



UNIVERSIDAD NACIONAL  
AVENIDA DE  
MEXICO

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN  
INGENIERÍA**

FACULTAD DE INGENIERÍA

**PROCESAMIENTO DE INFORMACIÓN PARA UN  
SISTEMA DE INTELIGENCIA TECNOLÓGICA  
CASO DE APLICACIÓN  
“DESALACIÓN DE AGUA DE MAR”**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN INGENIERÍA**

INGENIERÍA DE SISTEMAS - INVESTIGACIÓN DE OPERACIONES

P R E S E N T A:

**ROBERTO SOLIS GONZÁLEZ**

TUTOR:

**M. I EUGENIO LÓPEZ ORTEGA**

2008





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**JURADO ASIGNADO:**

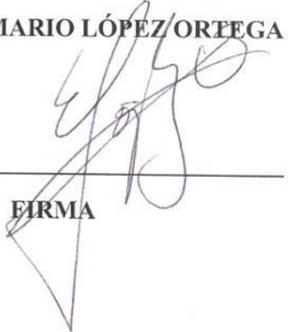
Presidente: Dr. Suárez Rocha Javier  
Secretario: Dr. Aceves García Ricardo  
Vocal: M. I. López Ortega Eugenio Mario  
1er. Suplente: Dr. Sánchez Guerrero Gabriel De Las Nieves  
2do. Suplente: Dra. Elizondo Cortes Mayra

Lugar donde se realizó la tesis:

Ciudad Universitaria, México, Distrito Federal

**TUTOR DE TESIS:**

**M. I. EUGENIO MARIO LÓPEZ ORTEGA**

  
\_\_\_\_\_  
**FIRMA**

**Agradecimiento:**

A mi tutor, el M.I. Eugenio López Ortega, por su paciencia y asesoría que sin ella no hubiera sido posible concluir este trabajo

**Dedicatoria:**

A mi hijo Leonardo

A mi esposa Raquel

A mis padres y hermanos

## ÍNDICE

INTRODUCCIÓN .....	3
PROBLEMÁTICA: .....	5
PROBLEMA .....	5
MÉTODO DE SOLUCIÓN .....	5
CAPÍTULO 1 .....	7
SISTEMA DE INTELIGENCIA TECNOLÓGICA (SIT) .....	7
INTRODUCCIÓN .....	7
1.1INTELIGENCIA COMPETITIVA (IC) .....	9
1.2INTELIGENCIA TECNOLÓGICA (IT) .....	10
1.3MODELOS DE INTELIGENCIA TECNOLÓGICA.....	11
1.4SISTEMA DE INTELIGENCIA TECNOLÓGICA (SIT) .....	13
1.4.1FASE DE IMPLANTACIÓN .....	13
1.4.2FASE DE OPERACIÓN.....	16
CAPÍTULO 2 .....	18
PROCESAMIENTO DE LA INFORMACIÓN .....	18
INTRODUCCIÓN .....	18
2.1 CIENCIOMETRÍA .....	19
2.2 ANÁLISIS DE PALABRAS ASOCIADAS.....	21
2.3 MAPAS TECNOLÓGICOS .....	24
CAPÍTULO 3 .....	27
ESCALAMIENTO MULTIDIMENSIONAL .....	27
INTRODUCCIÓN: .....	27
3.1 MODELO GENERAL DE ESCALAMIENTO MULTIDIMENSIONAL. ....	28
3.2 MODELO DE ESCALAMIENTO MÉTRICO .....	33
3.3 MODELO DE ESCALAMIENTO NO MÉTRICO .....	35
3.4 MODELO DE ESCALAMIENTO PONDERADO O DE DIFERENCIAS INDIVIDUALES (INDSCAL).....	38
3.5 ESCALAMIENTO DESDOBLADO (UNFOLDIG).....	42
CAPÍTULO 4 .....	46
CASO DE APLICACIÓN “DESALACIÓN DE AGUA DE MAR” .....	46
CONCLUSIONES.....	63
ANEXO .....	66
BIBLIOGRAFÍA .....	78

## INTRODUCCIÓN

El desarrollo experimentado por la tecnología durante las últimas décadas ha generado un importante aumento de la disponibilidad de información. En este contexto, la literatura técnico-científica en las distintas disciplinas y áreas del conocimiento muestra un crecimiento y una dinámica sin precedentes.

Por otra parte, la misma tecnología ha incrementado la accesibilidad a la información técnico-científica; a través del Internet, la rapidez para consultar una patente o cualquier otro tipo de documento (artículos en revistas arbitradas, presentaciones en congresos y conferencias, etc), es casi inmediata. Sin embargo, la amplísima cantidad de documentos disponibles ha hecho que el tiempo que se requiere para analizar la información contenida en ellos haya crecido a mayor ritmo que la disminución en el tiempo para su consecución.

El procesamiento automatizado de esta información con fines de facilitar su análisis, ha sido una tendencia desde hace más de 10 años. Los desarrollos alrededor del tema conocido como minería de textos son cada vez más dinámicos. El objetivo de estos desarrollos tecnológicos consiste en procesar una gran cantidad de documentos identificando en su contenido la información más relevante. En la actualidad existen herramientas que van desde el simple conteo de palabras e identificación de los conceptos adyacentes a una palabra clave, hasta aquellos que analizan la sintaxis de las oraciones para identificar los verbos y los complementos más relevantes y generar reportes dirigidos a facilitar el análisis de la información contenida en los documentos procesados.

Otro enfoque más reciente dirigido al análisis de este tipo de información, es lo que se conoce como inteligencia tecnológica. La inteligencia tecnológica es un concepto relativamente nuevo, que tiene como objetivos la búsqueda de la *buena* información del entorno externo a la organización para luego convertirla en conocimientos útiles para la toma de decisiones. Los conocimientos generados a

partir de la aplicación de la inteligencia tecnológica deberán ser capaces de identificar patrones y tendencias, cuya interpretación sirva a la toma de las decisiones adecuadas en el ámbito tecnológico de la organización.

Se puede definir a la inteligencia tecnológica como el conjunto de conceptos, métodos y herramientas que sirven para desarrollar, de forma coordinada, las actividades de búsqueda, obtención, análisis y difusión de la información técnico-científica relevante, de cara a la toma de decisión en la organización de acuerdo con su estrategia de mercado previamente establecida.

Este trabajo presenta el Sistema de Inteligencia Tecnológica (SIT) creado en el Instituto de Ingeniería de la UNAM (II), como herramienta de apoyo en la toma de decisiones de carácter estratégico.

Un problema a resolver dentro de la inteligencia tecnológica corresponde a la recolecta y procesamiento de la información técnico-científica relevante, de tal manera que los expertos en el campo de interés puedan comprenderla, analizarla y discutirla, de manera grupal, para generar conocimientos sobre las tendencias y los principales actores<sup>1</sup>.

El facilitar la estructuración de un volumen creciente de información técnico-científica, representa un desafío para los interesados en generar conocimientos acerca de las tendencias tecnológicas en campos específicos y los principales actores en dichos campos.

Esta situación ha promovido el desarrollo de nuevas disciplinas como la cienciometría, cuyo objetivo es identificar métricas que permitan homogeneizar el análisis de la creciente información técnico-científica.

---

<sup>1</sup> Rodríguez, M; Escorsa, P. (1998), Transformación de la información a la Inteligencia Tecnológica en la Organización Empresarial: Instrumento para la toma de decisiones estratégicas”, RECITEC-Recife, Vol. 2, Núm. 3, pp. 177-202, Brasil.

## **PROBLEMÁTICA:**

Una problemática a resolver dentro de la inteligencia tecnológica corresponde al procesamiento de la información técnico-científica relevante, de tal manera que los expertos en el campo de interés puedan comprenderla, analizarla y discutirla, de manera grupal, para generar conocimientos sobre las tendencias y asociaciones de los principales actores.

## **PROBLEMA**

Por lo que la representación de la información para dar una mayor comprensión sobre un tema científico, es un problema a resolver de la mejor manera dado que una buena representación permita extraer esquemas y modelos más significativos con objeto de presentar conocimientos implícitos, desconocidos anteriormente y potencialmente útiles.

## **MÉTODO DE SOLUCIÓN**

El usar del Sistema de Inteligencia Tecnológica propuesto por el instituto de Ingeniería de la UNAM, para recolectar y procesar información técnico-científica que resulte relevante para un área tecnológica determinada.

El utilizar como herramienta de análisis de la información la cienciometría y en especial el análisis de palabras asociadas para poder ver las correlaciones entre los parámetros bibliométricos.

Proponer como método para visualizar los datos con todos sus matices y sin perder la riqueza informativa. escalamiento multidimensional el cual simplifica todo lo posible el complejo patrón de asociación existente entre las entidades.

Así ver las ventajas del método de escalamiento multidimensional como propuesta crear mapas tecnológicos que permitan ver agrupaciones (cluster) y posesiones relativas de conceptos para encontrar mayor información y prever tendencias de comportamiento de un determinado tema científico.

El uso del Sistema de Cómputo de Inteligencia Tecnológica (SCIT) desarrollado por II como base computacional para el desarrollo de las tareas necesarias para la inteligencia tecnológica.

Finalmente, utilizar los aspectos relativos al procesamiento de la información técnico-científica señalados anteriormente en un caso de aplicación en el tema de desalación de agua de mar. Así presentar los principales reportes que genera el sistema y que son enviados a expertos para su análisis y generación de conocimientos acerca de las tendencias y actores en el tema.

El desarrollo de la tesis se sustentó en las siguientes etapas:

1. Conocimiento del concepto de inteligencia tecnológica y de la propuesta desarrollada en el IIUNAM.
2. Conocimiento del concepto de Cienciometría y análisis de palabras asociadas.
3. Análisis del método de escalamiento multidimensional para la creación de mapas tecnológicos SIT propuesto para el IIUNAM.
4. Aplicación del SCIT en un caso específico: Desalación de agua de mar.

# CAPÍTULO 1

## SISTEMA DE INTELIGENCIA TECNOLÓGICA (SIT)

### OBJETIVO:

Este capítulo proporciona los conceptos de inteligencia competitiva, inteligencia tecnológica, así mismo se muestran brevemente algunos de los modelos de inteligencia tecnológica que sean desarrollado, así dando lugar a la descripción del Sistema de Inteligencia Tecnológica (SIT) propuesto por el Instituto de Ingeniería de la UNAM, describiendo las fases y etapas involucradas en el.

### INTRODUCCIÓN

Antes de las últimas cuatro décadas, el conocer la evolución y resultados de los esfuerzos tecnológicos era relativamente sencillo, dado que la comunidad científica y tecnológica era reducida en número y en países. Los principales trabajos científicos aparecían en un volumen manejable de publicaciones y la velocidad de aparición de tecnologías era más lenta.

Actualmente, la situación global del entorno competitivo, la gran dinámica en el desarrollo de nuevas tecnologías y la gran cantidad de información disponible, obligan a los decisores en las empresas y en los organismos públicos a gestionar la información de manera más rápida, estratégica y efectiva.

Esta situación dio un gran auge al acopio y manejo de información para evaluar el entorno, los competidores y la evolución tecnológica y de mercado relevante para el desarrollo de las organizaciones. Hacia finales de los años ochenta se acuñó el concepto de *vigilancia tecnológica* que consiste en el análisis del entorno tecnológico de la empresa.

La *vigilancia tecnológica* es el esfuerzo sistemático y organizado que realiza una empresa para la observación, captación, análisis y difusión precisa de información sobre los avances del estado de la técnica, particularmente las áreas tecnológicas que le generan oportunidades o amenazas para su desarrollo.

La vigilancia implica observación, verificación y puesta al día sobre los desarrollos habitualmente en áreas de interés bien definido para un propósito específico (Coates et al. 1986 citado por Alan Porter). Implica escrutar el entorno apropiado en busca de información pertinente.

Ya que es esencial que la información que se distribuye tenga significado para sus destinatarios, especialmente los que toman decisiones, la característica esencial para que la información sea útil y tenga significado, es su valorización.

El esfuerzo colectivo y organizado de adquisición, tratamiento y difusión de la información que supone la vigilancia tecnológica, sólo se justifica en la medida que sea capaz de aportar conocimientos útiles a los tomadores de decisiones. Es en este caso cuando es posible hablar de la función de inteligencia, en el sentido Anglosajón de "intelligence", (Palop F et al, 1993)

La relevancia de contar con conocimientos del entorno de la organización para la adecuada toma de decisiones, motiva el desarrollo del concepto de inteligencia competitiva. Este concepto se gesta durante la década de los 80's y se consolida en los siguientes 10 años.

La inteligencia competitiva trata del proceso de obtención, análisis, interpretación y difusión de la información con valor estratégico sobre la industria y los competidores, que se transmite a los responsables de la toma de decisiones en el momento oportuno. En consecuencia, el objetivo de la inteligencia competitiva es ayudar a las empresas e instituciones en la toma de decisiones estratégicas y mejorar su capacidad competitiva e innovadora.

## 1.1 INTELIGENCIA COMPETITIVA (IC)

La Inteligencia Competitiva pretende determinar qué información sobre el entorno es la de mayor valor y que, por lo tanto, es necesario obtener, qué medios utilizar, a quién recurrir, cómo transmitirla y, sobre todo, cómo generar oportunamente un resultado incorporable a la toma de decisiones de la organización<sup>1</sup>.

El concepto de Inteligencia Competitiva según Kahaner lo define como:

*Un programa sistemático de recopilación y análisis de información acerca de las actividades de los competidores y en general de las tendencias de los negocios con el fin de lograr las metas establecidas.*

En la actualidad, los dos principales aspectos que definen metas de tipo estratégico de una organización son el comportamiento de los mercados y las tendencias tecnológicas dominantes. Lo anterior supone dos tipos de inteligencia competitiva:

- Inteligencia de mercado
- Inteligencia de tecnología

La inteligencia de mercado se orienta a la generación de conocimientos relacionados con el comportamiento del consumo y de los esfuerzos que realizan los competidores por permanecer y aumentar su participación en el mercado. En los países europeos este tipo de inteligencia se conoce como inteligencia económica.

Por su parte, la inteligencia tecnológica se vincula con el conocimiento relativo a las direcciones predominantes en el desarrollo tecnológico, a la identificación de los principales actores y a las acciones que realizan los competidores en los diversos temas tecnológicos.

---

<sup>1</sup> La integración entre la gestión del conocimiento y la inteligencia competitiva: la aportación de los mapas tecnológicos. Pere Escorsa, Ramón Maspons e Ivette Ortiz. Espacios. Vol. 21 (2) 2.000

## 1.2 INTELIGENCIA TECNOLÓGICA (IT)

La Inteligencia Competitiva enfocada hacia el conocimiento del entorno estratégico del progreso en Ciencia y Tecnología, se denomina Inteligencia Tecnológica, que puede definirse como:

*Un sistema para detectar, analizar y emplear información sobre eventos técnicos, tendencias y, en general, actividades o aspectos clave para la competitividad de la empresa, con el propósito de obtener una mejor explotación de la tecnología<sup>2</sup>*

La Inteligencia Tecnológica es un proceso enfocado a monitorear el ambiente tecnológico de una organización con la finalidad de que, con base en este conocimiento, los directivos cuenten con elementos para tomar decisiones más acertadas, en aspectos relacionados con innovación tecnológica<sup>3</sup>.

El proceso de Inteligencia Tecnológica, proporciona elementos de utilidad para tomar decisiones de mediano y largo plazo, brinda un método sistemático suficientemente lógico para generar conocimiento a partir de la información con la que se cuenta.

---

<sup>2</sup> Ashton, B.; Stacey, G. (1995), "Technical intelligence in business: understanding technology threads and opportunities", *International Journal of Technology Management*, Vol. 10, N°1

<sup>3</sup> Guía para implantar un sistema de inteligencia Tecnológica en organizaciones que realizan I &D. Linares Gerson G. ,Tesis de maestría ,2004.

### **1.3 MODELOS DE INTELIGENCIA TECNOLÓGICA**

Si bien las actividades de Inteligencia orientadas hacia el beneficio de los negocios tienen una trayectoria de muchos años, sólo desde 1986, con la creación de en Estados Unidos de la Society of Competitive Intelligence Professionals (SCIP), se observa un publicaciones orientadas a dar a conocer los beneficios de la inteligencia tecnológica aplicados a negocios, y las recomendaciones generales para su implantación.

Algunas empresas han realizado distintos esfuerzos para utilizar la Inteligencia tecnológica. A continuación se describen brevemente algunos modelos que han implementado algunas empresas para realizar inteligencia tecnológica.

La corporación Motorola es pionera entre las organizaciones con respecto a las prácticas avanzadas de procesos de inteligencia competitiva. Este carácter previsor de su cultura, les ha llevado a ser destacados practicantes en herramientas de previsión tecnológica, como los “technology road maps”.

Otra empresa pionera es Elf-Atochem filial del grupo petroquímico Elf Aquitaine, esta tiene una red de observadores formada por 70 personas, de los cuales una docena (expertos) están dedicados a la inteligencia tecnológica. De 800 referencias identificadas semanalmente, 85 son retenidas por su valor, para su análisis por el sistema de vigilancia tecnológica.

Scheumberg empresa de petrolíferos, gas y agua extendida por más de 100 países, la inteligencia tecnológica se apoya fuertemente del Internet como describe S. Guthery el Technology match una aplicación que aprovecha la red de recursos internos y externos para facilitar a los directivos, investigadores e ingenieros información sobre el estado del arte en dominios técnicos específicos. En 1994 establecieron un grupo de coordinación de inteligencia tecnológica, la cual encuestó a los técnicos, que identificaron 26 áreas de técnicas de interés y

determinaron con ellas tecnologías que deberían ser objeto de vigilancia tecnológica.

La empresa L'Oréal grupo cosmético francés desarrollo un sistema de 1,500 investigadores quienes elaboran un formulario sobre información sensible para la empresa, de ahí pasan a al dirección general de investigación, organizándose en un a base de datos, uno de cuyos subproductos es un boletín semanal, difundido a los responsables de investigación del grupo. Estos a su vez cierran el flujo de este sistema al interactuar con sus investigadores.

SOMFY empresa francesa de motores eléctrico para la elevación de persianas, puertas, utiliza un modelo de inteligencia muy organizado. Analiza la información de diferentes fuentes por especialistas temáticos que la clasifican por el servicio de documentación,. Como ejemplo en el caso de patentes se recuperan 1000 patentes al año de base de datos, se seleccionan 3000 y después de ser analizadas por tres especialistas del departamento de ingeniería, de proyectos de electromecánica y de electrónica, se realiza una síntesis y se difunde con una pagina por patente entre el departamento e ingeniería y marketing.

Daimler Benz Aerospace el socio Alemán del airbus europeo constituye un referente mundial en al gestión de información, destacando particularmente su empleo de la técnica prospectiva de escenarios. Como indica su vicepresidente F. Tessun la inteligencia tecnologica constituye para ellos una herramienta necesaria para observar lo que va a pasar en un futuro inmediato y si las previsiones realizadas en los planes son correctas. Para ello la técnica de escenarios les permite reflexionar sobre el futuro de diez a veinte años.

## 1.4 SISTEMA DE INTELIGENCIA TECNOLÓGICA (SIT)

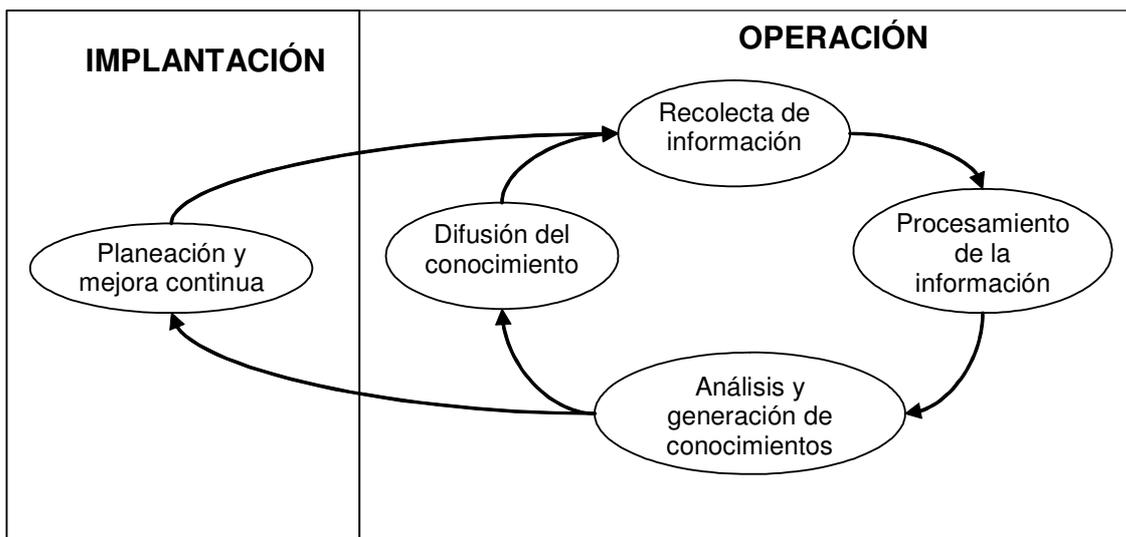
Como un esfuerzo del Instituto de Ingeniería de la UNAM se desarrolló el Sistema de Inteligencia Tecnológica (SIT) para llevar a cabo la implementación de la Inteligencia Tecnológica.

El Sistema de Inteligencia Tecnológica (SIT), se realiza mediante un ciclo sistemático y consta de dos fases, las cuales son:

Fase: Implantación

Fase: Operación

Figura 1.1 ■ Sistema de Inteligencia Tecnológica

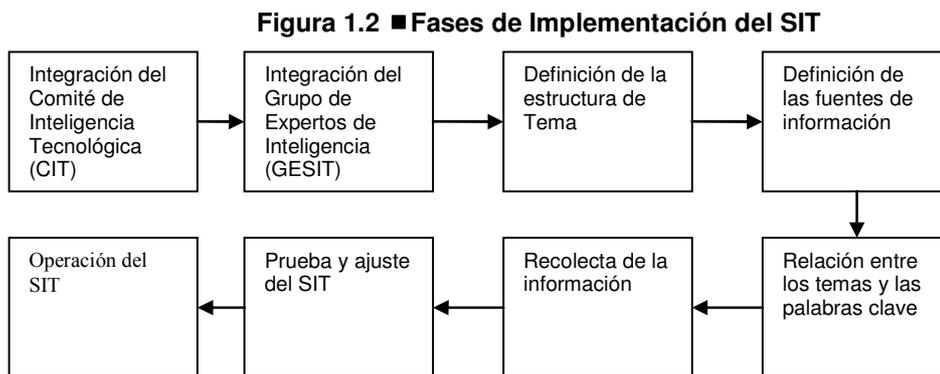


Fuente: López Eugenio. (2001). "Sistema de Inteligencia Tecnológica. El caso del área de catálisis FCC en la refinación del petróleo". RECITEC, Recife, v.5, n.2,

### 1.4.1 FASE DE IMPLANTACIÓN

La Fase de implantación corresponde al periodo donde el sistema es iniciado dentro de la organización y representa el paso de planeación.

La Fase de implantación del Sistema de Inteligencia Tecnológica esta formada por diferentes etapas que se muestran en la figura 1.2:



Fuente: López Eugenio. (2001). "Sistema de Inteligencia Tecnológica. El caso del área de catálisis FCC en la refinación del petróleo". RECITEC, Recife, v.5, n.2,

La primera etapa de la implementación del SIT es la integración del Comité de Inteligencia Tecnológica (CIT). Este comité lideraría la implementación del SIT y podría estar constituido por el nivel gerencial más alto de la organización. El CIT seleccionará los temas tecnológicos a analizar dentro del SIT.

La segunda etapa consiste en la creación del Grupo de Expertos de Inteligencia Tecnológica (GESIT) relacionado con el tema de investigación seleccionado.

El fundamento de la creación de este grupo es el conocimiento y experiencia acumulados por determinados especialistas en su campo de investigación, lo que los convierte en poseedores de criterios sólidos para poder valorar adecuadamente las características, tendencias y resultados de la actividad de investigación en dicho campo.

Este grupo tendría la responsabilidad de analizar la información previamente procesada con el fin de convertirla en conocimiento útiles a la toma de decisiones.

La tercera etapa es la definición de la estructura del tema, esta etapa es realizada por el GESIT y consiste en la descomposición del tema en subtemas. El objetivo de la estructuración del tema es poder clasificar la información a recopilar para facilitar el manejo de la información.

La cuarta etapa consiste en la definición de las fuentes de información. Teniendo definidos los temas y subtemas de interés por el GESIT, se pasa a la identificación de las fuentes de información correspondientes a dichos temas y subtemas. Para definir las fuentes de información se solicita a GESIT la enumeración y clasificación de las principales publicaciones, instituciones, investigadores, etc.

La quinta etapa define las palabras clave y los descriptores relacionados con cada subtema. Se solicita al GESIT la lista de palabras clave para cada tema y subtemas identificados. Estas palabras claves se utilizan para realizar la búsqueda de información en las fuentes definidas anteriormente.

La sexta etapa es la recolección de la información basada en las fuentes de información definidas en la cuarta etapa. La colección de la información es capturada en el software que fue creado para dar soporte a la operación del sistema.

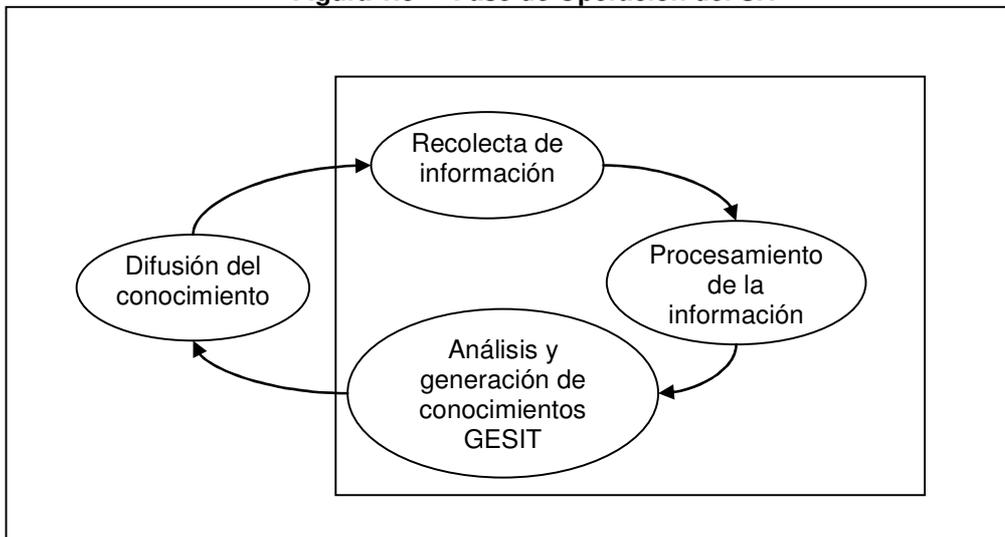
Los resultados obtenidos son presentados al GESIT para probar y ajustar la fase de Implantación. Cuando los miembros del GESIT están satisfechos con la información, se puede iniciar la fase de operación.

Cada proceso es documentado como se realiza al implantar un sistema de gestión de la calidad sustentado en la normas de la familia ISO 9000.

## 1.4.2 FASE DE OPERACIÓN

La Fase de operación del Sistema de Inteligencia Tecnológica esta formada por cuatro diferentes etapas las cuales se muestran en el esquema siguiente.

Figura 1.3 ■ Fase de Operación del SIT



Fuente: López Eugenio. (2001). "Sistema de Inteligencia Tecnológica. El caso del área de catálisis FCC en la refinación del petróleo". RECITEC, Recife, v.5, n.2,

La primer etapa está relacionada con la recolecta de la información que se requiere y que fue definida en la fase de implementación. Para ello es necesario realizar la búsqueda de la información en las fuentes de una manera sistematizada para facilitar su manejo y posterior análisis.

El procesamiento de la información recopilada corresponde a la segunda etapa con el fin de crear informes y reportes que son enviados al GESIT. Las técnicas que se utilizan para generar los reportes e informes están basadas en la Cienciometría y en métodos estadísticos de análisis multivariado. Los resultados de estos análisis se incorporan a los reportes presentados para el análisis por el grupo de expertos.

Para poder realizar el procesamiento de la información es necesario contar con un sistema computacional para el almacenamiento y procesamiento de la información. El instituto de Ingeniería desarrollo un Sistema de Cómputo de Inteligencia Tecnológica (SCIT), el cual será descrito en un capítulo posterior.

En la tercera etapa los integrantes del GESIT analizan los informes y reportes de la información que ha sido procesada en la etapa 2 para dar inicio a la generación de conocimiento. El identificar nuevas tendencias tecnológicas en cada subtema es un ejemplo de generación de conocimiento gracias al análisis de la información.

En la cuarta etapa se distribuyen los análisis y los conocimientos generados por el GESIT para dar inicio al ciclo nuevamente de la operación de del SIT. El objetivo de esta etapa es lograr que los conocimientos que generará el GESIT sean conocidos por toda la organización para poder sustentar las decisiones con relación al tema estudiando.

## **CAPÍTULO 2**

### **PROCESAMIENTO DE LA INFORMACIÓN**

#### **OBJETIVO:**

Este capítulo proporciona una orientación para el desarrollo de la etapa de procesamiento de la información de la fase de implementación del SIT propuesto por Instituto de Ingeniería, utilizando para ello herramientas de análisis tales como *cienciometría* y el análisis de palabras asociadas, que son la base para la construcción de los mapas tecnológicos con los cuales se busca generar conocimiento a partir del análisis de la estructuras mostradas en el mapa

#### **INTRODUCCIÓN**

Una vez realizada la recolecta de la información es necesario realizar el procesamiento de la información que es la parte central de la inteligencia tecnológica. Para llevarla a cabo esta etapa se aplican diferentes procesos y análisis a la información recopilada para darle un mayor valor y obtener conclusiones de utilidad para el tomador de decisiones.

En esta etapa se crean productos de información a la medida de las necesidades del decisor. Se desarrolla información avanzada para la toma de decisiones, incluyendo recomendaciones y nuevas propuestas generando así conocimiento.

Dependiendo de los objetivos de una empresa, pueden llevarse a cabo distintos tipos de inteligencia competitiva como se mencionó en el capítulo 1. Por lo que dependiendo el tipo de inteligencia que se desarrolle se debe utilizar las herramientas específicas, potencialmente útiles para la valorización de la información.

Para la inteligencia de mercado la técnicas utilizadas son:

- el modelo de las cinco fuerzas competitivas de Porter
- el análisis DAFO, el análisis del perfil del competido
- técnicas de benchmarking

Para la inteligencia Tecnológica la técnicas utilizadas son:

- Co-citaciones
- Cienciometría

## **2.1 CIENCIOMETRÍA**

La inteligencia tecnológica se sustenta importantemente en la explotación estadística de datos científicos y tecnológicos expresados a través de documentos. Para ello hace uso de la Cienciometría, la cual es una disciplina dirigida a medir el desarrollo científico y tecnológico a través del análisis de sus productos documentados: artículos y patentes.

La cienciometría permite una explotación rápida y eficiente de enormes cantidades de información procedentes de las bases de datos científicas o técnicas: artículos científicos, patentes, tesis doctorales, ponencias de congresos y otros documentos públicos.

La Cienciometría de alguna manera, intenta identificar las leyes que rigen la actividad científica, parte de la base de que los resultados de las investigaciones científicas y técnicas se plasman en forma escrita a través de artículos de revistas, memorias de patentes. El progreso científico y técnico queda registrado en escritos. El análisis de esta documentación escrita permitirá, pues, descubrir sus características y su evolución.

La Cienciometría estudia la incidencia que determinada disciplina científica-tecnológica tiene entre los trabajos de la comunidad investigadora, y por lo tanto,

los recursos que a priori se están destinando, o el número de patentes que se publican sobre determinada línea y qué empresas están detrás de tales trabajos.

Asímismo, la Cienciometría permite detectar las relaciones entre las distintas líneas de trabajo o entre centros técnicos y empresas entre sí. Cada artículo científico, patente, etc., es una asociación momentánea de problemas, técnicas y personas, ordenadas según una lógica de investigación de trabajo, por lo que cada artículo es una fuente multidimensional de datos con los que trabajar.

La Cienciometría se basa en el análisis y cómputo de determinados indicadores bibliométricos: autores de artículos, citas que aparecen en la bibliografía de cada artículo, palabras (key words) contenidas en los títulos de los artículos o en los resúmenes. Estos indicadores están presentes en los registros de las bases de datos. Mediante el recuento de estos indicadores se puede determinar:

- a) el crecimiento de cualquier campo de la ciencia, según la variación cronológica del número de trabajos publicados en él.
- b) el envejecimiento de los campos científicos, según la vida media de las referencias de sus publicaciones.
- c) la evolución cronológica de la producción científica, según el año de la publicación de los documentos.
- d) la productividad de los autores o instituciones, medida por el número de sus trabajos.
- e) la colaboración entre los científicos o instituciones, medida por el número de autores por trabajo o centros de investigación que colaboran.

f) el impacto o visibilidad de las publicaciones dentro de la comunidad científica internacional, medido por el número de citas que reciben éstas por parte de trabajos posteriores.

g) el análisis y evaluación de las fuentes difusoras de los trabajos, por medio de indicadores de impacto de las fuentes.

h) las trayectorias tecnológicas seguidas por empresas o países en un período determinado, de acuerdo con su actividad patentadora

## **2.2 ANÁLISIS DE PALABRAS ASOCIADAS**

Para poder detectar las relaciones existentes de un determinado tema científico, la Cienciometría usa el modelo de palabras asociadas (co-ocurrencia de palabras o co-word analysis) que surgió a mediados de los años ochenta, a partir de un proyecto francés desarrollado por el Centre de Sociologie de l'Innovation (CSI) de l'École des Mines de Paris y l'Institut National d'Information Scientifique et Technique (INIST) del CNRS, Su finalidad era procesar los términos científicos de una base de datos científica y multidisciplinaria, que agrupaba los trabajos de autores europeos principalmente (Callon, Courtial y Laville, 1991).

Las palabras asociadas, según los expertos, consiste en reducir el texto científico o técnico al conjunto de apariciones conjuntas entre las palabras que lo componen. Es decir, una especialidad de investigación puede ser identificada por su propio vocabulario, o más exactamente, por las particulares asociaciones que establece entre palabras.

El modelo de palabras asociadas se basa en el recuento del número de ocurrencias de una palabra clave o grupos de palabras clave en los documentos seleccionados, así como en la localización de las co-ocurrencias o citas coincidentes de las palabras clave seleccionadas.

El análisis de estos indicadores refleja la actividad científica y técnica de las organizaciones, países y empresas. Por ejemplo, puede ser útil seguir la pista del aumento o disminución de las patentes o publicaciones en un determinado período de tiempo para identificar tecnologías emergentes y en desarrollo.

Estas técnicas pueden también poner de manifiesto cualquier tipo de correlación y relación entre parámetros seleccionados, muy difíciles de detectar por una persona a partir de grandes cantidades de información. Así, es posible establecer las relaciones entre campos de actividades o tecnologías mediante el análisis de la co-ocurrencia de palabras clave. Por ejemplo, es posible identificar colaboraciones entre autores o instituciones en un cierto campo de investigación, reconocer tecnologías emergentes o detectar las distintas aplicaciones de una tecnología en varios mercados.

## LA MATRIZ DE CO-OCURRENCIAS

El primer paso para realizar un análisis de palabras asociadas consiste en la transformación de los datos en una matriz de asociación. Se dice que dos palabras co-ocurren cuando aparecen simultáneamente en el mismo documento.

Dos palabras estarán más asociadas entre sí cuanto mayor sea la co-ocurrencia entre ellas. Por tanto, la medida del enlace entre dos palabras de una red será proporcional a la co-ocurrencia de esas dos palabras en el conjunto de documentos que se tome como muestra.

**Tabla 3.1 ■ Matriz de asociaciones entre palabras clave**

	<i>Pal.1</i>	<i>pal.2</i>	<i>Pal.i</i>	<i>Pal.j</i>	<i>Pal.1000</i>
<i>Pal.1</i>	–	20	20	0	2
<i>Pal.2</i>	–	–	0	5	0
<i>Pal.i</i>	–	–	–	20	0
<i>Pal.j</i>	–	–	–	–	0
<i>Pal.1000</i>	–	–	–	–	–

La matriz de asociaciones o de co-ocurrencias es una matriz de adyacencia cuadrada simétrica. Cada elemento representa la asociación entre los descriptores. En la celdilla  $C_{ij}$  se coloca el número de documentos en los que la palabra "i" y la palabra "j" aparecen simultáneamente.

Como se observa en la celda,  $C_{12}$  es igual a 20, lo que significa que las palabras "i" y "j" aparecen juntas en un total de 20 documentos.

En teoría, a partir de esta matriz de adyacencia se podría reconstruir completamente la red cienciométrica que genera el campo científico en estudio, pero en la práctica no es conveniente, ya que los valores de las co-ocurrencias dependen del tamaño de la muestra. Por lo tanto es conveniente recurrir a la normalización de los valores de las co-ocurrencias.

## ÍNDICE DE ASOCIACIÓN

Para realizar la normalización se utiliza el **índice de asociación** o de equivalencia (Michelet B, 1988):

$$e_{ij} = \frac{C_{ij}^2}{C_i C_j}$$

Donde:  $e_{ij}$  = Índice de equivalencia o de asociación entre las palabras  $i$  y  $j$ .

$C_{ij}$  = Co-ocurrencia de las palabras  $i$  y  $j$ .

$C_i$  = Ocurrencia de la palabra  $i$ .

$C_j$  = Ocurrencia de la palabra  $j$ .

La ecuación anterior se puede reescribir de la siguiente forma:

$$e_{ij} = \frac{C_{ij}}{C_i} \times \frac{C_{ij}}{C_j}$$

Donde: El primer factor es la probabilidad de tener la palabra  $i$  cuando se tiene la palabra  $j$ ,

El segundo factor es la probabilidad de tener la palabra  $j$  cuando se tiene la palabra  $i$ .

Este índice es una medida de la relación "e" entre las palabras  $i$  y  $j$ .

Los valores de  $e_{ij}$  oscilan entre 0 y 1. Cuando dos palabras no aparecen juntas, su co-ocurrencia es nula y el índice de equivalencia vale cero. En cambio, cuando dos palabras siempre que aparecen lo hacen juntas, el índice de equivalencia es la unidad.

## 2.3 MAPAS TECNOLÓGICOS

Dada la imposibilidad de asimilar la enorme cantidad de información disponible, es crucial filtrar los hechos significativos. Con los mapas tecnológicos se agregan los datos de una forma que ningún experto es capaz de hacer.

Los mapas tecnológicos dan una visión de la estructura de relaciones que existe en un ámbito tecnológico. Estos mapas muestran un panorama detallado de las líneas de investigación en el área considerada, a través del examen de lo que se está publicando o patentando, con una reducción de la información, lo que es muy importante. Además, los mapas son independientes de las opiniones individuales.

Los mapas tecnológicos representan la proximidad o relación que existe entre los conceptos que conforman el campo tecnológico. Esta relación puede cuantificarse mediante diversos índices y métricas y, gracias a éstas, dibujarse obteniendo así los mapas tecnológicos. En este trabajo se utilizó el índice de asociación basado en el análisis de palabras asociadas.

Los métodos utilizados para generar los mapas tecnológicos se basan en la clasificación de información, o sea, la formación de clusters (clasificación de la información) y su representación en mapas bidimensionales de conceptos., el cual radica en la reducción de la dimensionalidad sin perder la riqueza informativa. La distancia en el mapa entre dos conceptos (referentes a tecnologías, productos, autores, empresas...) señalará la mayor o menor relación entre ellos.

Así gracias a los mapas tecnológicos es posible visualizar la estructura de un área científico-tecnológica estudiando la co-ocurrencia de los diferentes campos de los registros bibliográficos. En los mapas cada documento (podría ser una patente) ocupa un lugar en el espacio, en función de sus contenidos temáticos.

## **SEÑALES DÉBILES**

Tratados dinámicamente, es decir, comparando mapas de diferentes períodos, permiten seguir la evolución en el tiempo de las tecnologías y las líneas de investigación.

La detección de señales débiles sobre la aparición de tecnologías emergentes presenta un enorme interés. La identificación y adopción tempranas de una nueva tecnología puede significar para la empresa el logro de una ventaja competitiva que la distancie de sus competidoras.

Estas señales débiles pueden identificarse mediante el análisis de las palabras nuevas que aparecen en los diversos campos de los registros de las bases de datos -en los títulos, abstracts, descriptores, etc.- que no existían en períodos anteriores. También a partir de palabras todavía poco usadas pero cuya frecuencia está creciendo rápidamente.

Las palabras compuestas o multiterminos deben ser objeto de una atención especial ya que muchas veces una nueva tecnología o una nueva área comienza

a denominarse por un vocablo resultante de la unión de dos palabras (como opto-electrónica).

Es evidente que captar estas señales débiles puede tener gran importancia en la estrategia tecnológica de una empresa o área de investigación.

## **ANÁLISIS MULTIVARIANTE**

Estos métodos son llamados “métodos de reducción de la dimensión” porque su función es simplificar todo lo posible el complejo patrón de asociación existente entre las variables. Este proceso de simplificación consiste en la proyección de un gran número de dimensiones en un número menor de ellas, generalmente dos o tres.

El principal problema con que nos encontramos en el análisis multivariante radica en la forma mediante la cual visualizamos los datos con todos sus matices y sin perder la riqueza informativa. Dentro de las técnicas de análisis estadístico multivariante que cumplen con esta tarea, nos encontramos con tres básicas:

- 1) análisis de cluster o clustering,
- 2) análisis de componentes principales (PCA), y
- 3) escalamiento multidimensional (MDS).

En el siguiente capítulo se analizará el método de Escalamiento Dimensional como opción para crear los mapas tecnológicos

## **CAPÍTULO 3**

### **ESCALAMIENTO MULTIDIMENSIONAL**

#### **OBJETIVO:**

Mostrar el método de Escalamiento multidimensional como opción para construir los mapas tecnológicos, que ayuden a la generación de conocimiento su análisis por parte del grupo de expertos involucrado en un Sistema de Inteligencia Tecnológica

#### **INTRODUCCIÓN:**

El Escalamiento Multidimensional, es un método que en su origen fue concebido como un procedimiento muy general de representación de datos para la construcción de una configuración de puntos conocida una determinada información de proximidad entre cada par de elementos del estudio.

Hoy día, este método está constituido por un conjunto de procedimientos y técnicas que tratan de representar y explicar coeficientes de proximidad entre conjuntos de objetos mediante distancias, en un entorno multidimensional. Por tanto, se puede definir el Escalamiento multidimensional como un conjunto de técnicas para la representación de objetos basándose en la aproximación de coeficientes de proximidad mediante distancias en un espacio de dimensión mínima y el posterior análisis de la configuración obtenida.

Son muchos los algoritmos que se han desarrollado para obtener la configuración que mejor se ajuste a las disimilaridades de partida. Estos algoritmos pueden clasificarse, atendiendo a un enfoque confirmatorio o exploratorio, en dos grandes grupos: modelos no probabilísticos y modelos probabilísticos, distinguiendo en ambos casos entre modelos métricos, si las consideraciones entre distancias y disimilaridades son cuantitativas, y modelos no métricos, si son cualitativas. El método de estimación empleado y las diferentes medidas de bondad de ajuste,

son los principales factores que determinan la existencia de los diferentes métodos de escalamiento multidimensional que se conocen en la actualidad.

### 3.1 MODELO GENERAL DE ESCALAMIENTO MULTIDIMENSIONAL.

De modo general, podemos decir que el MDS toma como entrada una matriz de proximidades,  $\Delta \in M_{n \times n}$ , donde  $n$  es el número de estímulos. Cada elemento  $\delta_{ij}$  de  $\Delta$  representa la proximidad entre el estímulo  $i$  y el estímulo  $j$ .

$$\Delta = \begin{pmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{pmatrix}$$

A partir de esta matriz de proximidades el MDS nos proporciona como salida una matriz  $X \in M_{n \times m}$ , donde  $n$ , al igual que antes, es el número de estímulos, y  $m$  es el número de dimensiones. Cada valor  $x_{ij}$  representa la coordenada del estímulo  $i$  en la dimensión  $j$  (más adelante veremos el procedimiento para obtener esta matriz).

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

Cada fila  $[x_{i1}, \dots, x_{im}]$  de esta matriz contiene las coordenadas del estímulo  $i$  que delimitan el espacio  $m$ -dimensional: A partir de la matriz es posible situar los  $n$  estímulos en el espacio asignándoles los valores de coordenadas correspondientes. También es posible utilizar la matriz  $X$  para calcular las distancias entre dos estímulos  $i$  y  $j$  cualesquiera.

A partir de esta matriz  $X$  se puede calcular la distancia existente entre dos estímulos cualesquiera  $i$  y  $j$ , simplemente aplicando la fórmula general de la distancia de Minkowski:

$$d_{ij} = \left[ \sum_{a=1}^n (x_{ia} - x_{ja})^p \right]^{1/p}$$

donde p puede ser un valor entre 1 e infinito. A partir de estas distancias podemos obtener una matriz de distancias que denominamos D ∈ M<sub>n</sub> × n:

$$D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}$$

La solución proporcionada por el MDS debe ser de tal modo que haya la máxima correspondencia entre la matriz de proximidades inicial Δ y la matriz de distancias obtenidas D. Para que exista la máxima correspondencia MDS proporciona varias medidas, que veremos más adelante, y que nos informan sobre la bondad del modelo.

### **EJEMPLO PRÁCTICO**

Para introducir el concepto puede emplearse el siguiente ejemplo: Supongamos que se está interesado en el estudio de n ciudades respecto a su posición geográfica. Si se dispone de un mapa donde se encuentran representadas las ciudades la construcción de la correspondiente matriz de distancias entre las ciudades a partir de dicho mapa se reduce a una cuestión gráfica elemental.

Supongamos por el contrario que se dispone de una matriz cuadrada (n x n) formada por las distancias lineales entre cada pareja de ciudades y se pretende reconstruir el mapa partiendo de dicha matriz. Este problema, en dimensión arbitraria, tiene solución exacta sobre la base de los trabajos de Schoenberg (1935) y Young & Houselholder (1938).

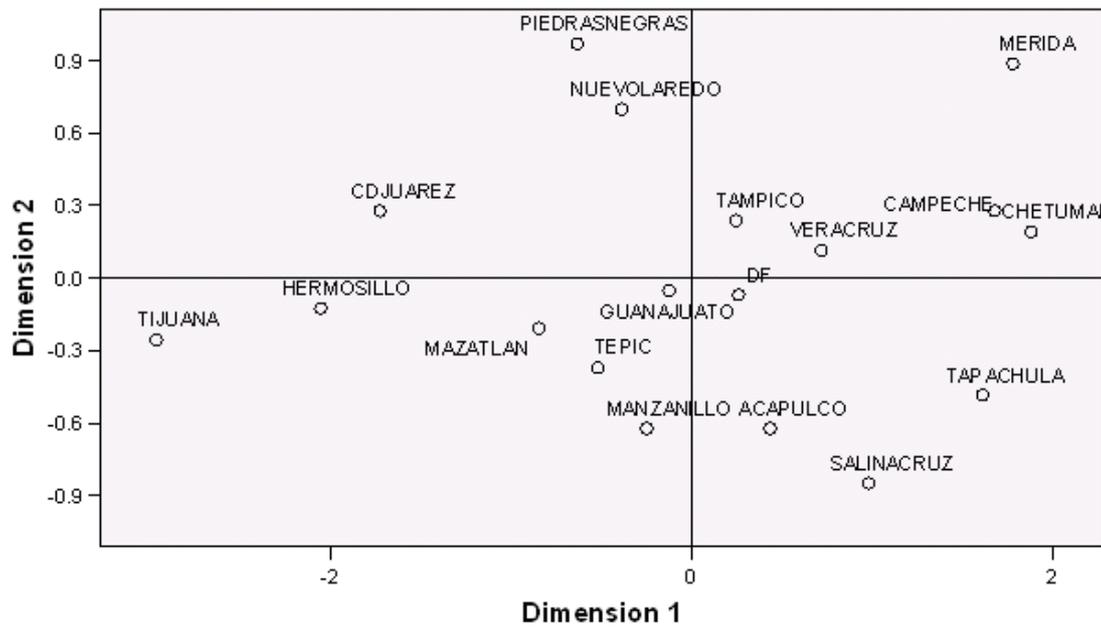
Consideremos la matriz Δ como las distancias de algunas ciudades de México. El objetivo es construir el mapa que exprese las distancias entre las ciudades.

**Tabla 3.2 ■ Matriz de distancias entre ciudades**

CIUDAD	VERACRUZ	TIJUANA	TEPIC	TAPACHULA	TAMPICO	SALINA CRUZ	PIEDRAS NEGRAS	NUEVO LAREDO	DF	MERIDA	MAZATLAN	MANZANILLO	HERMOSILLO	GUANAJUATO	CHETUMAL	CDJUAREZ	CAMPECHE	ACAPULCO
VERACRUZ	0	3,354	1,248	865	472	487	1,348	1,265	421	1,025	1,526	1,226	2,436	886	1,052	2,229	861	854
TIJUANA	3,354	0	2,071	4,145	2,981	3,644	2,458	2,611	2,880	4,338	1,758	2,527	857	2,542	4,299	1,309	4,174	3,135
TEPIC	1,248	2,071	0	2,002	977	1,508	1,405	1,237	807	2,265	292	459	1,195	537	2,228	1,675	2,102	1,064
TAPACHULA	865	4,145	2,002	0	1,411	439	2,497	2,072	1,247	1,274	2,280	1,840	3,386	1,627	1,257	2,977	1,066	1,089
TAMPICO	472	2,981	977	1,411	0	993	842	766	477	1,531	1,188	1,073	2,105	624	1,568	1,723	1,367	949
SALINA CRUZ	487	3,644	1,508	439	993	0	2,097	1,712	730	2,447	1,786	1,404	3,039	1,191	1,172	2,579	981	671
PIEDRAS NEGRAS	1,348	2,458	1,405	2,497	842	2,097	0	188	1,302	2,473	1,146	1,431	2,045	1,075	2,501	1,411	2,310	1,693
NUEVO LAREDO	1,265	2,611	1,237	2,072	766	1,712	188	0	1,180	2,290	1,158	1,294	2,075	892	2,318	1,338	2,127	1,068
DF	421	2,880	807	1,247	477	730	1,302	1,180	0	1,458	1,085	804	2,002	384	1,479	1,820	1,295	428
MERIDA	1,025	4,338	2,265	1,274	1,531	2,447	2,473	2,290	1,458	0	2,543	2,159	3,483	1,891	446	3,279	173	1,858
MAZATLAN	1,526	1,758	292	2,280	1,188	1,786	1,146	1,158	1,085	2,543	0	750	965	824	2,485	1,397	2,304	1,459
MANZANILLO	1,226	2,527	459	1,840	1,073	1,404	1,431	1,294	804	2,159	750	0	1,834	614	2,202	1,849	2,127	668
HERMOSILLO	2,436	857	1,195	3,386	2,105	3,039	2,045	2,075	2,002	3,483	965	1,834	0	1,723	3,488	766	3,300	2,486
GUANAJUATO	886	2,542	537	1,627	624	1,191	1,075	892	384	1,891	824	614	1,723	0	1,854	1,493	1,677	812
CHETUMAL	1,052	4,299	2,228	1,257	1,568	1,172	2,501	2,318	1,479	446	2,485	2,202	3,488	1,854	0	3,292	420	1,818
CDJUAREZ	2,229	1,309	1,675	2,977	1,723	2,579	1,411	1,338	1,820	3,279	1,397	1,849	766	1,493	3,292	0	3,090	2,179
CAMPECHE	861	4,174	2,102	1,066	1,367	981	2,310	2,127	1,295	173	2,304	2,127	3,300	1,677	420	3,090	0	1,786
ACAPULCO	854	3,135	1,064	1,089	949	671	1,693	1,068	428	1,858	1,459	668	2,486	812	1,818	2,179	1,786	0

Con las distancias establecidas entre cada elemento se busca la matriz  $X$  tal que  $D = X \cdot X^t$ . Utilizando el software SPSS se encuentran las matrices  $X$  de tal forma que podemos graficar los dos primeros componentes y dar la representación gráfica de las distancias de las ciudades como se muestra en la figura 3.1.

**Figura 3.4 ■ Gráfica de las distancias estimadas de las ciudades**



Este gráfico resulta una buena aproximación de la representación real de las distancias de las ciudades, como se observa en el siguiente gráfico 3.2 el cual es un mapa rotado de la República Mexicana

Figura 3.5 ■ Mapa rotado de la República Mexicana



El Stress obtenido para el ajuste de las distancias de las ciudades es de  
 $\text{Stress} = .09467$

Lo que indica que el ajuste es bueno, por lo que podemos dar como válida la representación de las distancias de las ciudades y así tener una representación gráfica de las ciudades.

## MODELOS DE ESCALAMIENTO MULTIDIMENSIONAL.

Existen dos modelos básicos de MDS que son: el modelo de escalamiento métrico y el modelo de escalamiento no métrico. En el primero de ellos consideramos que los datos están medidos en escala de razón o en escala de intervalo y en el segundo consideramos que los datos están medidos en escala ordinal.

### 3.2 MODELO DE ESCALAMIENTO MÉTRICO

Todo modelo de escalamiento parte de la idea de que las distancias son una función de las proximidades, es decir,  $d_{ij} = f(\delta_{ij})$ . En el modelo de escalamiento métrico partimos del supuesto de que la relación entre las proximidades y las distancias es de tipo lineal:  $d_{ij} = a + b\delta_{ij} + \xi$ .

El primer procedimiento de escalamiento métrico se debió a Torgerson (1952, 1958) quién se basó en un teorema de Young y Householder (1938), según el cual a partir de una matriz de distancias,  $D \in M_{n \times n}$ , se puede obtener una matriz  $B \in M_{n \times n}$  de productos escalares entre vectores. A su vez es posible descomponer la matriz  $B$  de productos escalares en el producto  $XX'$ , donde  $X$  es la matriz de coordenadas de los  $n$  estímulos en  $m$  dimensiones

El procedimiento consiste en transformar la matriz de proximidades  $\Delta \in M_{n \times n}$  en una matriz de distancias  $D \in M_{n \times n}$ , de tal forma que verifique los tres axiomas de la distancia euclídea:

1 No negatividad  $d_{ij} \geq d_{ii} = 0$

2 Simetría  $d_{ij} = d_{ji}$

3 Desigualdad triangular  $d_{ij} \leq d_{ik} + d_{kj}$

Los dos primeros axiomas son fáciles de cumplir, pero el tercer axioma no se cumple siempre. Este problema se conoce con el nombre de “estimación de la

constante aditiva". Torgerson solucionó este problema, estimando el valor mínimo de  $c$  que verifica la desigualdad triangular de la siguiente forma:

$$C_{\min} = \max_{i,j,k} (\partial_{ij} - \partial_{ik} - \partial_{kj})$$

De esta forma las distancias se obtienen sumando a las proximidades la constante  $c$ , es decir,  $d_{ij} = \partial_{ij} + c$ . Por ejemplo, supongamos que tenemos la siguiente matriz de proximidades:

$$\Delta = \begin{pmatrix} 0 & 1 & 5 \\ 1 & 0 & 2 \\ 5 & 2 & 0 \end{pmatrix}$$

Esta matriz no verifica la desigualdad triangular puesto que no se cumple que  $\partial_{31} \leq \partial_{32} + \partial_{21}$

$$5 \leq 2 + 1$$

Para calcular el valor mínimo de la constante aditiva  $c$  tendríamos que calcular todas las diferencias tal como se ha señalado anteriormente. En este caso se tendría que calcular  $\partial_{31} - \partial_{32} - \partial_{21}$  que es  $5 - 2 - 1 = 2$ . Estas diferencias las haríamos para todos los subíndices, obteniéndose que el valor mínimo de  $c$  es 2. La matriz de distancias sería en este caso:

$$D = \begin{pmatrix} 0 & 3 & 7 \\ 3 & 0 & 4 \\ 7 & 4 & 0 \end{pmatrix}$$

Una vez obtenida la matriz  $D \in M_{n \times n}$  es necesario transformarla en una matriz  $B \in M_{n \times n}$  de productos escalares entre vectores mediante la siguiente transformación:

$$b_{ij} = -\frac{1}{2} (d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2)$$

Donde

$$d_{i.}^2 = \frac{1}{n} \sum_j^n d_{ij}^2 \quad (\text{distancia cuadrática media por fila})$$

$$d_{.j}^2 = \frac{1}{n} \sum_i^n d_{ij}^2 \text{ ( distancia cuadrática media por fila)}$$

$$d_{..}^2 = \frac{1}{n^2} \sum_i^n \sum_j^n d_{ij}^2 \text{ ( distancia cuadrática media de la matriz)}$$

Una vez llegado a este punto, lo único que queda es transformar la matriz B en una matriz X ∈ M<sub>n</sub>×<sub>m</sub> tal que B=X·X', siendo X la matriz que nos da las coordenadas de cada uno de los n estímulos en cada una de las m dimensiones. Cualquier método de factorización permite transformar B en X·X'.

En resumen el procedimiento consiste en transformar:

Δ (Proximidades) → D (Distancias) → B (Productos escalares) → X (coordenadas)

### 3.3 MODELO DE ESCALAMIENTO NO MÉTRICO

A diferencia del escalamiento métrico, el modelo de escalamiento no métrico no presupone una relación lineal entre las proximidades y las distancias, sino que establece una relación monótona creciente entre ambas, es decir, si

$$\partial_{ij} < \partial_{kl} \Rightarrow d_{ij} \leq d_{kl}$$

Esto es que a medida que aumenten los valores de las proximidades, los de las distancias también aumenten o en todo caso no disminuyen. Es decir, la relación no se establece en forma de sumas o productos, sino simplemente en términos de la preservación del orden de las proximidades. Por este motivo se etiqueta este modelo como no- métrico

Su desarrollo se debe a Shepard (1962) quién demostró que es posible obtener soluciones métricas asumiendo únicamente una relación ordinal entre proximidades y distancias. Posteriormente Kruskal (1964) mejoró el modelo. El procedimiento se basa en los siguientes pasos:

- 1) Transformación de la matriz de proximidades en una matriz de rangos, desde 1 hasta (n (n - 1))/2.

2) Obtención de una matriz  $X \in M_{n \times n}$  de coordenadas aleatorias, es decir se sitúan los estímulos al azar en un espacio de  $r$  dimensiones (donde  $r$  es especificado por el usuario). A partir de esta matriz  $X$  inicial se calculan las distancias entre estímulos.

3) Comparación de las distancias calculadas en paso anterior con los rangos de las proximidades, transformándolas si es necesario para que sus rangos coincidan con estos, a las distancias obtenidas tras estas transformaciones se las denomina disparidades ( $\hat{d}_{ij}$ ). A partir de aquí podemos definir el error cuadrático como:

$$e_{ij}^2 = \left( d_{ij} - \hat{d}_{ij} \right)^2$$

4) Determinar una función de bondad de ajuste para evaluar cuánto se aproximan las distancias obtenidas a partir de  $X$  a las disparidades obtenidas a partir de la transformación de esas distancias. Esta función se conoce como nombre de Stress su formula es:

$$Stress = \sqrt{\frac{\sum_{i,j} \left( d_{ij} - \hat{d}_{ij} \right)^2}{\sum_i \sum_j \hat{d}_{ij}^2}}$$

Mientras mayor sea la diferencia entre las distancias y las disparidades, es decir, entre  $d_{ij}$  y  $\hat{d}_{ij}$ , mayor será el Stress y por tanto peor será el modelo. Por tanto, el Stress no es propiamente una medida de la bondad del ajuste, sino una medida de la no bondad o “maldad” del ajuste. Su valor mínimo es 0, mientras que su límite superior para  $n$  estímulos es  $\sqrt{1 - \left(\frac{2}{n}\right)}$

También se suele utilizar una variante del Stress que se denomina S-Stress, definida como:

$$S - Stress = \sqrt{\frac{\sum_{i,j} (d_{ij}^2 - \hat{d}_{ij}^2)^2}{\sum_i \sum_j \hat{d}_{ij}^2}}$$

5) Minimización del Stress. Dado que se partió de una matriz de coordenadas aleatorias, es de suponer que el ajuste nunca es bueno al principio. Por ello, se hace necesario llevar a acabo un proceso iterativo que vaya minimizando el valor del Stress. Estos se consigue alternando los valores de las coordenadas X de modo que la diferencia entre las distancias y disparidades derivadas a partir de ellos sea más pequeña que en el paso anterior.

La forma de llevar a cabo dicha reducción es sumar a la matriz X inicial una matriz de valores añadidos. Cada elemento de esta matriz contiene un valor que se sumara a la coordenadas del estímulo i en la dimension a. Este valor se determina mediante la expresión.

$$-\alpha \left( \frac{\partial S}{\partial x_{ia}} \right)$$

Donde:  $\left( \frac{\partial S}{\partial x_{ia}} \right)$  es la derivada del Stress con respecto a la coordenada a-ésima del estímulo i

$\alpha$  es una constante que representa el tamaño del paso

En la derivación de estos valores aparece el llamado problema del “mínimo local” y el “mínimo global”. El procedimiento iterativo se detiene cuando la función de Stress alcanza uno de los mínimos (donde la derivada vale 0) , sin embargo lo que interesa es hallar mínimo global. Los programas de escalamiento más recientes garantizan que la configuración adoptada como inicial se encontrará lo más cerca posible del mínimo global.

Otra medida que se suele utilizar es el coeficiente de correlación al cuadrado (RSQ), que nos informa de la proporción de variabilidad de los datos de partida que es explicada por el modelo. Los valores que puede tomar oscilan entre 0 y 1, al ser un coeficiente de correlación al cuadrado. Valores cercanos a 1 indican que el modelo es bueno y valores cercanos a 0 indican que el modelo es malo. Su expresión es:

$$RSQ = \frac{\left[ \sum_i \sum_j (d_{ij} - d_{..}) (\hat{d}_{ij} - \hat{d}_{..}) \right]^2}{\left[ \sum_i \sum_j (d_{ij} - d_{..})^2 \right] \left[ \sum_i \sum_j (\hat{d}_{ij} - \hat{d}_{..})^2 \right]}$$

### 3.4 MODELO DE ESCALAMIENTO PONDERADO O DE DIFERENCIAS INDIVIDUALES (INDSCAL)

Este modelo utiliza como entrada varias matrices de proximidades, por lo general, una por sujeto. El modelo fue originalmente ideado por Carroll y Chang (1970) para el caso métrico. Esta técnica se conoce como escalamiento multidimensional ponderado o WMDS (Weigthed Multidimensional Scaling). Por introducir diferencias individuales también se le denomina, escalamiento de diferencias individuales

El modelo de escalamiento de diferencias individuales (INDSCAL) supuso la segunda revolución dentro del MDS, tras la aparición del modelo no métrico, por cuanto supone una generalización del modelo euclidiano, de tal modo que obtienen una representación a partir de varias matrices de proximidades asumiendo que están difieren entre sí de forma sistemática y no aleatoria. Es decir, en lugar de considerar las diferencias entre matrices como sesgos en las respuestas de los sujetos, INDSCAL las contempla como diferencias preceptuales y cognitivas en el proceso de generación de las respuestas.

Aunque existen actualmente versiones no métricas del algoritmo INDSCAL, aquí se revisara la versión métrica.

Cada proximidad  $\delta_{i,j,k}$ , nos indicara la proximidad entre los estímulos  $i$  y  $j$  estimada por el sujeto  $k$ , Existen otras posibilidades en las que las proximidades de cada matriz corresponden a un tiempo diferente, o a proximidades estimadas en diferentes condiciones o en base a atributos diferentes de los estímulos.

El modelo considera que la relación entre las proximidades y las distancias es lineal. La novedad esta en la forma de estimar la distancia que es una modificación de la distancia euclidiana ordinaria, tratándola como una distancia euclidiana ponderada.

$$d_{ijk} = \sqrt{\sum_{a=1}^m w_{ka} (x_{ia} - x_{ja})^2}$$

Donde  $w_{ka}$  es el peso de l sujeto k-ésimo en la dimensión a- ésima

Así pues las dimensiones pueden verse encogidas en función del peso que tengan para el sujeto. Podemos interpretar pues el modelo INDSCAL como aquel en el que las diferencias individuales entre los sujetos surgen de las diferencias en los pesos otorgados a cada una de las distintas dimensiones que componen la solución común.

En la formula de distancia anterior si todos los pesos son iguales, la configuración de distancias entre los estímulos para cada sujeto será la del grupo total, es decir la solución común a todos los sujetos

A la configuración común de todos los sujetos se le conoce como espacio del grupo y suele diferir de la configuración propia de cada sujeto.

Cuando representamos la distancia entre los estímulos en función del peso que cada una de las dimensiones tiene para un individuo concreto, la configuración

de estímulos se verá encogida en aquellas dimensiones que tienen menor peso para el individuo. A esta configuración de distancias propia de este individuo se le conoce como el espacio del sujeto.

Así podemos decir que el INDSCAL representa las diferencias entre los juicios emitidos por los sujetos en términos de la importancia que cada uno de ellos otorga a cada una de las dimensiones que componen la solución, pero todas las dimensiones son comunes a todos los sujetos.

El procedimiento parte, al igual que el modelo métrico, de las proximidades, que se convierten en distancias absolutas ( $d_{ij,k}$ ) mediante una constante aditiva. Las distancias calculadas para cada sujeto se convierten luego en productos escalares  $b_{ij,k}$  tales que:

$$b_{ij,k} = \sum_{a=1}^r y_{ia,k} y_{ja,k}$$

Donde

$$y_{ia,k} = \sqrt{w_{ka} x_{ia}}$$

Con lo que la ecuación anterior nos queda

$$b_{ij,k} = \sum_{a=1}^r w_{ka} x_{ia} x_{ja}$$

Esta ecuación puede considerarse como un caso especial del modelo CANDECOMP (canónica decomposition para la descomposición de tablas de N vías). El modelo descompone una tabla de 3 vías y 3 modos en un conjunto de parámetros para cada vía que se combinan de forma multiplicativa para cada dimensión a y de forma aditiva para el total de dimensiones. En el caso de INDSCAL la segunda y tercera vía han de ser idénticas pues se refiere al mismo conjunto de estímulos.

La utilización del modelo permite la estimación de los valores de los productos escalares mediante regresión lineal utilizando un algoritmo especial llamado Mínimos Cuadrados Alternantes (ALS). El algoritmo procede a estimar los valores de los parámetros  $w_{ka}, x_{ia}, x_{ja}$  por mínimos cuadrados, manteniendo uno de ellos fijos y los otros dos libres de forma alternante. Cuando transcurridas una serie de iteraciones, el ajuste entre los datos y la solución es satisfactorio se fija el mismo valor para la segunda y tercera vías y se estima el valor de la primera vía (representada por los  $w_{ka}$ ).

La salida ofrecida por el modelo se da en dos matrices distintas

1. Una matriz de coordenadas  $X(n \times r)$  semejante a las proporcionadas por el modelo métrico o no métrico. Esta matriz representa el espacio de los estímulos para el total de los sujetos
2. Una matriz de pesos  $W(m \times r)$  que contienen los pesos otorgados por cada uno de los  $m$  sujetos a cada una de las  $r$  dimensiones. Esta matriz representa el espacio de estímulos y espacio de sujetos debe tomarse en el sentido de que son dos espacios distintos por lo que no es posible representar ambos en un único gráfico.

El espacio de los individuos tiene una serie de propiedades interesantes o para la interpretación de la solución proporcionada por INDSCAL

- Si elevamos al cuadrado el peso otorgado por un sujeto a una dimensión determinada, el valor obtenido se corresponde con la proporción de varianza en los datos del sujeto que es explicada por esa dimensión
- Si sumamos todos los pesos al cuadrado para un mismo sujeto el valor obtenido es la proporción de varianza en los datos del sujeto que es explicada por la solución proporcionada por INDSCAL es decir este valor es igual a  $RSQ$  para ese sujeto.

Es posible obtener en ocasiones valores negativos para los pesos de la matriz  $w$ , dado que solo se permiten pesos positivos, la presencia de los valores negativos puede indicar un mal ajuste del modelo a los datos. No obstante si los valores son muy pequeños pueden tomarse simplemente como aproximaciones a un valor cero de peso. En este último caso no existen ningún problema de ajuste.

A partir de las dos matrices anteriores es posible recuperar el espacio de estímulos individual para cada uno de los sujetos (espacio – sujeto). Esto se consigue simplemente multiplicando cada coordenada del espacio de estímulos total por la raíz cuadrada del peso asignado por el sujeto a esa dimensión, las nuevas coordenadas muestran el espacio del grupo encogido en aquellas dimensiones que resultan ser menos relevantes para el individuo.

Una razón por la que el modelo INDSCAL resulta tan útil en la practica es que las dimensiones no son arbitrarias sino que su orientación es siempre significativa o interpretable, a diferencia de lo que ocurre en modelos de escalamiento en dos vías, estas dimensiones no son rotables dado que el peso se otorgan es específicamente a la dimensión y no a ninguna otra, esto facilita mucho la interpretación de las soluciones aunque se tengan mas de dos dimensiones.

### **3.5 ESCALAMIENTO DESDOBLADO (UNFOLDIG)**

Todos los modelos de escalamiento que hemos visto hasta ahora trabajaban con matrices cuadradas, con el mismo número de estímulos en filas y columnas, también hemos utilizado matrices cuadradas al aplicar el modelo INDSCAL, aunque las tratamos como una como una única matriz en tres vías y dos modos las matrices cuadradas especialmente constituyen el tipo de entradas mas común en MDS. Sin embargo existen modelos de MDS para matrices de datos que no son cuadrados.

La característica fundamental de una matriz rectangular es que las entidades representadas en las filas y las columnas (generalmente sujetos y estímulos) son diferentes. Por tanto un análisis MDS de una matriz rectangular deberá representar conjuntamente ambas entidades, esto representa una propiedad sumamente interesante de este tipo de modelos. Recordemos que INDSCAL representa un espacio para los sujetos y otro para los estímulos, pero no ambos conjuntamente

Los datos de entrada para este tipo de MDS suelen ser puntuaciones de preferencia otorgados por un grupo de sujetos para un conjunto de estímulos aunque también es posible utilizar otro tipo de puntuaciones. Dos son los modelos MDS más utilizados para matrices rectangulares:

El modelo vectorial. En este una de las entidades (generalmente los estímulos) se representan como puntos en el espacio, mientras que la otra (generalmente los sujetos) se representa como vectores en ese mismo espacio. De este modo, la proyección de las posiciones de los estímulos sobre el vector de un sujeto (cuyo extremo indica la máxima preferencia) deberá reflejar las preferencias de este sujeto.

En el modelo desdoblado, tanto los sujetos como los estímulos se representan como puntos. Los puntos que representan a los sujetos en el espacio de solución indican la zona donde se encontraría la máxima preferencia de cada sujeto de tal modo que a medida que nos alejamos de uno de esos puntos en cualquiera de las direcciones, la preferencia va disminuyendo. Si alguno de los estímulos está próximo a un punto ese estímulo es el ideal para el sujeto representado por el punto ideal.

Este modelo tiene su origen en el modelo unfolding de Coombs que fue generalizado para el caso multidimensional por Bennet y Hays y Coombs asumiendo un nivel de medida ordinal (no métrico). Actualmente el análisis puede llevarse a cabo también con los niveles de medida superiores.

La matriz de entrada de un análisis desdoblado es una matriz rectangular de preferencia en dos vías y dos modos (generalmente sujetos X estímulos) donde cada entrada  $p_{ij}$  de la matriz corresponde a la preferencia expresada por el sujeto  $i$  por el estímulo  $j$ . dado que cada fila de la matriz contienen las puntuaciones de preferencia de una fuente de datos distinta (generalmente un sujeto) es habitual suponer aquí que los datos de cada fila son condicionales. El análisis entiende que esta matriz rectangular es un trozo de la diagonal de una matriz de proximidades incompleta, esto implica que el análisis da por perdida gran cantidad de la información contenida en la matriz de proximidades, por lo que este modelo es mas propenso a no converger o soluciones degeneradas.

El modelo asume que al proximidad del estímulo  $j$  al punto ideal del sujeto  $i$  (es una función de la preferencia del sujeto  $i$ -ésimo por el estímulo  $j$ -ésimo ( $p_{ij}$ ). Formalmente

$$\pi_{ij} = f(p_{ij}) = d_{ij}^2$$

Donde

$$d_{ij}^2 = \sum_{a=1}^r (y_{ia} - x_{ja})^2$$

Siendo  $y_{ia}$  la coordenada del sujeto  $i$ -ésimo en la dimensión  $n$ -ésima y  $x_{ja}$  la coordenada del estímulo  $j$ -ésimo en la dimensión  $n$ -ésima. Ambas coordenadas son estimadas por el algoritmo

La función  $f$  puede ser lineal (caso métrico) o monótona caso no métrico

## **PROGRAMA ALSCAL**

ALSCAL es un algoritmo que permite realizar el escalamiento multidimensional sus excelentes resultados prácticos han hecho que sea implementado en SAS y en SPSS entre otros, de ahí que su utilización se haya difundido siendo uno de los modelos más empleados.

ALSCAL (Alternating Least Square Scaling) fue desarrollado por Takane, Young y DE LEew (1997) como un programa de MDS enteramente basado en el algoritmo de mínimos cuadrados alternantes, tal y como fue modificado por Young, T Akane y Lewyckyj (1978). Para obtener una solución a partir de los datos, el programa opera en forma iterativa, donde cada iteración consta de dos pasos.

1. Fase de escalamiento óptimo. En está fase se transforman los datos de modo que se ajusten las distancias cuadráticas estimadas por el modelo.
2. Fase de estimación de estimación de modelo. En esta fase se estiman los parámetros del modelo de tal modo que las distancias cuadráticas estimadas por el mismo se ajusten a los datos óptimamente escalados

Dado que en cada uno de estos pasos el ajuste se lleva acabo mediante una aproximación por mínimos cuadrados, el algoritmo resultante se llama de mínimos cuadrados alternantes.

ALSCAL es uno de los programas más versátiles de los que se pueden disponer para llevar a cabo análisis MDS. Contiene un gran número de opciones que nos permiten:

1. Analizar datos en escala nominal, ordinal, de intervalo o de razón
2. Utilizar matrices completas o con datos perdidos
3. Utilizar matrices de proximidad simétricas o asimétricas
4. Utilizar datos condicionales o incondicionales
5. Aplicar los modelos con replicación o sin replicación
6. Utilizar datos de tipo continuo o discreto
7. Aplicar gran variedad de modelos de escalamiento diferentes( métrico, no métrico, ponderado (INDSCAL), desdoblado,

Es por eso que se utiliza en este trabajo para poder explorar todas las posibilidades que este programa nos da para enriquecer el análisis de la información en el sistema de inteligencia tecnológica. El detalle de algoritmo se presenta con mayor detalle en el Anexo

## CAPÍTULO 4

### CASO DE APLICACIÓN “DESALACIÓN DE AGUA DE MAR”

**OBJETIVO:** Este capítulo muestra el caso de aplicación de la inteligencia tecnológica desarrollado por el Instituto de Ingeniería de la UNAM sobre el tema de desalación de agua de mar. Mostrando el calculo de varios indicadores cuantitativos, la creación de la matriz de co-ocurrencia y la creación de los mapas tecnológicos en el tema de desalación de agua entregados al grupo de expertos.

### INTRODUCCIÓN

La aplicación del Sistema de Inteligencia Tecnológica por parte del Instituto de Ingeniería de la UNAM, se realizó sobre el tema de Desalación de Agua de Mar. Para ello se siguió la metodología explicada en los capítulos anteriores, así llevando a cabo las diferentes etapas de la operación de SIT.

La primera etapa relacionada con la recolecta de la información referente al tema de desalación de agua de mar se realizó considerando los artículos científicos contenidos en la página de ISI Web of Knowledge.

En las búsquedas realizadas con en la base de ISI Web of Knowledge se utilizaron como términos de búsqueda el nombre de publicaciones, el nombre de instituciones y las palabras clave que proporcionó el GESIT en la etapa de implementación del sistema.

Como resultado de estas búsquedas se encontraron 1100 artículos relacionados con la desalación de agua de mar en el periodo de 2003 a 2007.

Para capturar la información previamente se realizó la depuración de los campos de palabras clave, autores y instituciones que se encontraron en los 1,100 artículos que tenían diferentes abreviaturas o errores de captura.

Como ejemplo la figura 5.1 muestra las abreviaturas corregidas del campo de palabra clave de los 1,100 artículos encontrados

**Tabla 5.1 ■ Abreviaturas corregidas del campo palabra clave**

Siglas	Concepto
CDM	Clean development mechanism
NPP	Nuclear power Plant
NF	Nanofiltracion
RO	Reverse Osmosis
AFM	Atomic force microscopy
MSF	Multi stage flash
SAR	Sodium Adsorption Radio
FTIR	Fourier transform infrared
AOP	Advanced oxidation process
DSC	Differential scanning calorimetry
EDCs	Endocrine disrupting chemicals
EPS	Extracellularpolymeric substances
SMP	Soluble microbial products
NHR	Nuclear heating reactor
HTR	High-temperature gas-cooled reactor
NOM	Natural organic matter
DSS	Duplex stainless steel
ICI	Interference color imaging technique
SCM	Surface corrosion potential mapping
EPS	Extracellular polymeric substances
FTT	Finite time thermodynamics
MSW	Municipal solid waste
MF	Microfiltration
AOC	Assimilable organic carbon
BAC	Bio-activated carbon
SDI	Silt density index
FO	enton oxidation
CFD	Computational fluid dynamics
PDMA	Pendant-drop mechanical analysis
MED	Multiple effect distillation
DBPs	Disinfection by-products
DOM	Dissolved organic matter
UF	ltrafiltration
VI	eparation ofchromium
DOM	Dissolved organic matter
SDI	Silt density index
MFI	Modified fouling index
PAC	Powdered activated charcoal

La figura 5.2 muestra las palabras clave en singular que se pasaron a plural para tener un solo término que evidentemente hacen alusión a un mismo concepto.

**Tabla 5.2 ■ Palabras clave que se cambiaron de singular a plural**

Palabra clave	Plabra clave sorregida
Anthocyanin	anthocyanins
brackish water	brackish waters
membranes processes	membrane processe
advanced oxidation process	advanced oxidation processes
antiscalant	antiscalants
brine discharge	brine discharges
brine	brines
chemical treatment	chemical treatments
chloramine	chloramines
coagulant	coagulants
colloid	colloids
composite membranes	composite membraness
cost	costs
diffusion coefficient	diffusion coefficients
dual-purpose plant	dual-purpose plants
economic	economics
effluent	effluents
environmental	environments

Después de la depuración de los campos se capturó la información de los 1,100 artículos en el Sistema de Cómputo de Inteligencia (SCIT) desarrollado por el Instituto de Ingeniería de la UNAM, que es la infraestructura computacional que da soporte a los procesos de captura y análisis de la información involucrados en la fase de operación del Sistema de Inteligencia Tecnológica, el cual contiene módulos de captura, y módulos para crear los indicadores de actividad y el análisis de palabras asociadas necesarias para el procesamiento de la información con el objetivo de convertirlo en inteligencia.

La figura 5.0 muestra la interfaz del Sistema de Cómputo de Inteligencia (SCIT) utilizado en el estudio de caso.

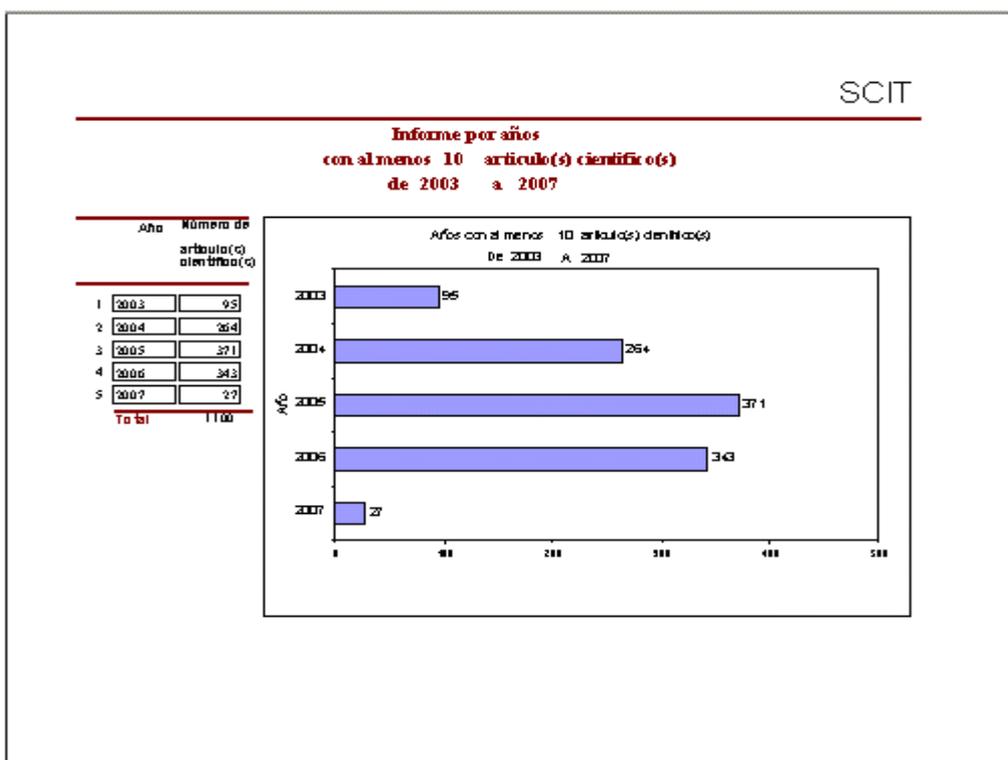
**Figura 5.0 ■ Interfaz del Sistema de Computó de Inteligencia Tecnológica (SCIT)**



Posteriormente se realizó la segunda etapa relacionada con el procesamiento de la información recopilada, con ello se generaron los informes de conteo de diferentes campos que fueron enviados al grupo de expertos (GESIT) para su análisis.

La figura 5.1 muestra el informe por año, en el cual se puede observar que la producción de artículos sobre desalación de agua de mar ha aumentado en los últimos años. En 2003 se publicaron 95 artículos y en 2006 se publicaron 343; por lo que el tema ha crecido en relevancia.

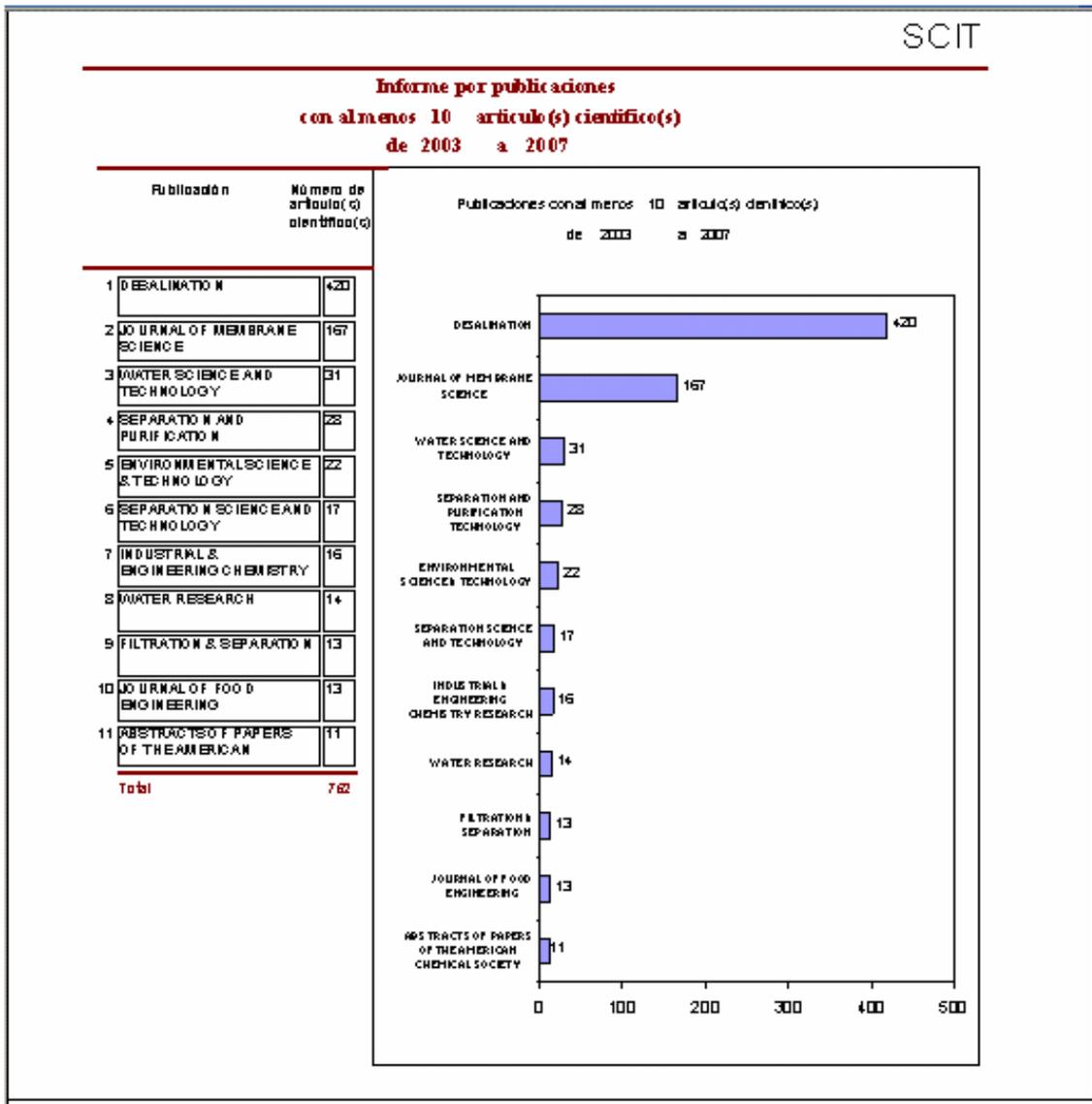
Figura 5.1 ■ Informe de por años



Fuente: Elaboración Propia a partir de SCIT con datos de la base de datos ISI Web of Knowledge.

La figura 5.2 muestra el informe de publicaciones, en el que se identifica que la principal publicación con documentos sobre desalación de agua de mar es la publicación de Desalination. El informe también muestra que existen otras 11 publicaciones en las que se han escrito al menos 10 artículos científicos sobre desalación de agua de mar en el periodo de 2003 a 2007.

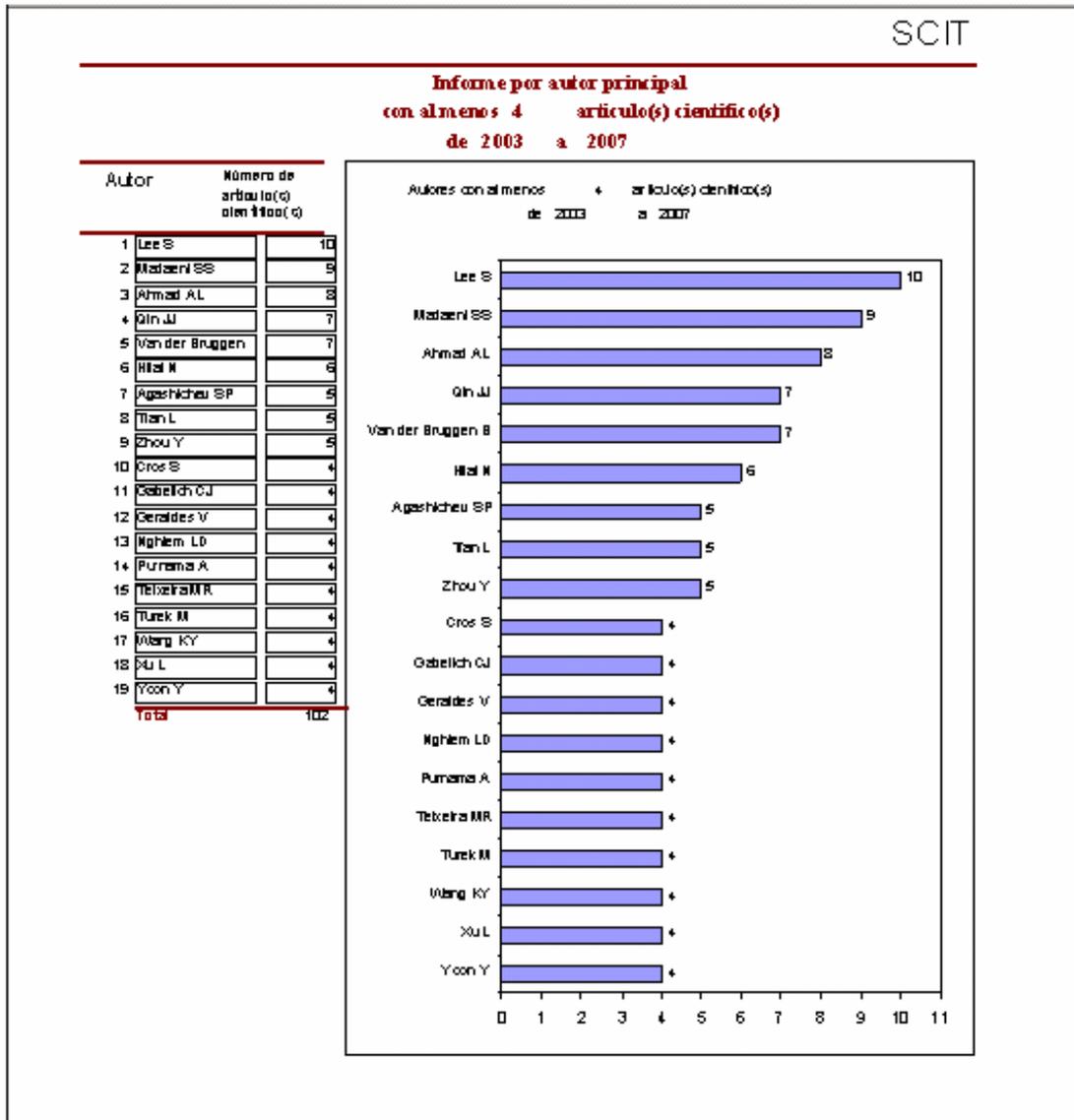
Figura 5.2 ■ Informe de Publicaciones



Fuente: Elaboración Propia a partir de SCIT con datos de la base de datos ISI Web of Knowledge.

La figura 5.3 muestra el informe de autor, en el que se identifica que autor más prolífico en el tema de desalación de agua de mar es el autor Lee S, además se muestra que por lo menos hay 19 autores que han publicados al menos 4 artículos sobre desalación de agua de mar en el periodo de 2003 a 2007.

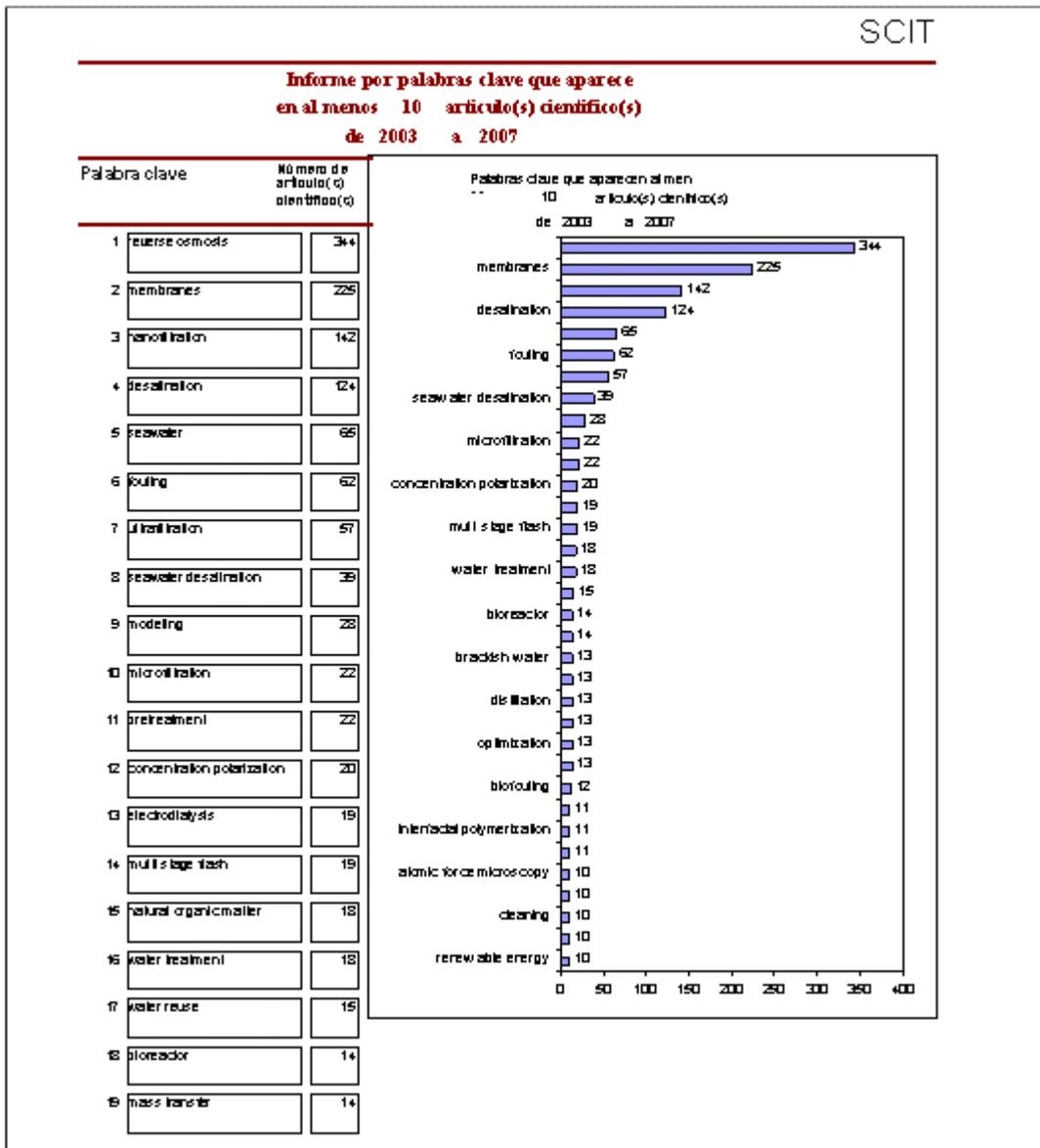
Figura 5.3 ■ Informe de autor principal



Fuente: Elaboración Propia a partir de SCIT con datos de la base de datos ISI Web of Knowledge.

En la figura 5.4 se muestra el informe por palabra clave, en el que se observa que la palabra clave más utilizadas en los artículos científicos son Reverse Osmosis, membranes, nanofiltration, además se observa que empieza usarse palabras clave como renewable energy en más de 10 artículos relacionados con desalación de agua de mar.

Figura 5.4 ■ Informe de palabra clave



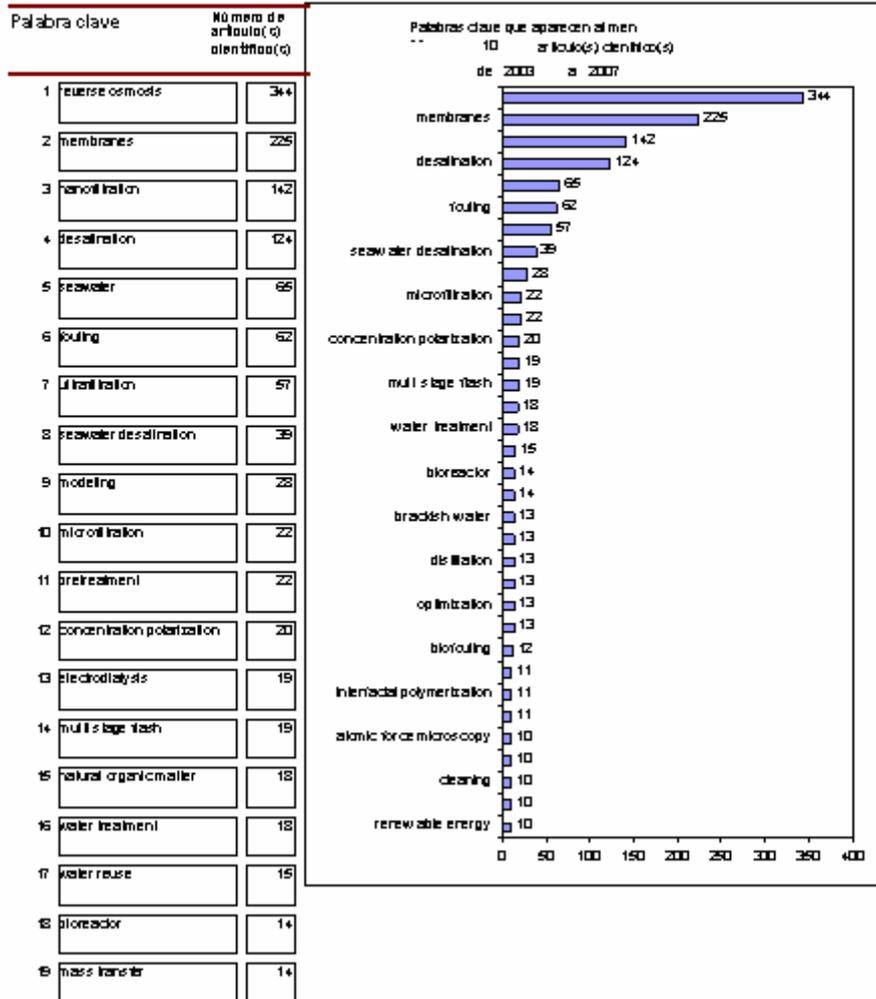
Fuente: Elaboración Propia a partir de SCIT con datos de la base de datos ISI Web of Knowledge.

En la figura 5.5 se muestra el informe por instituciones, en el que se observa que la institución asociada a las autores con más artículos publicados sobre desalación de agua de mar son Nation Univ Singapore, además se observa que hay otras 8 instituciones relacionadas con al menos 10 artículos científicos sobre desalación de agua de mar.

Figura 5.5 ■ Informe de instituciones

SCIT

**Informe por palabras clave que aparece  
en al menos 10 artículo(s) científico(s)  
de 2003 a 2007**



Fuente: Elaboración Propia a partir de SCIT con datos de la base de datos ISI Web of Knowledge.

Estos reportes fueron entregados al grupo de expertos (GESIT) para así seguir con el ciclo del Sistema de Inteligencia Tecnológica

## **Análisis de palabras asociadas**

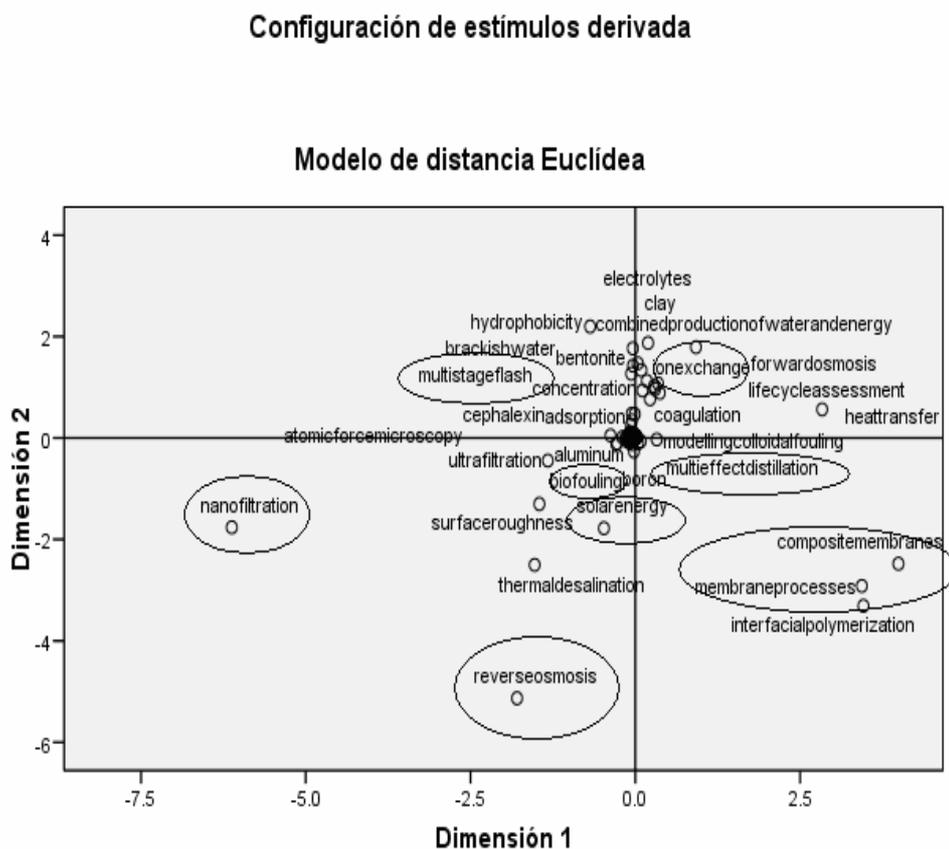
Asímismo para la etapa de procesamiento de la información se realizó un análisis de palabras asociadas de diferentes campos bibliográficos de la información capturada en el SCIT

Para llevar a cabo el análisis de palabras asociadas relacionadas a las palabras clave, se generó en el SCIT, la matriz de asociación de las palabras clave que tenían mayor frecuencia de aparición, la cual se muestra en la figura 5.6



Con la matriz normalizada se utilizó el método de escalamiento multidimensional del software estadístico SPSS así obteniendo el mapa tecnológico del tema desalación de agua de mar basado en las relaciones de las palabras claves obtenidas de los 1100 artículos que se encontraron relacionados al tema.

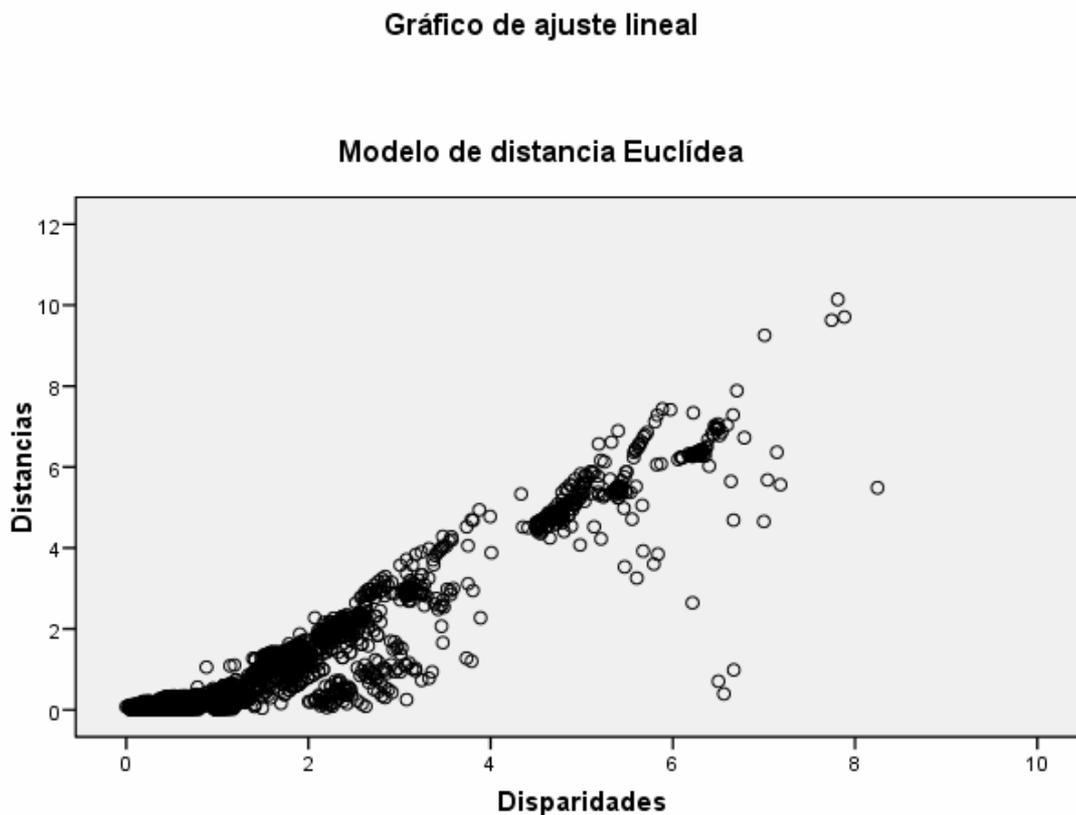
**Figura 5.6 ■ Mapa tecnológico de las palabras claves**



Dado que para esta representación el STRESS = .27413 Y RSQ = .93126 damos como valido el ajuste

La figura 5.7 muestra el ajuste lineal entre las disparidades a partir de nuestros datos y las distancias a partir de la representación de dos dimensiones arrojada por el modelo, el cual muestra un ajuste regularmente bueno

**Figura 5.7 ■ Ajuste lineal de modelo de palabras clave**



El mapa de la figura 5.6 es entregado al GESIT para que se analice junto con los reportes anteriores informes para dar seguimiento a la etapa de generación de conocimiento del Sistema de inteligencia.

Interpretación del Mapa tecnológico basado en las palabras clave.

Con el análisis del mapa tecnológico de las palabras clave realizado para el tema de desalación de agua de mar fue posible identificar las principales técnicas de desalación vigentes en el periodo de 2003 a 2007 las cuales son:

- Multi Effect Distillation
- Multi Stage Flash
- Osmosis inversa

También se puede observar que ligado al tratamiento de agua para poder utilizar los métodos de desalación de agua de mar está el tema de la contaminación del agua por los tratamientos previos a la desalación expresado en palabras clave como "*biofouling*".

Además, se pudo observar que el tema de las membranas está relacionado con la desalación de agua de mar ya que algunas de las técnicas de desalación utilizan membranas para llevar a cabo el proceso de desalinación.

Un tema que se observa relacionado con la desalación de agua de mar es la fuente de energía que utilizan las diferentes técnicas de desalación, estas expresadas por las palabras clave como "*solar energy*" y *renewableenergy*.

Adicionalmente se identificó como un tema de gran peso en el proceso de desalación de agua de mar el de la "*nonofiltración*" que por sí solo puede ser un tema de investigación futura.

Estos resultados junto con el análisis detallado del GESIT ayudan a la generación de conocimiento que es el objetivo de Sistema de Inteligencia Tecnológico.

También se elaboro se realizó un análisis de palabras asociadas de diferentes campo de autor con las palabras claves con la finalidad de identificar de manera más clara los temas que cada autor trabaja en mayor medida y así facilitar su seguimiento en futuras investigaciones en un tema específico.

Para llevar a cabo este análisis se generó en el SCIT la matriz de asociación de los 10 autores con más documentos publicados y las palabras clave la cual se muestra en la figura 5.8

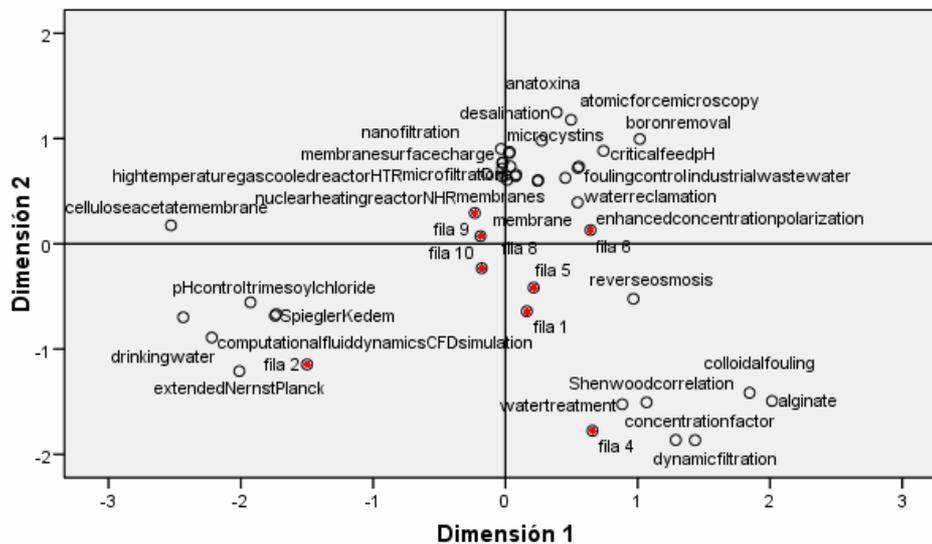
Con esta matriz se utilizó el método escalamiento desdoblado, para obtener el mapa tecnológico de autores con relación palabras clave, con el objetivo de identificar los campos que son más cercanos a los autores con más publicaciones, y así poder facilitar el seguimiento del trabajo de estos autores.

**Tabla 5.8 ■ Matriz de asociación 10 Autores con más artículos y las palabras clave**

PC	Agashichev SP	Ahmad AL	Hilal N	Lee S	Madaeni SS	Qin JJ	Teixeira MR	Tian L	Van der Bruggen B	Zhou Y
alginate	0	0	0	1	0	0	0	0	0	0
anatoxin-a	0	0	0	0	0	0	1	0	0	0
atomic force microscopy	0	0	1	0	0	0	0	0	0	0
boron removal	0	0	0	0	0	1	0	0	0	0
cellulose acetate membrane	0	1	0	0	0	0	0	0	0	0
colloidal fouling	0	0	0	1	0	0	0	0	0	0
computational fluid dynamics (CFD) simulation	0	1	0	0	0	0	0	0	0	0
concentration factor	0	0	0	1	0	0	0	0	0	0
critical feed pH	0	0	0	0	0	1	0	0	0	0
desalination	0	0	0	0	0	0	0	0	0	1
drinking water	0	1	0	0	0	0	0	0	0	0
dynamic filtration	0	0	0	1	0	0	0	0	0	0
enhanced concentration polarization	1	0	0	0	0	0	0	0	0	0
extended Nernst-Planck	0	1	0	0	0	0	0	0	0	0
fouling control	0	0	0	0	0	1	0	0	0	0
high-temperature gas-cooled reactor (HTR)	0	0	0	0	0	0	0	1	0	0
industrial wastewater	0	0	0	0	0	1	0	0	0	0
membrane	0	0	1	0	1	0	0	0	0	0
membrane surface charge	0	0	0	0	0	0	1	0	0	0
membranes	0	0	1	0	1	0	0	0	0	0
microcystins	0	0	0	0	0	0	1	0	0	0
microfiltration	0	0	0	0	0	0	0	0	1	0
nanofiltration	0	0	1	0	0	0	1	0	1	0
nuclear heating reactor (NHR)	0	0	0	0	0	0	0	1	0	0
pH control	0	1	0	0	0	0	0	0	0	0
polyamide	0	0	0	0	0	0	0	0	0	1
polyamide-urethane	0	0	0	0	0	0	0	0	0	1
pressure-driven membrane processes	0	0	0	0	0	0	0	0	1	0
reverse osmosis	1	0	0	1	1	1	0	0	0	0
seawater desalination	0	0	0	0	0	0	0	1	0	0
Shenwood correlation	0	0	0	1	0	0	0	0	0	0
Spiegler-Kedem	0	1	0	0	0	0	0	0	0	0
thin-film composite membrane	0	0	0	0	0	0	0	0	0	1
trimesoyl chloride	0	1	0	0	0	0	0	0	0	0
water reclamation	0	0	0	0	0	1	0	0	0	0
water treatment	0	0	0	1	0	0	0	0	0	0

## Configuración de estímulos derivada

### Modelo de distancia Euclídea de difs. individuales (ponderadas)



Interpretación del Mapa tecnológico de autores en relación con palabras clave.

Con el análisis del mapa tecnológico de los autores con relación a las palabras clave realizado fue posible identificar los temas de desalación de agua de mar más relacionados con los 10 autores con mayor número de artículos científicos publicados en el periodo de 2003-2007:

Lee S es un autor relacionado con el punto fila 4 que claramente tiene una cercanía con los temas (*Dynamic filtration, water treatment, concentration factor*)

Agashichev SP , Madaeni SS y Qin JJ relacionados con los puntos fila 1 fila 5 y fila 6 son autores que tienen una mayor cercanía a los temas de (*reverse osmosis, membranas, enhanced concentration polarization*)

Ahmad Al autor relacionado con el punto fila 2 claramente tiene relación con otros tópicos de la desalación de agua de mar entre los más cercanos están los

de simulación de fluidos representado por la palabra clave “*computational fluid dynamics CFD simulation*”

Van der Bruggen y Zhou Y tienen una cercanía mayor a los temas de membranas que esta mas identificado en el cuadrante superior izquierdo del mapa tecnológico.

## **CONCLUSIONES**

La necesidad de información para respaldar la toma de decisiones en materia técnico- científica ha facilitado el desarrollo de nuevos conceptos como los de, inteligencia competitiva y inteligencia tecnológica.

El concepto de inteligencia competitiva y en particular la inteligencia tecnológica dan respuesta a la necesidad de información dando un valor adicional en el análisis de la información generando conocimiento que faciliten la toma de decisiones en temas tecnológicos-científicos.

El Sistema de Inteligencia Tecnológica (SIT) desarrollado en el IIUNAM propone una manera de realizar Inteligencia tecnológica de manera sistematizada basada en la gestión de calidad la cual documenta cada uno de los procesos necesarios para la inteligencia tecnológica.

El SIT propone un proceso de dos fases, los cuales son implantación y operación que dan sustento metodológico a etapas y tareas necesarias par realizar inteligencia tecnológica.

Una vez desarrollada la fase de implementación, y para poder llevar a cabo la fase de operación del SIT, es necesario llevar a cabo las etapas de *Recolecta de la información* y *Procesamiento de la información*.

La etapa *de procesamiento de la información* se centra en el análisis analítico de la información recolectada, y su conversión en inteligencia. Para ello se hace uso del concepto de Cienciometría para obtener indicadores que permiten ver las tendencias y relaciones en el área técnico – científica analizada. Los indicadores de actividad de la Cienciometría proporcionan información sobre los autores y las instituciones más relevantes en el tema que han venido desarrollando con mayor frecuencia trabajos sobre un tema. El análisis de palabras asociadas de la Cienciometría muestra las relaciones que existen entre los diferentes conceptos. Con la creación de la matriz de distancias entre conceptos, se muestra las relaciones fuerte y débiles de un tema técnico - científico.

El uso del método de escalamiento multidimensional facilita la transformación de la matriz de palabras-asociadas en mapas tecnológicos que revelan la manera en que relacionan los conceptos de un tema. La utilización de estos mapas brinda al experto un criterio más para lograr un mejor entendimiento sobre las interrelaciones de diferentes campos de investigación y de los principales actores en los mismos, la evolución de frentes de investigación, así como el impacto de intervenciones externas. Esto permite hacer recomendaciones más acertadas, ya sea en política científica o de información.

El desarrollo del Sistema de Cómputo de Inteligencia Tecnológica (SCIT) facilita la captura de la información, la creación de indicadores de actividad así como la creación de las matrices de distancias a partir del análisis de palabras asociadas necesarias para poder elaborar posteriormente los mapas tecnológicos.

Con el SIT y el SCIT se cuenta con herramientas para facilitar la transformación de información en inteligencia que a su vez apoya la toma de decisión.

Es importante dejar en claro que para realizar Inteligencia Tecnológica se requiere la necesidad de que los resultados obtenidos en el procesamiento de la información sean analizados por un grupo de expertos (GESIT) en el tema que se va a investigar, capaz de valorar adecuadamente los resultados e

interpretar los mapas de forma correcta. Ya que el análisis realizado por una persona que carezca del conocimiento suficiente sobre el tema obtendrá resultados poco fiables.

Con este trabajo se da un paso importante en la operación de la Inteligencia Tecnológica, por lo que no pretende ser un trabajo definitivo si no un punto de arranque para profundizar en el desarrollo de las herramientas que faciliten las tareas necesarias para realizar Inteligencia Tecnológica.

Una de las líneas de investigación futura que se plantean de este trabajo, consiste en la experimentación de técnicas complementarias que ya han sido aplicadas a la creación de mapas tecnológicos, como por ejemplo las redes neuronales.

Así mismo, se plantea la necesidad de seguir trabajando en el SCIT para mejorar las etapas de captura y procesamiento de la información, creando indicadores que muestren de una mejor manera las tendencias de un tema científico-técnico que faciliten la toma de decisiones.

## **ANEXO**

### **ALGORITMO ALSCAL**

El algoritmo ALSCAL, como todos los diseñados para resolver el problema que se plantea el MDS es iterativo.

Una vez iniciada la configuración inicial, se inicia el algoritmo de optimización en la que se distinguen dos fases:

Algoritmo de optimización

Fase 1. (Escalamiento óptimo). Se transforman los datos de modo que se ajustan las distancias cuadráticas estimadas por el modelo.

Fase 2 (Estimación del modelo). Se obtienen las dos variables fundamentales que son las coordenadas de los estímulos y los pesos de los individuos, en el caso de que se trate de un escalamiento multidimensional ponderado. De tal modo que las distancias cuadráticas estimadas se ajusten a los datos óptimamente escalados.

Estos dos pasos se van alternando hasta que se llega a una solución lo suficientemente precisa. Dado que, como se verá, cada paso es un procedimiento de mínimos cuadrados, el algoritmo completo se llama Alternating Least Squares Scaling (ALSCAL).

## Configuración Inicial

La obtención de la configuración inicial implica la realización de varios pasos que se detallan a continuación.

**C.1)** Partiendo de la configuración inicial de estímulos o distancias  $O_k$  entre los diferentes estímulos del sujeto  $k$  se construye una nueva matriz de proximidades transformadas  $O_k^*$  añadiendo una constante  $c_k$  a dichas medidas de proximidad. Así los elementos de la nueva matriz  $O_k^*$  son de la forma

$$O_{ijk}^* = O_{ijk} + c_k$$

De tal forma que cumpla la desigualdad triangular

$$O_{ijk}^* + O_{jlk}^* \geq O_{ilk}^*$$

Y se mantenga lo no negatividad de la distancia

$$O_{ijk}^* \geq 0$$

Donde

$O_{ijk}^*$  es la mediada de proximidad ajustada entre los estímulos  $i$  y  $j$  del sujeto  $k$

$O_{jlk}^*$  es la medida de proximidad ajustada entre los estímulos  $j$  y el  $l$  del sujeto  $k$

$O_{ilk}^*$  es la medida de proximidad ajustada entre los estímulos  $i$  y el  $l$  del sujeto  $k$ .

La constante  $c$  ha de ser tan pequeña como sea posible, este paso es necesario para lograr que la matriz  $B_k^*$  descrita posteriormente sea semidefinida positiva

**C.2)** Se calcula una matriz  $B_k^{**}$  con el doble centrado  $O_k^*$  de la matriz de estímulos ajustada para cada sujeto. Cada elemento de la matriz se calcula como:

$$b_{ijk}^{**} = -\frac{1}{2}(O_{ijk}^{*2} - O_{i.k}^{*2} - O_{.jk}^{*2} + O_{.k}^{*2})$$

Donde

$O_{i.k}^*$  Son las medias de los renglones para la proximidades ajustadas para el sujeto k

$O_{.k}^*$  Son las medias de las columnas para la proximidades ajustada para el sujeto k

$O_{.k}^*$  Es la media para el sujeto k

El centrado doble es necesario para convertir distancias en productos escalares el cual es requerido para calcular una configuración inicial que utiliza el procedimiento de Young-Householder-Torgeson

**C.3)** Las matrices individuales de cada sujeto son normalizadas esto para que tengan la misma varianza dando lugar a la matriz  $B_k^*$ . Los elementos de esta matriz son:

$$b_{ijk}^* = \frac{b_{ijk}^{**}}{\left[ \frac{\sum_i \sum_j (b_{ijk}^{**})^2}{(n)(n-1)} \right]^{1/2}}$$

Donde

$n$  = número de estímulos

$(n)(n-1)$  = número de elementos fuera de la diagonal en la matriz  $B^{**}$

El denominador es la desviación estándar de los productos escalares de la matriz  $B^{**}$ .  $B_k^*$  es una matriz con elementos  $b_{ijk}^*$  que son los productos escalares para el sujeto k'. Es así que la normalización de las matrices de los sujetos compara la contribución de cada uno de forma individual a la formación de una matriz de los productos escalares promedio.

**C.4)** Se calcula una matriz promedio  $B^*$  para el conjunto de  $m$  individuos  $B^*$ . Los elementos de la matriz son

$$b_{ij.}^* = \frac{\sum_k b_{ijk}^*}{m}$$

Donde

$m$  = es el número de sujetos

**C.5)** La matriz  $B^*$  es utilizada para calcular la matriz  $X$  que cumpla que:

$$B^* = XX'$$

Donde

$X$  es una matriz de  $n$  estímulos y  $r$  dimensiones

$X'$  es la transpuesta de la matriz  $X$ .

La matriz  $X$  **es la configuración inicial**

**C.6)** Para el ALSCAL ponderado se calculan matrices iniciales de los pesos  $W_k$  para cada sujeto. Las matrices iniciales  $W_k$  son matrices de  $r \times r$  donde  $r$  es el número de dimensiones. Después las diagonales de  $W_k$  formaran los renglones de la matriz  $W$  que es una matriz de  $n \times r$ .

Las matrices  $W_k$  son determinadas como:

$$B_k^* = YW_kY'$$

Donde

$$Y = XT \text{ y } TT = I$$

Donde  $T$  es una rotación ortogonal de la configuración  $X$  para una nueva orientación  $Y$ .  $T$  se calcula por el procedimiento de Schönemann-de Leeuw discutido, Takane, y Lewyckyj (1978).

$T$  gira a  $X$  de modo que  $W_k$  es tan diagonal como sea posible (es decir, con elementos fuera de la diagonal sean tan cercanos a cero como sea posible). Esto es que los elementos fuera de diagonal diagonales salen del modelo (se asume que los pesos de los sujetos solamente afectan las dimensiones del espacio del estímulo).

## Algoritmo de optimización

El algoritmo de optimización consta, como ya se ha indicado, de dos fases:

### Fase 1. Escalamiento Óptimo

En la fase de escalamiento óptimo se realizan los siguientes pasos:

#### F1.1.- Calculo de distancias.

Con las matrices  $X$  y  $W$  obtenidas, se calculan las distancias euclidianas ponderadas el siguiente modo:

$$d_{ijk}^2 = \sum_{a=1}^r w_{ka} (x_{ia} - x_{ja})^2$$

Donde:

$w_{ka}$  es el peso que le asigna el sujeto  $k$  a la dimensión  $a$

$x_{ia}$  es la coordenadas del estímulo  $i$  en la dimensión  $a$

$x_{ja}$  es la coordenada del estímulo  $j$  en la dimensión  $a$

#### F1.2. Transformación monótona.

Como segundo paso para obtener el escalamiento óptimo, se utiliza la transformación monótona de mínimos cuadrados de Kruskal para datos ordinales. Este procedimiento permite obtener la matriz de disparidades que son una transformación monótona de los datos que deben parecerse tanto como se a posible (en términos de mínimos cuadrados) a la matriz de distancia previamente calculada.

Idealmente las distancias deberían ordenar los estímulos de la misma manera que los datos originales, pero como en la práctica no será así, obtendremos las disparidades que si están en el mismo orden que los datos y que se ajustarán tanto a los datos iniciales como sea posible.

### F1.3 Normalización

Las disparidades computadas previamente ahora se normalizan por las razones técnicas relacionadas con el algoritmo de alternancia de los mínimos cuadrados (véase Young, Leeuw, y Takane, 1979). Durante el curso del proceso de la optimización, queremos reducir al mínimo una medida de error llamado S STRESS. Pero el procedimiento monótono de la regresión descrito arriba reduce al mínimo solamente el numerador de la fórmula de SSTRESS. Así, la fórmula de abajo se aplica para reajustar la longitud de las disparidades del vector para reducir al mínimo el SSTRESS:

$$D_k^{*N} = D_k^* (D_k' D_k) (D_k' D_k^*)^{-1}$$

Donde:

$D_k^*$  es un vector columna con  $\frac{n(n-1)}{2}$  elementos conteniendo todas las disparidades para el sujeto K

$D_k$  es un vector columna con  $\frac{n(n-1)}{2}$  elementos conteniendo todas las distancias para el sujeto k

$D_k' D_k$  Es la suma de las distancias al cuadrado

$D_k' D_k^*$  Es la suma de productos cruzados

El vector de las disparidades normalizado  $D_k^{*N}$  es una estimación condicional de mínimos cuadrados para las distancias; es decir, es la estimación de mínimos cuadrados para una iteración dada. Los valores anteriores de  $D^*$  son substituidos por valores de  $D^{*N}$ , y los pasos subsecuentes utilizan las disparidades normalizadas.

## F1.4 Cálculo del STRESS

Se utiliza la fórmula de Takane- Young – de Leeuw

$$SSTRESS(1) = S = \left\{ \frac{1}{m} \sum_{k=1}^m \left[ \frac{\sum_i \sum_j (d_{ijk}^2 - d_{ijk}^{*2})^2}{\sum_i \sum_j d_{ijk}^{*4}} \right] \right\}^{1/2}$$

Donde:

$d_{ijk}^*$  son las medidas de la disparidad normalizadas computadas previamente, y el  $d_{ijk}$  se computa como se mostró también anteriormente. Así se computa SSTRESS de las disparidades normalizadas y del sistema anterior de coordenadas y de pesos.

## F1.5 Terminación

Al final de los 4 pasos se realiza el test de terminación para determinar si se ha obtenido o no la solución final

El valor actual de SSTRESS se compara al valor de SSTRESS de la iteración anterior. Si la mejora es menor que un valor, la iteración para si no, el programa procede al paso siguiente. (Este paso se salta en la primera iteración.)

## Fase 2. Estimación del modelo

En ALSCAL los pesos y las coordenadas no se pueden solucionar en forma simultánea, así que lo hacemos sucesivamente. Así, la fase de estimación del modelo engloba dos subfases de minimización separadas:

### F2.1 Estimación de los pesos de los sujetos

Este paso se salta para el modelo euclidiano no ponderado. En este paso una estimación condicional de los pesos se computa en cada iteración por mínimos cuadrados.

Se calculan los pesos de la configuración, supuestas constantes las disparidades y la configuración:

$$W \cong D^* P (P^* P)^{-1}$$

La derivación de la fórmula es como sigue: Hemos encontrado disparidades tales que

$$d_{ijk}^* \cong d_{ijk}^2$$

Donde

$$d_{ijk}^2 = \sum_{a=1}^r w_{ka} (x_{ia} - x_{ja})^2$$

Sea  $P_{ija}$  la distancia no ponderada entre los estímulos  $i$  y  $j$  según lo proyectado sobre la dimensión  $a$ , es decir:

$$P_{ija} = (x_{ia} - x_{ja})^2$$

Entonces

$$d_{ijk}^2 \cong d_{ijk}^2 = \sum_{a=1}^r w_{ka} P_{ija}$$

En notación matricial esto es expresado como

$$D^* = WP'$$

Donde

$D^*$  es ahora una matriz de  $m \times \frac{n(n-1)}{2}$  teniendo un renglón para cada sujeto y una columna para cada par de estímulos.

$W$  es una matriz  $m \times r$  teniendo un renglón para cada sujeto y una columna para cada par de estímulos.

$P'$  tiene un renglón para cada dimensión y una columna para cada par de estímulos

Se resuelve para  $W$

$$WP' \cong D^*$$

Como

$$WP'P(P'P)^{-1} = D^*P(P'P)^{-1}$$

Por lo tanto

$$W = D^*P(P'P)^{-1}$$

Con lo que tenemos la estimación condicional de los mínimos cuadrados para  $W$ . De hecho hemos reducido al mínimo SSTRESS a este punto concerniente a los valores previamente computados para las coordenadas del estímulo y el escalamiento óptimo. Substituimos los viejos pesos sujetos por los valores nuevamente estimados.

## **F2.2 Estimación de las coordenadas de lo estímulos,**

Suponiendo fijos los valores de los pesos estimados en la etapa anterior, así como las disparidades calculadas  $D^*$ .

Las coordenadas son determinados una a la vez reduciendo al mínimo el "SSTRESS" con respecto a una coordenada dada. La ecuación que permite que solucionemos para una coordenada  $x$  dada se deriva como sigue

$$S = SSTRESS = \left\{ \frac{1}{m} \sum_{k=1}^m \frac{\sum_i \sum_j (d_{ijk}^2 - d_{ijk}^{*2})^2}{\sum_i \sum_j d_{ijk}^{*4}} \right\}$$

El s-stress en relación con una coordenada dimensional,

Teniendo en cuenta que  $\sum_i \sum_j d_{ijk}^{*4}$  es una constante con respecto a  $x_{le}$ , podemos hacer:

$$c_k = \frac{1}{\sum_i \sum_j d_{ijk}^{*4}}$$

Por lo tanto,

$$S = \left[ \frac{1}{m} \sum_k c_k \sum_i \sum_j (d_{ijk}^2 - d_{ijk}^{*2})^2 \right] = \left[ \frac{1}{m} \sum_k c_k S_k \right]$$

Donde

$$S_k = \sum_i \sum_j (d_{ijk}^2 - d_{ijk}^{*2})^2$$

Hemos separado S en una suma de funciones  $S_k$  cada una de las cuales puede optimizarse por separado respecto a  $x_{le}$ .

Como

$$d_{ijk}^2 = \sum_a w_{ka} (x_{ia} - x_{ja})^2$$

Entonces

$$S_k = \sum_i \sum_j \left[ \sum_a w_{ka} (x_{ia} - x_{ja})^2 - d_{ijk}^{*2} \right]^2$$

Teniendo en cuenta que la coordenada respecto a la que se está optimizando es la e entonces:

$$S_k = \sum_i \sum_j \left[ \sum_{a \neq e} w_{ka} (x_{ia} - x_{ja})^2 + w_{ke} (x_{ie} - x_{je})^2 - d_{ijk}^{*2} \right]^2$$

Si definimos

$$a_{ijke}^2 = \frac{d_{ijk}^{*2} - \sum_{a \neq e} w_{ka} (x_{ia} - x_{ja})^2}{w_{ke}}$$

Entonces:

$$\begin{aligned} S_k &= w_{ke}^2 \sum_i \sum_j \left[ a_{ijke}^2 - (x_{ie} - x_{je})^2 \right]^2 \\ &= w_{ke}^2 \sum_i \sum_j \left( a_{ijke}^2 - x_{ie}^2 + 2x_{ie}x_{je} - x_{je}^2 \right)^2 \end{aligned}$$

Definiendo:

$$b_{ijke}^2 = a_{ijke}^2 - x_{je}^2$$

Que es una constante cuando estamos optimizando con respecto a la coordenada e, podemos expresar de la siguiente forma:

$$S_k = w_{ke}^2 \sum_i \sum_j \left( b_{ijke}^2 - x_{ie}^2 + 2x_{ie}x_{je} \right)^2$$

Ahora hemos aislado los términos constantes ( $b_{ijke}^2$ ) de los términos de la variable (" $x_{ie}$ ")

A continuación para optimizar, calculamos la derivada parcial de S respecto a la coordenada xie

$$\frac{\partial S}{\partial x_{ie}} = \frac{\partial}{\partial x_{ie}} \frac{1}{m} \sum c_k S_k$$

Así

$$\frac{\partial S}{\partial x_{ie}} = \frac{1}{m} \sum c_k \frac{\partial S_k}{\partial x_{ie}} \dots \dots \dots (1)$$

Donde

$$\frac{\partial S}{\partial x_{ie}} = \frac{\partial}{\partial x_{ie}} \left[ w_{ke}^2 \sum_i \sum_j \left[ b_{ijke}^2 - x_{ie}^2 + 2x_{ie}x_{je} \right]^2 \right]$$

Si denominamos

$$\gamma_{ijke} = (b_{ijke}^2 - x_{ie}^2 + 2x_{ie}x_{je})$$

Entonces

$$\begin{aligned} \frac{\partial S_k}{\partial x_{le}} &= 2w_{ke}^2 \sum_i \sum_j \gamma_{ijke} \frac{\partial}{\partial x_{le}} \gamma_{ijke} \\ &= 2w_{ke}^2 \sum_i \sum_j \gamma_{ijke} \left( \frac{\partial}{\partial x_{le}} b_{ijke}^2 - \frac{\partial}{\partial x_{le}} x_{ie}^2 + \frac{\partial}{\partial x_{le}} 2x_{ie}x_{je} \right) \\ &= 2w_{ke}^2 \sum_j \gamma_{ljke} (0 - 2x_{le} + 2x_{je}) \\ &= 4w_{ke}^2 \sum_j \gamma_{ljke} [x_{je} - x_{le}] \\ &= 4w_{ke}^2 \sum_j [b_{ljke}^2 - x_{le}^2 + 2x_{le}x_{je}] [x_{je} - x_{le}] \\ &= 4w_{ke}^2 \sum_j [x_{le}^3 - 3x_{le}^2x_{je} + 2x_{le}x_{je}^2 - b_{ljke}^2x_{le} + b_{ljke}^2x_{le}] \end{aligned}$$

Reordenando los términos

$$\frac{\partial S_k}{\partial x_{le}} = 4w_{ke}^2 \sum_j [x_{le}^3 - 3x_{le}^2x_{je} + 2x_{le}x_{je}^2 + b_{ljke}^2x_{le} - b_{ljke}^2x_{le}] \dots\dots\dots(2)$$

Que sustituyendo en (1) da una ecuación con una sola incógnita  $X_{le}$  que se iguala a cero y se resuelve. Todas las variables se consideran constantes excepto  $X_{le}$ . después procedemos a estimar el valor para otro coordenada.

Así sucesivamente obtenemos los valores para cada coordenada del punto l. Después nos trasladamos a un nuevo punto y procedemos hasta que los nuevos coordenadas para todos los estímulos se estimen. Entonces se volverá al principio del algoritmo de optimización y se comenzará otra iteración.

## **BIBLIOGRAFÍA**

Ashton, B.; Stacey, G., "Technical intelligence in business: understanding technology threads and opportunities", International Journal of Technology Management, Vol. 10, N°1 1995

Borg Ingwer, "Modern multidimensional scaling, Theory and Applications", Springer 2005

Callon M., Courtial, J. P. y Penan, H. "Cienciometría. El estudio cuantitativo de la actividad científica: de la bibliometría a la vigilancia tecnológica." Gijón: Trea 1995

De Moya Félix - "Atlas de la Ciencia Española: propuesta de un sistema de información científica" Universidad de Granada 2002

Escorsa Pere a, Maspons Ramon e Ortiz Ivette. "La integración entre la gestión del conocimiento y la inteligencia competitiva: la aportación de los mapas tecnológicos". Espacios. Vol. 21 (2) 2000

Faba Pérez Cristina, Guerrero Bote Vicente P. , De Moya Félix "Fundamentos y técnicas cibernéticas . Mérida : Sociedad de la Información/ Junta de Extremadura, 2004.

Ferreiro Aláez Luis "Bibliometría.Análisis bivalente", EYPSA 1993

Fred R. Mcfadden University "Modern database Management", Addison-Wesley Educational Publisher, Inc 1999

Guzmán - Sánchez, MV; Sotolongo - Aguilar, G\* "Mapas tecnológicos para la estrategia empresarial. Situación tecnológica de la neisseria meningitidis" Instituto Finlay, La Habana Cuba.

Kassirer , J. P., Edward, M. D. y Champion, M. D. (1994). Peer Review: Crude and Understudied, but Indispensable. *Journal of the American Medical Association*, vol. 272, no. 2, pp. 96-97.

Linares Gerson G. "Guía para implantar un sistema de inteligencia Tecnológica en organizaciones que realizan I&D", Tesis de maestría, Universidad Nacional Autónoma de México 2004.

López Eugenio. "Sistema de Inteligencia Tecnológica. El caso del área de catálisis FCC en la refinación del petróleo". *RECITEC*, Recife, v.5, n.2, p.187-198 2001.

Luukkonen, T. y Stahle, B.. "Quality evaluations in the management of basic and applied research". *Research Policy*, vol. 19, no. 4, pp. 357-368. 1990

Madera Jaramillo Ma. de Jesús "Una aplicación Web para la obtención de indicadores bibliométricos en ciencia y tecnología" , Tesis de maestría, Universidad Nacional Autónoma de México.

Martin, B. R. "The Use of Multiple Indicators in the Assessment of Basic Research". *Scientometrics*, vol. 36, no. 3, pp. 343-362. 1996

Novalis Susan "Access 2002 programación VBA" ANAYA 2002

Pérez López Cesar, "Métodos estadísticos avanzados con SPSS". Thomson 2005

Polanco Xavier "Infometría e Ingeniería del Conocimiento: Exploración de Datos y Análisis de la Información en vista del Descubrimiento de Conocimientos". Segundo Taller Iberoamericano sobre Indicadores en Ciencia y Tecnología, 24-26 de Abril de 1996, Cartagena de las Indias, Colombia

Real Deus José Eulogio "Escalamiento multidimensional", La muralla 2001

Rodríguez Salvador Marisela, "Análisis de Patentes en la Inteligencia Competitiva y Tecnológica: el Caso de los Materiales Avanzados", 2, 8 PUZZLE 2005

Rodríguez, M. (1998), "La Inteligencia Tecnológica: Elaboración de mapas tecnológicos para la identificación de líneas recientes de investigación en materiales avanzados y sinterización", Tesis Doctoral, Universidad Politécnica de Cataluña.

Schiffman Susan S. "Introduction to Multidimensional Scaling, Theory, methods and applications", Academia Press, 1981

Solesbury W. "Scientific research: demystifying peer review. Research Evaluation", vol. 6, no. 1, pp. 19-23. (1996).

Spinak, E. Diccionario Enciclopédico de Bibliometría, Cienciometría e Informetría. Caracas: UNESCO-CII/II. (1996).

Suárez-Balseiro Carlos A. Maura Sardo Mariano. " Bibliometría Evaluativo Seguimiento y Evaluación de la Actividad Científica y Tecnológica (SACT)" , Universidad de Puerto Rico, 2005.

Torgerson Warren S. Theory and methods of scaling , Robert E. Krieger Publishing Company. 1985

Young Forrest W. Multidimensional: History, theory and applications 1987