada a 5 empresas simuladas en el instante $t = t_1$</p>	<p>11(b) Fila imagen asociada a la empresa 3 en el instante $t = t_1$</p>
<p>11(c) Serie temporal asociada a un indicador de la empresa 1 para diez instantes de análisis.</p>	<p>11(d) Fila imagen asociada a la empresa 1 en el instante $t = t_1$</p>

Figura 11. Representación asociada de indicadores para diferentes entidades laborales de un propio sector o de un territorio.

3.2.2. Patrón Pentágono.

Este patrón es una figura bidimensional en forma de pentágono tal y como se muestra en la Figura 12. En cada una de las aristas del pentágono se representan los resultados obtenidos en una componente y por cada ángulo se ubican los resultados de cada elemento. Si el elemento tiene una evaluación de Mal el radio alcanzará solamente al pentágono más pequeño, si es de Regular llegaría a uno más exterior, si es de Bien hasta el pentágono siguiente, si es de Excelente alcanzaría al pentágono exterior y si No Procede la evaluación, entonces se representaría como un punto en el centro del pentágono.

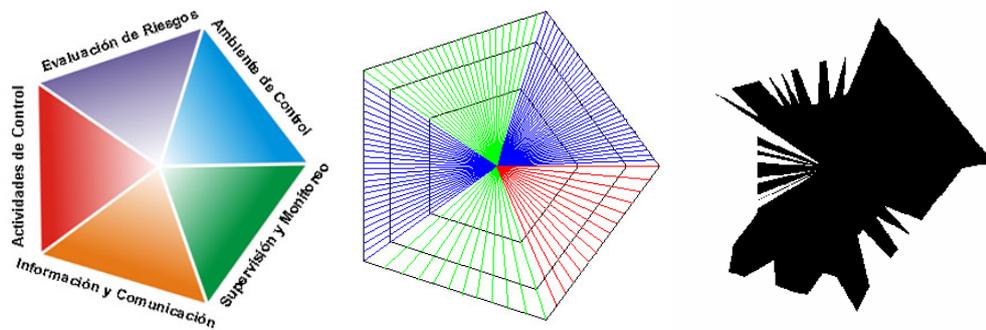


Figura 12. Pentágonos de visualización del Control Interno.

Una vez ubicados todos los puntos o evaluaciones de los elementos se unen por líneas rectas formando una línea quebrada. Para este caso se estudia el área relativa que comprende la línea quebrada formada, para ello se utiliza la siguiente fórmula:

$$A_r = 100 \left[\frac{A_r}{A_{máx}} \right]$$

El área relativa comprendida por la figura formada es igual a su área total sobre el área total (área del pentágono) multiplicada por 100 para llevarlo a un por ciento. El valor de Ar indica el estado integral de la empresa en todas sus componentes, mientras mayor sea este valor a 100, más cerca está la empresa del valor ideal.

Tomando también los datos obtenidos del cuestionario aplicado en Neurociencias se elaboraron dos patrones, uno de ellos representando las evaluaciones con evidencia o respaldo, similar a la nueva información que sería obtenida mediante Lógica Difusa. (Figura 12) El área relativa

de la Figura 13b es mayor que la de la Figura 13a, lo que representa un por ciento también mayor del valor ideal de la entidad.

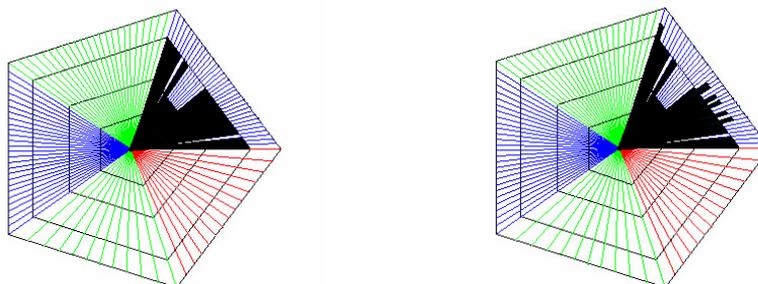


Figura 13. a) Evaluaciones sin respaldo. 12 b) Evaluaciones con respaldo.

Aunque aparentemente ambos son muy similares, es necesario entender que las evaluaciones respaldadas, ya fuera por comentarios o por evidencia adjunta, aportan nueva información que antes no se tenía en cuenta en la empresa para dar un nivel real de la evaluación; ahora gracias a técnicas de Inteligencia Artificial pueden ser procesadas y realizarse comparaciones y estudios para futuras decisiones en la entidad. El campo del Reconocimiento de Patrones asociados a lo estructural (*forma*) podría ser interesante para estudios posteriores en el procesamiento de estas imágenes.

3.3. Aplicación de la Herramienta Tropes

Tropes es una herramienta, construida por Microsoft Windows, para el análisis semántica de textos, el cual lleva implícito el análisis cualitativo y cuantitativo de datos textuales, análisis de encuestas (de preguntas abiertas), estudio de la competencia, ayuda en la redacción de resúmenes de textos, análisis del discurso político, análisis de las entrevistas. Ver Anexo 5.

En el texto se analizarán letras del alfabeto, caracteres de puntuación, etc. En caso de querer una secuencia de palabras se deberá unir usando “_”. Ejemplo (Republica_de_Cuba). Esta herramienta es capaz de leer ficheros en diferentes formatos como: HTML, XML, entre otros; también recupera páginas Web (Ficheros HTML).Éste cuenta con una base de conocimientos, donde están definido una serie de parámetros. La dificultad mayor en el uso de este software, es que hay que tener licencia.

Tropes dispone de múltiples herramientas de análisis, de entre las cuales cabe destacar que permiten clasificar automáticamente las palabras del texto; contraer el texto, ya sea de forma

automática o parametrizable; visualizar los fragmentos pertinentes, por temas o por categorías; filtrar temas en función de su relevancia; localizar rápidamente las series cronológicas en el interior de un documento y de comprender inmediatamente la estrategia del narrador y crear escenarios semánticos que permiten fabricar herramientas de interpretación y explorar las clasificaciones personalizadas.

3.3.1. Experimentos utilizando Tropes.

Después de haber estudiado la Guía Evaluativa del Control Interno en Cuba (Estándar Internacional COSO). Se ha seleccionado de esta guía una de sus cinco componentes: Ambiente de Control, de la cual se seleccionan también un conjunto de palabras claves (integridad, competencia, control, confianza, estructura_organizativa, valor_ético), pues se hará un análisis de esta información utilizando el Tropes.

Se introducirán los comentarios obtenidos en el estudio realizado en Neurociencias referidos también al primer componente del COSO; aquí se analizarán cuáles constituyen clases semánticas, y la relación que existe entre ellas y las preguntas del cuestionario.

Análisis #1: Palabra propia: Integridad (Se repite dos veces, pero no constituye una clase semántica a pesar que aparece como nombre principal en la temática.)

Análisis # 2: Palabra propia: competencia (Aparece 3 veces en el texto, una vez en mayúscula y dos veces en 2 minúscula. Constituye una clase semántica, por lo que aporta ideas). Ver figuras 14, 15 y 16



Figura 14. Gráfico de Esferas de la representación de relaciones semánticas de la herramienta Tropes.

En el gráfico de esferas se puede ver la relación entre las clases equivalentes, a la izquierda las precedentes (nivel, dirección) y a la derecha las clases sucesoras (profesional, conocimiento), siendo la clase profesional la de más relación con la palabra escogida anteriormente.

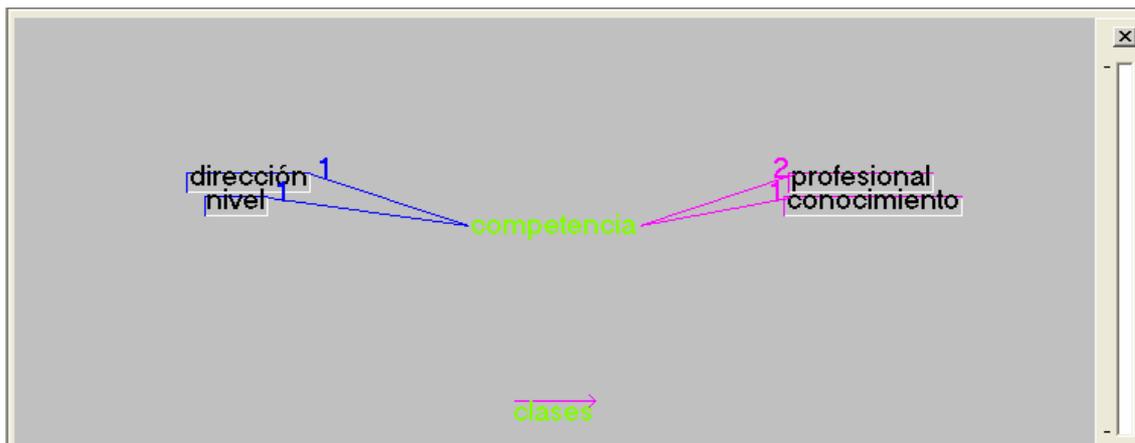


Figura 15. Gráfico de Estrellas de la representación estadística de relaciones semánticas de la herramienta Tropes.

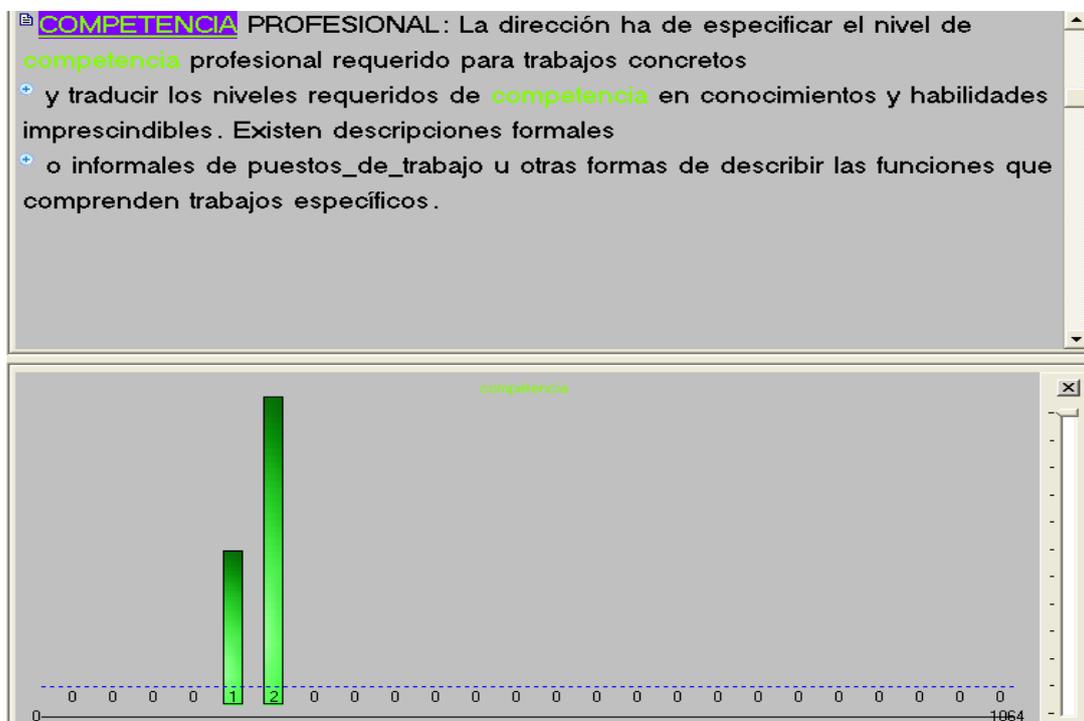


Figura 16. Gráfico de Barras de la representación estadística de relaciones semánticas de la herramienta Tropes.

En el gráfico estrella, se expresa la misma información, pero con la cantidad de veces que la palabra aparece en el texto, indica el número de relaciones (frecuencia de concurrencia) que hay entre las clases equivalentes.

Análisis # 3: Palabra propia: control (Aparece 8 veces en el texto, constituye una clase semántica, la cual se relaciona con 11 clases equivalentes.). Ver Anexo 6.

Análisis # 4: Palabra propia: confianza (No aparece en el texto por lo que no aporta ninguna información.)

Análisis # 5: Palabra propia: valores éticos.

Como se ha observado, cuando es una palabra compuesta o idea, agrupas todas las clases relacionada, con las ideas, en una misma, de ahí se hace otro análisis con las clases equivalentes, en este caso se analiza la palabra “norma”. Ver Anexo 7.

Finalmente se tiene un histograma que ubica las clases en el lugar que le corresponde en el texto, mostrando de dónde se obtuvo dicha información.

Con estos experimentos, en la serie de figuras se demuestra la conveniencia de incorporar al SI una herramienta similar, capaz de realizar comparaciones y estudios semánticos de las palabras asociadas a los comentarios y documentos adjuntos que respaldan las evidencias. De esta forma se podría crear la Ontología asociada.

3.4. Conclusiones.

En este capítulo mediante la utilización de diversas herramientas para la comparación de resultados, se arribó a que la propuesta del Sistema de Información para el Control Interno, es factible y brinda más información relacionada para el Control Interno de cualquier entidad que permitió resolver algunas dificultades presentadas en CECI. Además, los experimentos desarrollados en la entidad seleccionada, permitieron discernir futuras acciones que permitan crear la Ontología requerida para el desarrollo de un Sistema de Información.

CONCLUSIONES

- El estudio del estado del arte permitió la sistematización en los conocimientos necesarios sobre Inteligencia Artificial y Ciencias Empresariales así como la argumentación de los modelos a utilizar para la extracción de la información.
- Para la entrada, almacenamiento, procesamiento de Datos y Metadatos, así como la recuperación de información, se propone el diseño del Sistema de Información Audintec, que permitiría la obtención de información significativa y oportuna a los usuarios para la Toma de Decisiones asociadas al Control Interno.
- Mediante la utilización de las diferentes herramientas y la contrastación de resultados con criterio de expertos de un campo interdisciplinario del saber, se ha comprobado la factibilidad del diseño propuesto.

RECOMENDACIONES

- Continuar profundizando en el estudio del tema desarrollado para identificar, contrastar y validar otros modelos que permitan obtener indicadores y/o patrones significativos y útiles.
- Desarrollar el diseño y la implementación de esta propuesta.

BIBLIOGRAFÍA

- Argandoña, G. (2002) Fortalecimiento de la función de Auditoría Interna. Gerente Business Expansion & Marketing.
- Baptista, A. C. E. (2004). *Uso de RDF y bases de datos de metadatos nativas dentro del proyecto Omnipaper*. Do Porto, Facultad de Ingeniería de la Universidad de Do Porto; pp. 166-169.
- Blanco, H. (2000) El Factor Gerencial y el Perfeccionamiento de la Empresa Estatal Cubana. Economía y Desarrollo. No. 1 / Vol. 126 / Ene.-Jun. / 2000
- Blanco L. (2004) *Complejidad, caos y administración de empresas. Un acercamiento desde los sistemas de información y conocimiento*. Memorias en CD ROM del Evento: "45 Aniversario de los Estudios de Economía" La Habana 2004.
- Durand E. (2004). *El diseño de un modelo integrado de información sobre control interno*. Uniersia Business Review. Actualidad Económica. Cuarto Trimestre. pp .88-100.
- Escorza, P; Maspons, R; Ortiz, I. (2000), La integración entre la Gestión del Conocimiento y la Inteligencia Competitiva: la aportación de los mapas tecnológicos, Espacios, Vol. 21, Nº 2, Caracas.
- Espín R., Fernández E., Mazcorro G. Lecich M.I. (2007) *A fuzzy approach to cooperative n-person games*. European Journal of Operational Research, 176, p.1735-1751.
- Faúndez, U. A. (2003). *Análisis de información: características, metodologías, proyecciones*. Instituto de Ciencias Políticas. Chile, Universidad de Chile.
- Founier, I. D. (2006). *Descripción de los recursos de información en Internet: Formato Dublín Core*. ECIMED. Vancouver.
- Friedman J., Hastie T., Tibshirani R. (2001) *The elements of Statistical Learning*. Froedman, Hastie & Tibshirani Inc. New York.
- García A.; Contreras S.; Inda E.; Jorge R.; González O.E. (2006). *AUDINTEC: soluciones informáticas para la gestión empresarial*. Memorias en CD-ROM de la IV Semana Tecnológica – FORDES. Ministerio de Informática y Comunicaciones. Cuba. Noviembre 2006.
- García A.; Contreras S.; Inda E.; Jorge R.; Valdespino A.; González O.E. (2007). *AUDINTEC: soluciones informáticas para la gestión gubernamental*. (2007). Memorias en CD-ROM de la Convención INFORMÁTICA 2007. Cuba. III Simposio Internacional Gobierno en Línea. Cuba. Noviembre 2007.
- Grandío A. (1996) *"Empresa, mercado y necesidades: una síntesis en Ciencias Sociales"*. Tesis Doctoral (1996) – Capítulo 2. Publicada en el sitio <http://www3.uji.es/~agrandio/Tesis/Teo0.htm>. consultada Enero 2007.

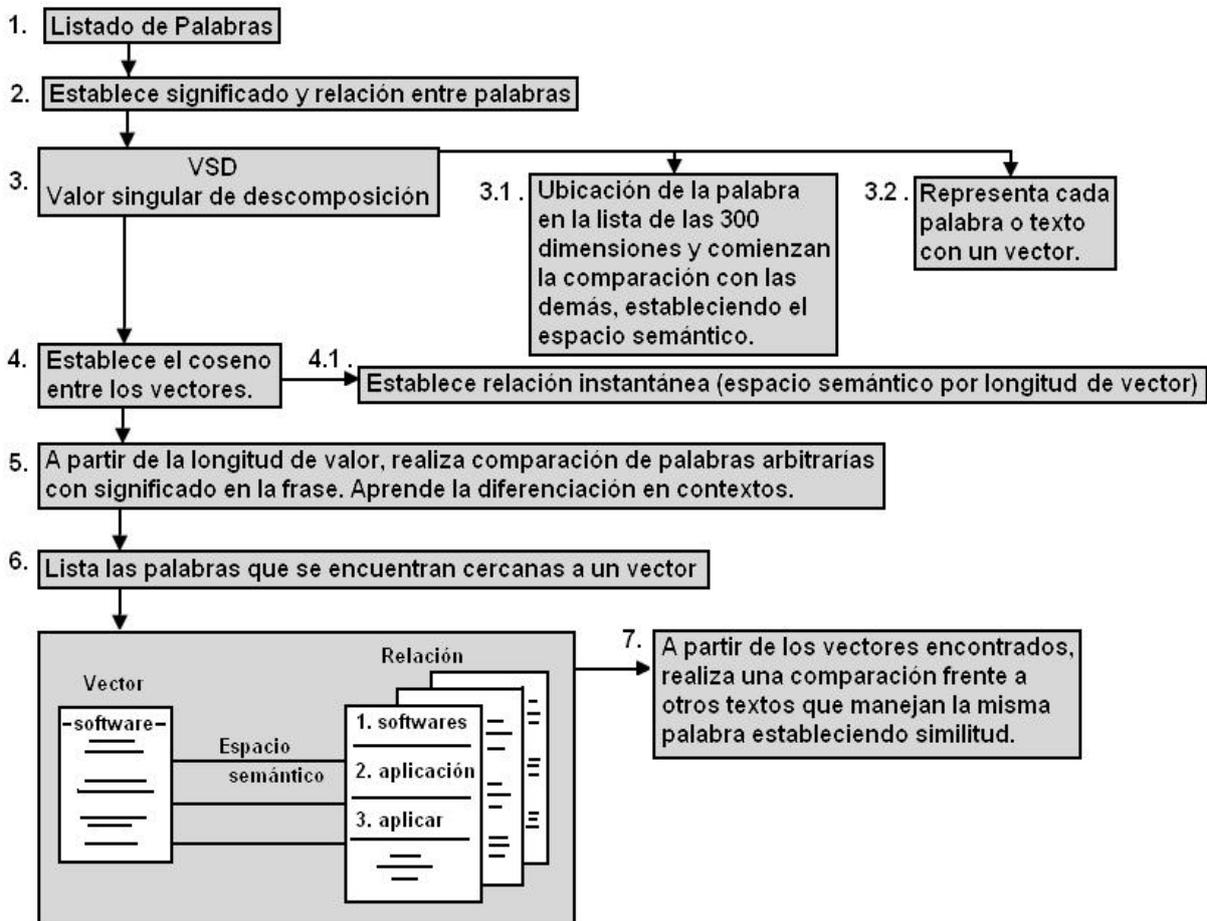
- Ibáñez, P. C. (2008). El informe COSO y sus repercusiones: un enfoque moderno del control interno. *Revista de Contabilidad y Tributación*: pp. 203-220.
- Lloyd Bierstaker, James; Thibodeau, Jay C. (2002). *An investigation of narratives versus questionnaires for identifying internal control weaknesses and strengths with implications for decision aid effectiveness*. Department of Accounting and Finance. College of Management. University of Massachusetts-Boston. USA.
- Martinez, R. G. (2002). *Algoritmos TDIDT aplicados a la minería de datos inteligentes*. Revista del Instituto Tecnológico de Buenos Aires. Buenos Aires, Universidad de Buenos Aires. Vol. 26: pp. 39 - 57.
- Mesarina M., Srinivasmurthy V.K., Lyons N., and Sayers C. (2007). *A Management and Performance Framework for Semantic Web Servers*. WWW 2007, May 8–12, 2007, Banff, Alberta, Canada. ACM 978-1-59593-654-7/07/0005. Canada.
- Michard Alain, Rizk, Antoine. (2005). *Semantic Interoperability Strategy: The European XML Clearinghouse*. Feasibility Study. December 2005. IDABC.
- Peza, M. B. d. I. (2007). *El Control Interno como Herramienta en la Toma de Decisiones y Protección de las Empresas*. H. C. Miranda. Mexico.
- Pinto, F. J. (2007). *Evaluación del Sistema de recuperación de información Lemur Utilizando Distintos Tipos de Indexación Automática*. Departamento de Ciencias de la Computación. España, Universidad de la Coruña.
- Toledo J.C. (2003). *“El control interno de la administración: modernización de los sistemas como una emergencia”*. Memorias CD-ROM del VIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Panamá, 28-31 Octubre. 2003.
- Treadway J.C. (1987). *Report of the Nacional Comisión on Fraudulent Financial Reporting (COSO Report)*. October 1987. [<http://www.coso.org/Publications/NCFFR.pdf>] -enero 2007
- Vallez M. y Pedraza-Jimenez, R. (2007). *El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines* [en línea]. "Hipertext.net", núm. 5, 2007. <<http://www.hipertext.net>> [Consulta: 12 may. 2007].
- Winston, P. H. (1994). *Inteligencia Artificial*. 3era. Edición. Adison Wesley Publishing - Iberoamericana.
- _____. COBIT 4.0 (2005). *Objetivos de Control, Directrices Gerenciales, Modelos de Madurez*. IT Governance Institute. Sitio Web: www.itgi.org. ISBN 1-933284-37-4.
- _____. (2001) *Internal Control. Comptroller's Handbook*. January 2001. Comptroller of the Currency Administrator of National Banks. Management.
- _____. PricewaterhouseCoopers LLP. (2006). *Private companies: are your internal controls supporting your business strategy?* BS-BS-06-0570-A. US.
- _____. AICPA. (2007). *Internal Control Issues in Derivatives Usage: An Information Tool*. COSO publications. Durham. 2.

ANEXOS

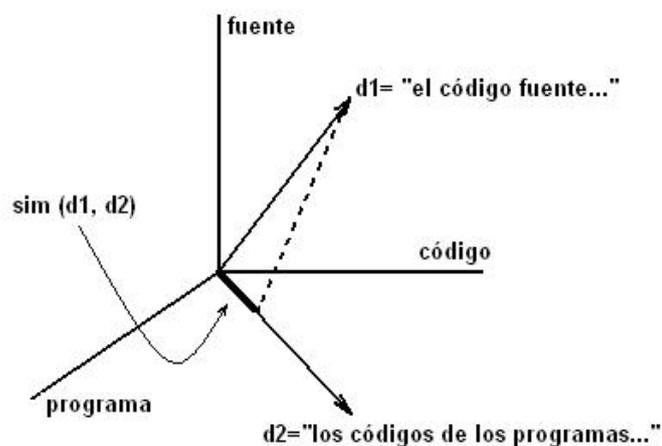
Anexo 1: Los 15 elementos del formato Dublin Core.

- Título: nombre por el que es conocido formalmente el recurso
- Creador: persona, organización o servicio que tiene responsabilidad principal sobre el contenido
- Materia: tema del contenido del recurso expresado de forma controlada
- Descripción: explicación del contenido en texto libre
- Editor: persona, organización o servicio que hacen disponible el recurso
- Colaborador: persona, organización o servicio que hace contribuciones al contenido del recurso
- Fecha: fecha asociada a la creación o disponibilidad del recurso. La práctica recomienda seguir el formato marcado por la ISO AAAA-MM-DD
- Tipo: naturaleza o género del contenido del recurso. Se recomienda tomar el valor para este elemento de un vocabulario controlado
- Formato: manifestación física o digital del recurso. Por ejemplo “tipo de soporte”. También puede utilizarse para determinar el hardware, software o equipamiento necesario para su visualización
- Identificador: referencia unívoca al recurso. Se recomienda usar algún sistema de identificación formal, como el URI (Uniform Resource Identifier), la URL (Uniform Resource Locator), el DOI (Digital Object Identifier) o el ISBN.
- Fuente: Secuencia de caracteres utilizado para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.
- Relación: referencia a un recurso relacionado. Para estos dos elementos también se recomienda el uso de un sistema de identificación formal para referirse a los recursos.
- Lengua: idioma del contenido intelectual del recurso de forma codificada
- Cobertura: extensión o alcance del contenido, tanto espacial como temporal. Se recomienda el uso de un sistema de identificación formal.
- Derechos de autor: información sobre los derechos de los recursos.

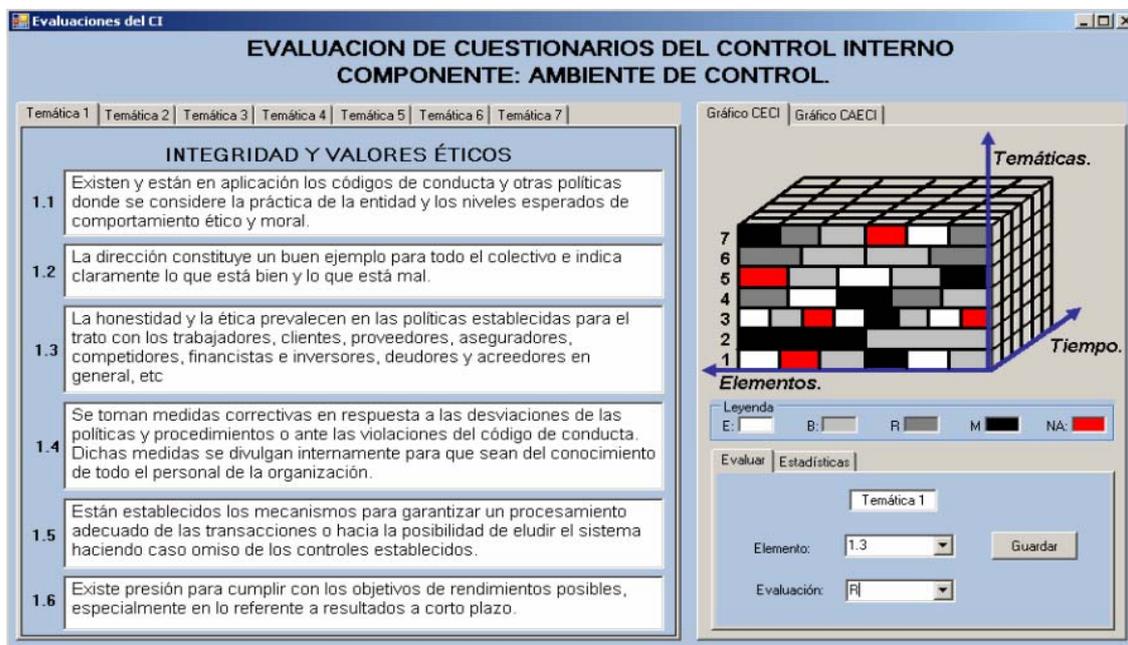
Anexo 2: Proceso de indexación semántica.



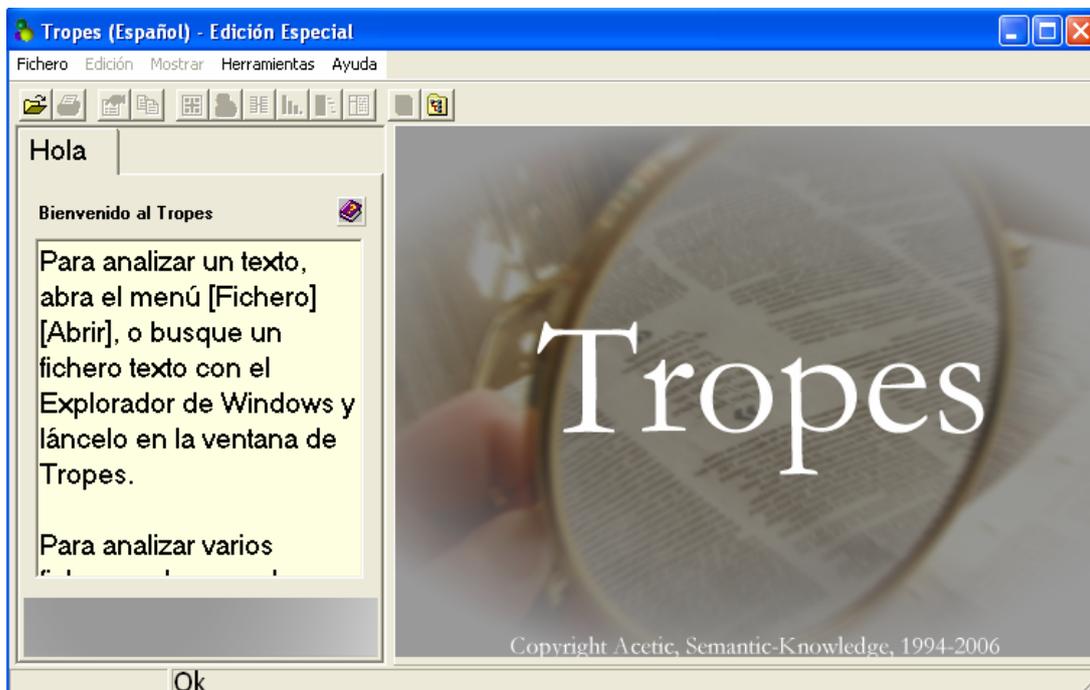
Anexo 3: Representación gráfica de la similitud de documentos.



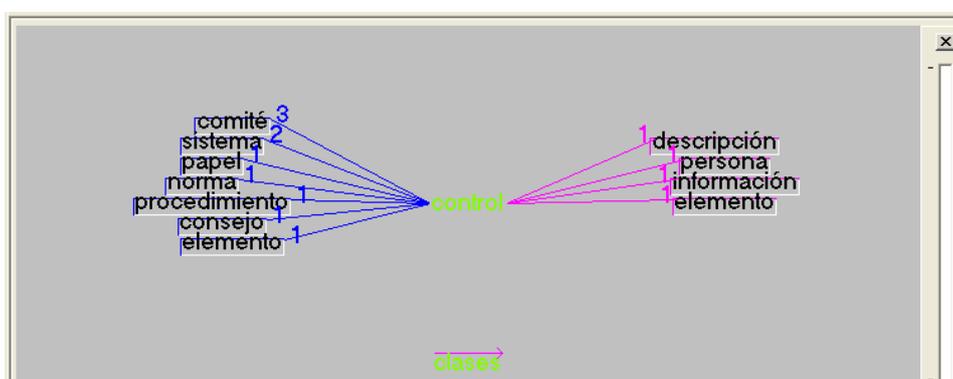
Anexo 4: Interfaz de la herramienta “Gráfica Evaluativa”.



Anexo 5: Versión 7.0.1 de la herramienta Tropes.

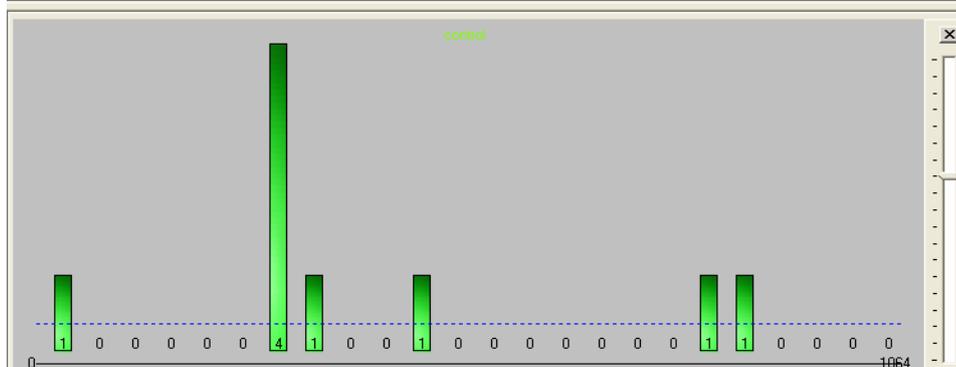


Anexo 6: Graficas del estudio semántico de la palabra “control”.

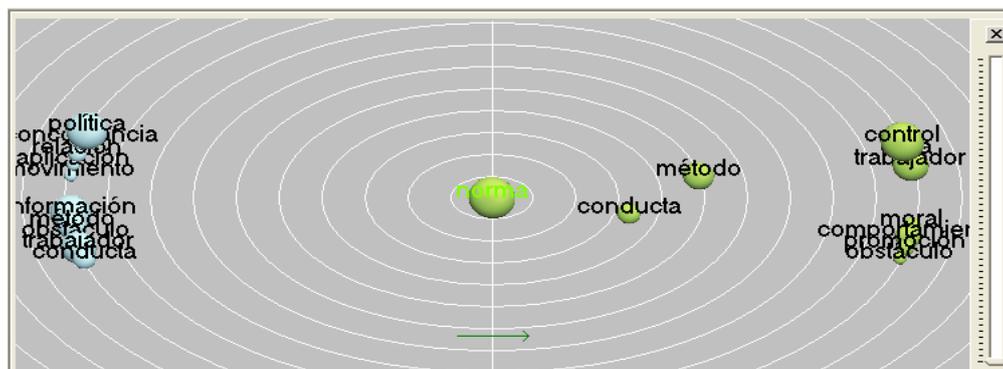


AMBIENTE DE CONTROL. INTEGRIDAD Y VALORES ÉTICOS: La dirección ha de transmitir el mensaje de

- **COMITÉ DE CONTROL:** Un Comité de Control activo y efectivo es un elemento de control importante.
- Dado que la dirección tiene la posibilidad de eludir los sistemas de control, el comité tiene un papel importante para asegurar un CONTROL INTERNO efectivo.
- El Comité examina constructivamente las decisiones tomadas por la dirección (por ejemplo, iniciativas estratégicas, transacciones importantes),
- y busca explicaciones para resultados pasados (por ejemplo, variaciones del

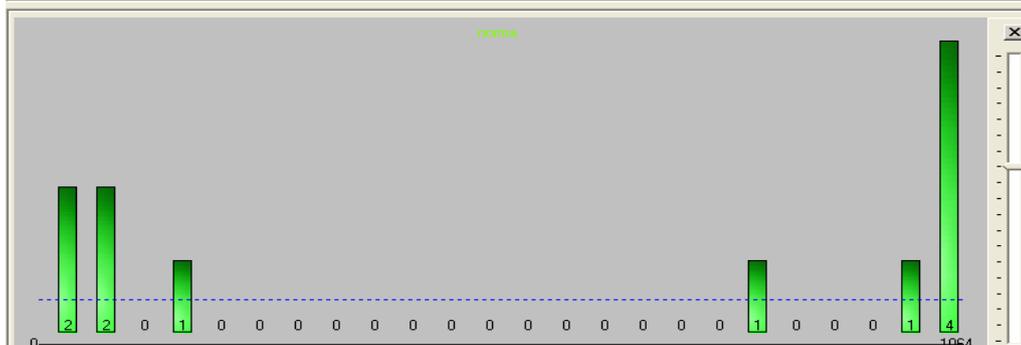


Anexo 7: Graficas del estudio semántico de la palabra “norma”.



INTEGRIDAD Y VALORES ÉTICOS: La dirección ha de transmitir el mensaje de que la integridad y los valores éticos no pueden,

- bajo ninguna circunstancia, ser contravenidos. Los trabajadores deben captar y entender este mensaje.
- La dirección ha de demostrar continuamente, mediante mensaje y su comportamiento, un compromiso con los valores éticos.
- Existen y están en aplicación los códigos de conducta y otras políticas donde se considere la práctica de la entidad
- Se toman medidas correctivas en respuesta a las desviaciones de las políticas y



Anexo 8:
Evaluación realizada en el Centro de Neurociencias de Cuba.
Fecha: 09 de enero 2008. - Evaluador: XXXXX
COMPONENTE: AMBIENTE DE CONTROL

Código		E	B	R	M	N/A
1. INTEGRIDAD Y VALORES ÉTICOS						
La dirección ha de transmitir el mensaje de que la integridad y los valores éticos no pueden, bajo ninguna circunstancia, ser contravenidos. Los trabajadores deben captar y entender este mensaje. La dirección ha de demostrar continuamente, mediante mensaje y su comportamiento, un compromiso con los valores éticos.						
1.1	Existen y están en aplicación los códigos de conducta y otras políticas donde se considere la práctica de la entidad y los niveles esperados de comportamiento ético y moral. (Si existe, por favor, adjuntar evidencia de estos documentos).		B			
	Se aplica el Plan de Trabajo Mensual (y Trimestral) y evalúa mensualmente el desempeño de cada trabajador y se analiza con cada uno de ellos previamente. En los casos Muy Bien y Excelente se defienden dichas evaluaciones ante el Consejo de Dirección del centro.					
1.2	La dirección constituye un buen ejemplo para todo el colectivo e indica claramente lo que está bien y lo que está mal.		B			
	Se evidencia un clima de trabajo satisfactorio y de pertenencia. Todos los trabajadores de la entidad tienen cargos en las organizaciones políticas, de masas y la ANEC.					
1.3	La honestidad y la ética prevalecen en las políticas establecidas para el trato con los trabajadores, clientes, proveedores, aseguradores, competidores, financistas e inversores, deudores y acreedores en general, etc.		B			
1.4	Se toman medidas correctivas en respuesta a las desviaciones de las políticas y procedimientos o ante las violaciones del código de conducta. Dichas medidas se divulgan internamente para que sean del conocimiento de todo el personal de la organización.		B			
	No ha existido problema pero está previsto si así lo fuera.					
1.5	Están establecidos los mecanismos para garantizar un procesamiento adecuado de las transacciones o hacia la posibilidad de eludir el sistema haciendo caso omiso de los controles establecidos.		B			
	Se utiliza el Sistema Integral Informático ACCES y sistemáticamente aplican el Sistema de Auditoría de los diferentes subsistemas.					
1.6	Existe presión para cumplir con los objetivos de rendimientos posibles, especialmente en lo referente a resultados a corto plazo.		N P			
	Se trabaja al día no existiendo atrasos.					

2. COMPETENCIA PROFESIONAL						
La dirección ha de especificar el nivel de competencia profesional requerido para trabajos concretos y traducir los niveles requeridos de competencia en conocimientos y habilidades imprescindibles.						
2.1	Existen descripciones formales o informales de puestos de trabajo u otras formas de describir las funciones que comprenden trabajos específicos.		B			
	Se comprueba que no existe interrelación entre funciones.					
2.2	Se analizan los conocimientos y las habilidades requeridas para realizar los trabajos adecuadamente.					
	Existe un dominio de las actividades que realiza cada trabajador del área está previsto un Plan de Capacitación en temas relacionados con sus actividades.		B			

3. COMITÉ DE CONTROL

Un comité de control activo y efectivo es un elemento de control importante. Dado que la dirección tiene la posibilidad de eludir los sistemas de control, el Comité tiene un papel importante para asegurar un Control Interno efectivo.

3.1	El Comité examina constructivamente las decisiones tomadas por la dirección (por ejemplo, iniciativas estratégicas, transacciones importantes), y busca explicaciones para resultados pasados (por ejemplo, variaciones del presupuesto).				M	
3.2	Se establecen Comités dependientes del Consejo en casos justificados por la necesidad de prestar atención más detalladas o directas a asuntos específicos.				M	
3.3	Los integrantes tienen los conocimientos suficientes, experiencia en el negocio de la entidad y tiempo para realizar sus funciones adecuadamente.				M	
3.4	Se celebran reuniones con los responsables financieros y/o contables, auditores internos y externos con la necesaria frecuencia y oportunidad.				M	
3.5	Se suministra información a los miembros del Consejo ó el Comité de forma suficiente y oportuna para permitir la supervisión de los objetivos y las estrategias de la dirección, la situación financiera y los resultados operativos de la entidad y las condiciones de los acuerdos significativos.				M	
3.6	Se evalúa suficiente y oportunamente por parte del Consejo o del Comité de Control la información más sensible, las investigaciones y los hechos significativos e inusuales,				M	
3.7	Está establecida la forma de dejar evidencia de que los superiores deben dar el ejemplo.				M	
3.8	El Consejo supervisa y realiza el seguimiento necesario de las acciones que deberían ser tomadas como consecuencia de sus conclusiones, incluyendo en su caso investigaciones especiales.				M	
M : Está por implementarse el Sistema de Control. Está designada la trabajadora que atenderá este Comité, pero aún no ha comenzado a funcionar dicho Comité de Control.						

4. ATMÓSFERA DE CONFIANZA MUTUA					
La filosofía de la dirección y el estilo de gestión suelen tener un efecto omnipresente en una entidad. Éstas son, naturalmente intangibles, pero se pueden buscar indicadores positivos ó negativos.					
4.1	La dirección actúa con cuidado y solamente después de analizar los riesgos y los beneficios potenciales de una operación que conlleva riesgos.		B		
4.2	Existe rotación de la plantilla en las funciones claves, por ejemplo, operativas, contabilidad, proceso de datos, auditoria interna.		B		
4.3	Existe preocupación de la dirección hacia el proceso de datos y las funciones de contabilidad, así como por la fiabilidad de la presentación de la información financiera y la salvaguarda de los activos.		B		
4.4	Se celebran con frecuencia reuniones de la dirección con las divisiones o filiales.		B		
4.5	Hay frecuente interacción entre la administración principal y la operativa, particularmente cuando operan desde localizaciones geográficamente apartadas.		B		
B: Existe una relación de confianza mutua donde los métodos de trabajo tienen resultados dentro del área económica.					

5. ESTRUCTURA ORGANIZATIVA					
La estructura organizativa no deberá ser tan sencilla que no pueda controlar adecuadamente las actividades de la entidad ni tan complicada que inhiba el flujo necesario de información. Los ejecutivos deben comprender la experiencia y los niveles de conocimientos requeridos en función de sus cargos.					
5.1	La estructura organizativa es apropiada y tiene capacidad de suministrar el flujo necesario de información para gestionar sus actividades.		B		
5.2	Están diseñadas las responsabilidades de los principales directivos y de las expectativas respecto a las actividades de sus áreas de responsabilidad y ambas son del conocimiento de éstos.		B		
5.3	Tienen los directivos los conocimientos y la experiencia adecuados para cubrir las responsabilidades asignadas.		B		
5.4	Se modifica la estructura organizativa cuando existen cambios en las circunstancias que lo requieran.		B		
5.5	Disponen de suficientes trabajadores con capacidad de dirección y supervisión.		B		
B: La estructura organizativa en el área contable es apropiada y responde a los flujos necesarios de información de la entidad.					

6. ASIGNACIÓN DE AUTORIDAD Y RESPONSABILIDAD					
La asignación de responsabilidad, la delegación de autoridad y el establecimiento de políticas conexas ofrecen una base para el seguimiento de las actividades y el sistema de control y establecen los papeles respectivos de las personas dentro del sistema.					
6.1	Se asigna responsabilidad y se delega autoridad para tratar objetivos organizativos, funciones operativas y requisitos de regulación, incluyendo la responsabilidad sobre los sistemas de información y autorizaciones para efectuar cambios.		B		
6.2	Las normas y los procedimientos relacionados con el control, incluyen la descripción de funciones, son los idóneos para los directores de departamentos y los supervisores.		B		
6.3	La entidad dispone del personal necesario en número, conocimientos y experiencia, para llevar a cabo su misión.		B		
6.4	Existe correspondencia en la delegación de autoridad en relación con las responsabilidades asignadas.		B		

7. POLÍTICAS Y PRÁCTICAS EN PERSONAL.					
Las políticas de recursos humanos son esenciales para la selección y la permanencia de los trabajadores competentes que permitan llevar a cabo los planes de la entidad y así lograr la consecución de sus objetivos.					
7.1	Se han implementado las políticas y procedimientos para la contratación, formación y promoción de los trabajadores.		B		
7.2	Se informa a los trabajadores de sus responsabilidades y lo que se espera de ellos.		B		
7.3	Son suficientes y adecuadas las acciones correctivas que se toman en respuesta a las desviaciones en las políticas y procedimientos aprobados.		N P		
7.4	Las políticas de personal están orientadas a la observancia de las normas éticas y morales de la entidad.		B		
7.5	Las verificaciones de los historiales de los candidatos a formar parte del personal de la entidad, en cuanto a comportamiento previo o actividades que se consideren no aceptables por la entidad, son suficientes y apropiados.		B		
7.6	Son idóneos los criterios de retención de los trabajadores, los criterios de promoción y de las técnicas para la recopilación de la información en relación con el código de conducta u otras directrices de comportamiento.		B		
	R:Las políticas de esta organización corresponden al Departamento de Recursos Humanos por lo tanto el resultado de este Departamento es con referencia solamente al pago del salario el cual se ejecuta de acuerdo a lo establecido.				

GLOSARIO

Acceso físico: En el control de acceso, tener acceso hacia áreas físicas o entidades.

Acceso lógico: El acto de ganar acceso hacia la información de la computadora. El acceso puede estar limitado a “sólo leer”, pero derechos de acceso más extensivos incluyen la habilidad de arreglar los datos, crear nuevos archivos, y borrar archivos existentes. (ver acceso físico).

Actividad de control: Las actividades de control son políticas y procedimientos establecidos para enfrentar los riesgos y lograr los objetivos de la entidad. Los procedimientos que una organización ejecuta para tratar el riesgo se llaman actividades de control interno. Las actividades de control interno son una respuesta al riesgo en tanto que son diseñadas para contener la parte poco certera del resultado que ha sido identificado.

Auditoría: Revisión de las actividades de una organización y de las operaciones para asegurar que éstas están siendo ejecutadas o están funcionando de acuerdo con los objetivos, el presupuesto, las reglas y normas. El objetivo de esta revisión es identificar, en intervalos regulares, desviaciones que pudieran necesitar una acción correctiva.

Auditoría Interna:

- Los medios funcionales a través de los cuales los gerentes de una entidad reciben de fuentes internas la seguridad de que todos los procesos contables están operando de manera en la que puedan minimizar la probabilidad de la ocurrencia de un fraude, error o prácticas ineficientes o antieconómicas. Tiene muchas de las características de la auditoría externa, pero puede llevar a cabo las directivas provenientes de la gerencia a la que porta. (Normas de auditoría INTOSAI).
- Una actividad independiente y objetiva que da a una organización una seguridad sobre el grado de dominio de sus operaciones, que da sus consejos para mejorarlas y que contribuye a crear valor anadido. Ayuda esta organización a alcanzar sus objetivos, evaluando, con un enfoque sistemático y metódico, sus procedimientos de gestión de riesgos, de control y de dirección de empresa, y haciendo propuestas para reforzar su eficacia. (IIA, IFACI).
- La auditoría interna es una actividad de avalúo establecida dentro de una entidad como servicio para la misma. Sus funciones incluyen, entre otras cosas, examinar, evaluar y monitorear cuan adecuada y efectiva es la contabilidad de los sistemas de control interno (IFAC).
- La auditoría interna es la que se realiza por el auditor interno bajo la supervisión administrativa de la propia empresa; se realiza atendiendo todo lo que se revisará por la externa para uso propio de la empresa y es preventiva. Su objetivo es evaluar y emitir opinión a la Administración interna a la empresa con el objetivo de prevenir.

Auditoría Externa: Una auditoría llevada a cabo por un cuerpo que es externo e independiente del auditado, siendo el propósito dar una opinión o un informe sobre las cuentas de los estados financieros, la regularidad y legalidad de las operaciones, y/o el manejo financiero. Es la que revisa todo el accionar de la empresa en lo financiero, en la gestión de sus recursos y en sus sistemas de control; evalúa y juzga, tiene facultad para dar Fe Pública. La efectúa un auditor calificado y certificado, independiente o ajeno a la institución. Su objetivo es juzgar y emitir opinión profesional sobre la entidad a terceros.

Audidores Internos: Examinan la efectividad del sistema de control interno y recomiendan mejoras, pero no tienen responsabilidad primaria por implantarlo o mantenerlo.

Ciclo de valoración del riesgo: Es un proceso continuo e interactivo para identificar y analizar cambios en las condiciones, oportunidades y riesgos y realizar las acciones necesarias al respecto,

Componente de Control Interno: Uno de los cinco elementos del control interno. Los componentes del control interno de una entidad son: entorno de control, evaluación de riesgo, actividades de control, información y comunicación, y seguimiento. (COSO 1992).

Control: Cualquier acción tomada por la gerencia, el consejo y otras partes para gestionar el riesgo y aumentar las posibilidades de que los objetivos y metas se alcancen. La gerencia planifica, organiza y dirige la gestión con suficientes acciones para proporcionar una garantía razonable de que los objetivos y metas se alcanzan (IIA).

- Un sustantivo, utilizado como sujeto, ejemplo: existencia de un control, una política o procedimiento que sea parte del control interno. Un control puede existir dentro de estos cinco componentes.
- Un sustantivo, utilizado como objeto, ejemplo efectuar control - el resultado de políticas y procedimientos diseñados para controlar; este resultado puede o puede no ser un control interno efectivo.
- Un verbo, ejemplo: controlar, regular, establecer o implantar una política que afecte el control (COSO 1992).

Control Interno. El control interno es un proceso integrado que afecta a la gerencia y al personal de la entidad y está diseñado para dar seguridad razonable de que en la búsqueda de su misión, los siguientes objetivos generales serán conseguidos:

- Ejecutar las operaciones de forma ordenada, ética, económica, eficiente y efectiva, cumpliendo con las obligaciones de contabilidad y todas las leyes aplicables, así como las regulaciones y la salvaguarda de los recursos contra la pérdida.
- Una garantía independiente y objetiva y una actividad consultora diseñada para proporcionar valor añadido y mejorar la operativa de la organización. Ayuda a la organización a cumplir sus objetivos mediante el uso de un enfoque sistemático y disciplinado para evaluar y mejorar la eficacia en los procesos de gestión del riesgo, control y administración (IIA).

COSO: Comité de organizaciones patrocinadas, un grupo de varias organizaciones de contabilidad. En 1992 publicó un estudio significativo sobre el control interno llamado Control Interno- un marco integrado. El informe - que coincide con el llenado de CAECI - es muchas veces llamado Informe COSO.

Documentación: La documentación de la estructura de control interno es la prueba material y escrita de los componentes del proceso de control interno que incluye la identificación de la estructura de una organización, las políticas y las categorías operacionales, los objetivos relacionados y las actividades de control.

Efectivo: Se refiere al cumplimiento de los objetivos o al grado al que los resultados de una actividad cumplen con los objetivos o con los efectos para los que se realizó una actividad.

Efectividad: Es el grado al que los objetivos son logrados, y la relación entre el impacto deseado y el impacto verdadero que recibe la entidad. (Normas de auditoría INTOSAI). El grado en el que los objetivos establecidos han sido logrados bajo la óptica de costo-efectividad. (glosario).

Eficiente: Se refiere a los recursos utilizados para lograr los objetivos. Significa que el mínimo de recursos como entradas deben ser utilizados para obtener una determinada cantidad y calidad de salidas, o el máximo producto con una cantidad y calidad determinadas utilizadas como entradas.

Eficiencia: La relación entre la salida, en términos de bienes, servicios y otros resultados y los recursos utilizados para producirlos. (Normas de auditoría INTOSAI). Utilización de los recursos financieros, humanos y materiales de manera que se puedan maximizar las salidas por un determinado número de recursos, o minimizar los recursos invertidos para determinada cantidad y calidad de salidas.

Entorno de control: El entorno de control establece el tono de una organización, influyendo en la conciencia de control de su personal. Es el fundamento para todos los componentes del control interno, el que da disciplina y estructura.

Evaluación de riesgo: La evaluación de riesgo es el proceso de identificación y análisis de los riesgos relevantes al alcanzar los objetivos de la entidad y determinar una respuesta apropiada.

Entropía: Medida de la incertidumbre existente ante un conjunto de mensajes, de los cuales se va a recibir uno solo. Tendencia natural de la pérdida del orden.

Fraude: Interacción fuera de la ley entre dos entidades, donde una de las partes intencionalmente defrauda a la otra a través de representaciones falsas para ganar ventajas ilícitas o injustas. Involucra actos de engaño, conciliación falsa, utilizados para ganar alguna ventaja injusta o deshonestas.

Hiponimia: Relación donde una palabra posee todos los rasgos semánticos de otra más general pero que añade en su definición otros rasgos semánticos que la diferencian de la segunda.

Homonimia: Relación semántica que se establece entre dos palabras cuando estas presentan identidad formal (fónica o gráfica) pero diferencia en el significado.

Incertidumbre: La falta de habilidad para saber de antemano la exacta probabilidad o impacto de eventos futuros. (COSO ERM).

Indexación: Adición autónoma de datos al archivador como términos de índice.

Instituto de Auditores Internos (IAI). Esta es una organización que establece normas éticas y de práctica, provee capacitación y fortalece el profesionalismo entre sus miembros.

Entidad Fiscalizadora Superior (EFS). La organización pública de un Estado que, sin importar su diseño, constitución u organización, ejerce por ley, la más alta función pública de auditoría de ese Estado. (Normas de auditoría INTOSAI & IFAC).

Integridad: La calidad o el estado de un importante principio moral; rectitud, honestidad y sinceridad; el deseo de hacer las cosas correctamente, profesar y vivir de acuerdo con ciertos valores y expectativas. (COSO 1992).

Lematización: Es la operación de hallar el lema del cual se deriva una palabra concreta. En este contexto, su finalidad es agrupar palabras que tienen un contenido semántico muy próximo; por ejemplo: biblioteca, bibliotecas, o bibliotecario.

Meronomia: Relación semántica no-simétrica cuando el significado de una palabra constituye una parte del significado total de otra palabra.

Objetividad: Es una actitud mental que permite que los auditores externos e internos de las EFS realcen su actividad de tal manera que crean de honesta en su trabajo y que tengan compromisos con otros en detrimento de la calidad del mismo.

Operaciones: La palabra usada con "objetivos" o "controles" tiene que ver con la efectividad o la eficiencia de las actividades de una entidad, incluyendo la actuación y beneficio de los objetivos, y salvo guardando los recursos. (COSO 1992). Las funciones, procesos, y actividades con los que los objetivos de una entidad son alcanzados.

Organización Internacional de Instituciones fiscalizadoras superiores (INTOSAI). INTOSAI es la organización profesional de las instituciones fiscalizadoras superiores (EFS's) en los países que pertenecen a las Naciones Unidas o a sus agencias especializadas. Las EFS's juegan un rol central en la auditoría de cuentas gubernamentales y operaciones, y en la promoción de gerencias financieras sólidas y responsabilidad en los gobiernos.

Polisemia: Capacidad que tiene una sola palabra para expresar distintos significados.

Sinonimia: Relación de semejanza de significados entre determinadas palabras (llamadas sinónimos) u oraciones.

Riesgo: La posibilidad de que ocurra un evento adverso que afecte el logro de los objetivos. (COSO ERM).

Riesgo inherente: El riesgo que tiene una entidad de que en la ausencia de acciones gerenciales pueda mitigar la probabilidad del riesgo o su impacto. (COSO ERM).

Riesgo aceptable: La cantidad de riesgo a la que una entidad está preparada para exponerse antes de que una acción se juzgue necesaria. Una base amplia de riesgo que una compañía o entidad está dispuesta a aceptar en la búsqueda de su misión o su visión. (COSO. ERM).

Riesgo residual: El riesgo que permanece después de que la gerencia dé respuesta al riesgo.

Seguridad razonable: Es igual a un nivel satisfactorio de confianza bajo consideraciones dadas de costos, beneficios y riesgos. El concepto de que el control interno, sin importar su buen diseño y operación, no puede garantizar que los objetivos de la entidad sean alcanzados. Esto se debe a limitaciones inherentes de todos los sistemas de control interno. (COSO 1992).

Sistema de control interno (o proceso, o arquitectura): Un sinónimo de control interno, aplicado en una entidad. (COSO 1992).

Sistema de Información: Sistema de herramientas para crea, almacenar, intercambiar, buscar, recuperar o manipular información.

Tolerancia al riesgo: Es la variación relativa aceptable en la consecución de los objetivos. (COSO: ERM).

Tesoro: Es el puente entre el lenguaje del informado (documentalista) y el lenguaje del no informado (usuario).

Unidad de Control Interno: Departamento (o actividad) dentro de una entidad, a la que se le confía revisiones y valoraciones de los sistemas de la entidad y procedimientos para minimizar la probabilidad de un fraude, errores y prácticas ineficientes. La auditoría interna debe ser independiente dentro de una organización y reportar directo a la gerencia. Un departamento, división o equipo de consultores u otros profesionales que proporciona una garantía independiente y objetiva así como servicios de consultoría diseñados para añadir valor y mejorar la operativa de la organización. El Control Interno ayuda a la organización a cumplir sus objetivos mediante el uso de un enfoque sistemático y disciplinado para evaluar y mejorar la eficacia en los procesos de gestión del riesgo, control y administración (IIA).

Valores éticos: Los valores éticos son los que permiten que quien toma una decisión determine un curso apropiado de conducta, estos valores deben estar basados en lo que es "correcto", que puede ir más allá de lo que es legalmente requerido. (COSO 1992).

Valoración de riesgo: Significa estimar el significado de un riesgo y valorar la probabilidad de la ocurrencia de éste.