



Universidad Tecnológica de la Mixteca

Análisis de leyes y modelos del desarrollo de hardware y software

Tesis Doctoral

Presenta

Amando Alejandro Ruiz Figueroa

Bajo la dirección del

Dr. Pavel Makagonov

Huajuapán de León, Oaxaca. Diciembre de 2011



La miniaturización ha sido sinónimo de éxito en la microelectrónica; pero es necesario emigrar a nuevas tecnologías porque el límite está próximo.

Alejandro Ruiz

∞ Dedicatorias ∞

Al **todo poderoso**, por colocar las piezas exactas en los lugares exactos para la culminación de este trabajo.

A mi **padre**, porque aunque físicamente no está a mi lado, su gran recuerdo perdura en lo más profundo de mi alma (†).

A mi **madre**, por su abnegado amor en los momentos más difíciles, cuando todo lo creí perdido.

A mis **hermanos**, por abrir la senda que ha facilitado mi andar.

A mi **esposa**, por todo el amor en estos años juntos.

A mis **hijas**: por todo el tiempo robado.

Alejandro Ruiz

∞ Agradecimientos ∞

La felicidad por la obra terminada, y el placer de poner "punto final" a una "empresa" que hipoteca la vida entera, se manifiestan en reconocer y agradecer a quienes intervinieron e hicieron posible ésta investigación.

De manera muy especial quiero expresar mi gratitud al **Dr. Pavel Makagonov** por su inestimable ayuda, paciencia y acertada dirección; revisor infatigable, siempre sugiriendo ideas para el desarrollo de esta investigación.

Al **Dr. Modesto Seara Vázquez** rector del Sistema de Universidades Estatales de Oaxaca (SUNEO) por la visión futurista de este sistema de transformación en el estado de Oaxaca que ha dado a muchos de nosotros el beneficio de un trabajo estable y a la vez la oportunidad de realizar sueños que parecían inalcanzables.

Al comité revisor de la tesis, cuyas observaciones y aportaciones han enriquecido este trabajo: **Dr. Adolfo Guzmán Arenas** del Centro de Investigación en Computación del IPN; **Dra. Yulia Ledeneva** de la UAEM; **Dr. Luis Villaseñor Pineda** del Departamento de Ciencias de la Computación del INAOE; **Dr. Antonio Orantes Molina** del Instituto de Electrónica y Mecatrónica de la UTM; **Dr. Daniel Erasto Santos Reyes** de la División de Estudios de Posgrado de la UTM; **Dr. Yury Figatner** de la Universidad de la Sierra Sur.

Así también al **Dr. Konstantin Sboychakov**, **Dra. Virginia Berrón Lara** y demás personas que hoy se me escapan de la mente pero que en su momento aportaron su granito de arena para la culminación de este trabajo.

Finalmente a la Universidad Tecnológica de la Mixteca, Universidad de la Sierra Sur y al Consejo Nacional de



U.T.M.

Agradecimientos

A. Ruiz

Ciencia y Tecnología por el apoyo parcial en el proyecto N-39011A.

Alejandro Ruiz.

| | Número de página |
|---------------------|------------------|
| Frase. | i |
| Dedicatorias. | iii |
| Agradecimientos. | v |
| Contenido. | vii |
| Lista de figuras. | xi |
| Lista de tablas. | xv |
| Lista de acrónimos. | xvii |
| Resumen. | xix |
| Abstract | xxi |

Parte I

| | |
|---|----|
| Capítulo 1. Introducción. | 3 |
| 1.0. Formulación y justificación del problema. | 4 |
| 1.1. Objetivos y aportaciones del trabajo | 5 |
| 1.2. Hipótesis y enfoque. | 6 |
| 1.3. Restricciones. | 7 |
| 1.4. Estructura de la tesis. | 8 |
| Capítulo 2. Estado del arte. | 11 |
| 2.1. Ontologías. | 12 |
| 2.1.1. Términos de las ontologías. | 13 |
| 2.1.2. Clasificación de ontologías | 14 |
| 2.1.3. Ontologías existentes y librerías de ontologías | 15 |
| 2.1.4. Aplicaciones basadas en ontologías. | 16 |
| 2.2. Demandas prácticas de la bibliometría y la cienciometría. | 17 |
| 2.2.1. Origen y evolución histórica de la bibliometría. | 18 |
| 2.2.1.1. Leyes empíricas de la bibliometría. | 18 |
| 2.2.2. Evaluación de la actividad científica a través de la bibliometría. | 20 |
| 2.2.3. Indicadores bibliométricos | 20 |
| 2.2.3.1. Limitaciones de los indicadores bibliométricos. | 21 |
| 2.3. Modelos temporales de desarrollo tecnológico y científico. | 22 |
| 2.3.1. Revolución científica y tecnológica de la electrónica. | 23 |
| 2.3.1.1. El paradigma tecnológico del transistor. | 24 |
| 2.3.1.2. Cambio de paradigma y revolución tecnológica. | 25 |
| 2.3.1.3. Revolución tecnológica del transistor y su retroalimentación con la ciencia. | 26 |
| 2.3.2. El transistor como parte elemental de la integración a muy alta escala. | 27 |
| 2.3.3. Modelos temporales de desarrollo tecnológico y científico. | 33 |
| 2.3.4. Modelos auxiliares del ciclo de vida de un sistema. | 38 |
| 2.4. Formulación original de la ley Moore. | 42 |

| | |
|--|----|
| Capítulo 3. Metodología de construcción y algoritmo de la ontología. | 45 |
| 3.1. Análisis de la evolución de computación PCDS analizando textos cortos. | 46 |
| 3.1.1. Muestreo para modelos con conocimiento científico. | 46 |
| 3.1.2. Visual heuristic cluster analysis for texts. | 50 |
| 3.1.3. Método para análisis de datos con conocimiento pobre. | 51 |
| 3.1.4. Análisis de resultados. | 56 |
| 3.2. Método de construcción y algoritmo del prototipo rápido de ontología. | 57 |
| 3.2.1. Estructura del muestreo para creación y análisis del prototipo rápido de ontología. | 58 |
| 3.2.2. Método de construcción del prototipo rápido de ontología. | 59 |
| 3.2.3. Base de elementos de la ontología. | 61 |
| 3.2.3.1. Estructura de entrada de datos: cuerpo de textos. | 61 |
| 3.2.3.1.1. Textos estructurados jerárquicamente. | 62 |
| 3.2.3.1.2. Dominio. | 62 |
| 3.2.3.1.3. Datos. | 63 |
| 3.2.3.2 Estructura de salida: ontología. | 64 |
| 3.2.3.2.1. Relación. | 64 |
| 3.2.3.2.2. Palabra. | 64 |
| 3.2.3.2.3. Concepto. | 65 |
| 3.2.3.2.4. Tema. | 67 |
| 3.2.3.2.5. Ontología. | 67 |
| 3.2.4. Algoritmo para construir la ontología. | 68 |
| 3.2.4.1. Pre-procesamiento del cuerpo. | 68 |
| 3.2.4.2. Formación de la raíz. | 68 |
| 3.2.4.3. Recursión por nodos. | 69 |
| 3.2.4.4. Selección del sub-cuerpo. | 69 |
| 3.2.4.5. Selección de vocabulario. | 69 |
| 3.2.4.6. Agregado de documentos. | 69 |
| 3.2.4.7. Formación de conceptos. | 70 |
| 3.2.4.8. Selección de conceptos. | 70 |
| 3.2.4.9. Formación de temas. | 70 |
| 3.3. Análisis dinámico de la ontología. | 70 |
| 3.3.1. Estructura de la ontología de textos científicos. | 72 |
| 3.3.2. Construcción de la ontología. | 74 |
| 3.3.3. Resultados experimentales. | 76 |
| 3.3.4. Estabilidad de la ontología como medida de evolución. | 79 |
| 3.3.5. Conclusiones parciales. | 80 |

Parte II

| | |
|--|----|
| Capítulo 4. Modelos de desarrollo del HSW basado en el estudio de computación PCDS. | 85 |
| 4.1. Distribución temporal de resúmenes para computación PCDS. | 86 |
| 4.1.1. Distribución de resúmenes agrupados por año de publicación para computación PCDS. | 86 |
| 4.1.2. Comportamiento individual de computación PCDS. | 87 |



| | |
|---|-----|
| 4.1.3. Curva de distribución acumulada y curva S para computación PCDS. | 88 |
| 4.1.3.1. Ajuste del polinomio cúbico por mínimos cuadrados. | 89 |
| 4.1.3.2. Cálculo de los coeficientes del polinomio cúbico. | 90 |
| 4.1.4. Curva S de computación PCDS aproximada con un polinomio cúbico. | 92 |
| 4.1.5. Curva S de computación PCDS aproximada con la función logística. | 94 |
| 4.1.6. Superposición de las curvas cúbica y logística para computación PCDS. | 98 |
| 4.1.7. Curva S de VLSI y tolerancia a fallos aproximada con un polinomio cúbico y la función logística. | 100 |
| 4.1.7.1. Distribución de resúmenes agrupados por año de publicación para VLSI y tolerancia a fallos. | 100 |
| 4.1.7.2. Curva S y curva de distribución acumulada para VLSI y tolerancia a fallos. | 102 |
| 4.2. Nuevo enfoque de la ley de Moore. | 105 |
| 4.2.1. Conclusiones parciales. | 112 |
| | |
| Capítulo 5. Conclusiones y trabajo futuro. | 115 |
| Referencias. | 123 |
| Apéndice A. Glosario de términos. | 133 |
| Apéndice B. Listado de Programas. | 147 |
| Apéndice C. Tutorial de visual heuristic cluster analysis. | 157 |
| Apéndice D. Modelos 2004-2008 | 171 |
| Apéndice E. Material generado | 177 |

Lista de figuras.

| | Número de página | |
|--------------|--|----|
| Figura 2.1. | Revoluciones científicas de la evolución de la física. | 23 |
| Figura 2.2. | Evolución de la electrónica de estado sólido. | 24 |
| Figura 2.3. | Dibujo de la patente del Transistor de Efecto de Campo de Lilienfeld. | 27 |
| Figura 2.4. | Fotografía del primer transistor. | 29 |
| Figura 2.5. | Dibujo de la patente del transistor de punta de contacto. | 29 |
| Figura 2.6. | Fotografía del primer circuito integrado. | 30 |
| Figura 2.7. | Patente del primer circuito integrado. | 30 |
| Figura 2.8. | Evolución de los procesadores. | 32 |
| Figura 2.9. | Desarrollo del microprocesador basado en las distintas tecnologías de integración. | 33 |
| Figura 2.10. | Ciclo de vida de un sistema. | 39 |
| Figura 2.11. | Modelos de ciclo de vida de un sistema. | 39 |
| Figura 2.12. | Etapas del ciclo de vida de un sistema. | 40 |
| Figura 2.13. | Curvas S asociadas a un polinomio cúbico o a la función logística. | 40 |
| Figura 2.14. | Evolución de un sistema a salto. | 41 |
| Figura 2.15. | Curva tangente al conjunto de pequeñas curvas S. | 41 |
| Figura 2.16. | Gráfica costo-componente. | 42 |
| Figura 2.17. | Gráfica del número de componentes. | 43 |
| Figura 3.1 | Estimación para explorar la información en Internet. | 47 |
| Figura 3.2. | Errores comunes en archivos de texto plano. | 48 |
| Figura 3.3. | Corrección de errores en los archivos de texto plano. | 49 |
| Figura 3.4. | Archivos agrupados por año. | 49 |
| Figura 3.5. | Técnica para convertir DCP a DCE. | 52 |
| Figura 3.6. | Palabras comunes de la distribución de resúmenes sin agrupar. | 53 |
| Figura 3.7. | Tendencia de resúmenes agrupados por año de publicación. | 53 |
| Figura 3.8. | Matriz «texto / palabra» para la suma anual de resúmenes. | 54 |
| Figura 3.9. | Matriz «texto / texto» para la suma anual de resúmenes. | 54 |
| Figura 3.10. | Matriz «texto / texto» antes y después del clustering. | 55 |
| Figura 3.11. | Evolución temporal de los clústeres 1 y 3. | 55 |
| Figura 3.12. | Matriz texto / texto para patrones y textos anuales del clúster 1 y el clúster 3. | 57 |
| Figura 3.13. | Distribución de clústeres de palabras para distintos documentos. | 60 |
| Figura 3.14. | Distribución anual de clústeres de palabras. | 61 |
| Figura 3.15. | Jerarquía de los segmentos de texto. | 62 |
| Figura 3.16. | Estructura plana de clústeres. | 66 |
| Figura 3.17. | Algoritmo para construcción de la ontología. | 68 |
| Figura 3.18. | «Foto temporal» de la matriz «texto / palabras». | 73 |
| Figura 3.19. | Distribución de clúster de palabras para distintos documentos. | 75 |
| Figura 3.20. | Clúster de la matriz «palabra / palabra». | 76 |
| Figura 3.21. | Parte más importante de la ontología. | 77 |
| Figura 3.22. | Ontología para los períodos 1990 a 1997 y 1998 a 2004. | 78 |
| Figura 4.1. | Histograma de distribución de resúmenes para computación PCDS. | 87 |
| Figura 4.2. | Histograma de distribución de resúmenes por tema. | 88 |
| Figura 4.3. | Distribución acumulada para computación PCDS. | 89 |

| | | |
|--------------|---|-----|
| Figura 4.4. | Curva cúbica para computación PCDS. | 94 |
| Figura 4.5. | Curva logística para computación PCDS. | 98 |
| Figura 4.6. | Superposición de las curvas logística y cúbica para computación PCDS. | 99 |
| Figura 4.7. | Discrepancia entre la función logística y la curva cúbica para computación PCDS. | 100 |
| Figura 4.8. | Histograma de distribución de resúmenes para VLSI y Tolerancia a fallos. | 101 |
| Figura 4.9. | Distribución acumulada para VLSI y Tolerancia a fallos. | 102 |
| Figura 4.10. | Curva cúbica para VLSI y tolerancia a fallos. | 103 |
| Figura 4.11. | Curva logística para VLSI y tolerancia a fallos. | 104 |
| Figura 4.12. | Superposición de las curvas logística y cúbica para VLSI y Tolerancia a fallos. | 104 |
| Figura 4.13. | Esquema de la tecnología de doble núcleo. | 106 |
| Figura 4.14. | Conjunto de curvas S de la evolución del tamaño de los dispositivos electrónicos. | 107 |
| Figura 4.15. | Puntos con mayor grado de concavidad. | 108 |
| Figura 4.16. | Conjunto de curvas S y recta tangente. | 109 |
| Figura 4.17. | Metodología para hallar los puntos tangentes con la ecuación de recurrencia. | 110 |
| Figura 4.18. | Conjunto de curvas S aproximadas mediante parábolas. | 112 |
| Figura B.1. | Salida del programa B.1.7. | 151 |
| Figura B.2. | Salida del programa B.1.8, primera iteración. | 153 |
| Figura B.3. | Salida del programa B.1.8, segunda iteración. | 153 |
| Figura B.4. | Salida del programa B.1.8, tercera iteración. | 154 |
| Figura B.5. | Salida del programa B.1.8, cuarta iteración. | 154 |
| Figura B.6. | Salida del programa B.1.8, quinta iteración. | 156 |
| Figura C.1. | Menú de VHCA. | 158 |
| Figura C.2. | Construcción de la base de datos de textos completos. | 158 |
| Figura C.3. | Creación del diccionario básico. | 159 |
| Figura C.4. | Selección de textos de la base de datos. | 159 |
| Figura C.5. | Modificación del diccionario básico. | 160 |
| Figura C.6. | Ventana para ingresar al régimen de clasificación orientado a dominio. | 160 |
| Figura C.7. | Menú para distribución de textos por dominio y sub-dominio. | 161 |
| Figura C.8. | Ventana para elegir los criterios de selección de las palabras. | 161 |
| Figura C.9. | Ventana para guardar las matrices TW, TT y WW. | 162 |
| Figura C.10. | Apertura de archivos con el block de notas. | 163 |
| Figura C.11. | Menú para analizar el Diccionario Orientado a Dominio. | 163 |
| Figura C.12. | Apertura de VHCA. | 164 |
| Figura C.13. | Ventana que muestra la apertura de la matriz texto/texto. | 164 |
| Figura C.14. | Matriz simétrica TT. | 165 |
| Figura C.15. | Agrupación de dos clústeres importantes. | 165 |
| Figura C.16. | Resultados de los clústeres en forma de árbol. | 166 |
| Figura C.17. | Coloreado de clústeres. | 166 |
| Figura C.18. | Resultados del clustering. | 167 |
| Figura C.19. | Asignación de índices. | 167 |
| Figura C.20. | Carga del primer clúster. | 168 |
| Figura C.21. | Creación de DOD. | 168 |
| Figura C.22. | Creación de matrices. | 169 |
| Figura C.23. | Creación de la matriz texto/DOD. | 169 |



| | | |
|--------------|--|-----|
| Figura C.24. | Resultados del clustering. | 170 |
| Figura C.25. | Calculo de factores | 170 |
| Figura D.1. | Histograma de distribución de los resúmenes VLSI-FT período de 1998 al 2010. | 172 |
| Figura D.2. | Distribución acumulada para VLSI y FT período 1998-2010. | 173 |
| Figura D.3. | Aproximación de la curva S de VLSI-FT con un polinomio cúbico | 174 |
| Figura D.4. | Aproximación de la curva S de VLSI-FT con la función logística para VLSI y TF. | 175 |
| Figura D.5 | Superposición de curvas para VLSI y TF. | 176 |

Lista de tablas.

| | Número de página |
|---|------------------|
| Tabla 2.1. Escalas de integración para circuitos integrados | 32 |
| Tabla 3.1. Características de los resúmenes de computación PCDS agrupados por año de publicación. | 50 |
| Tabla 3.2. Características de los resúmenes de computación PCDS agrupados por períodos. | 50 |
| Tabla 4.0 Cantidad de resúmenes y datos acumulados para el periodo 1990-2004 para VLSI-FT. | 101 |
| Tabla 4.1. Evolución de los dispositivos electrónicos basado en el tamaño de su terminal de control. | 106 |
| Tabla 4.2. Puntos seleccionados que presentan el mayor grado de concavidad. | 108 |
| Tabla D.1. Datos de VLSI y TF para el periodo 1998-2010. | 172 |

Lista de Acrónimos.

- ACM:** Association for Computing Machinery.
- AMD:** Advanced Micro Devices.
- ASLIB:** Association for Information Management.
- BJT:** Bipolar Junction Transistor –Transistor de Unión Bipolar–.
- CI:** Integrated Circuit –Circuito Integrado–.
- D69-2002:** Artículo número 69 de computación distribuida publicado en el año 2002.
- D950:** Microprocesador INTEL.
- DOD:** Diccionario Orientado a Dominio.
- DCE:** Documento con Conocimiento Enriquecido.
- DGP:** Documento con Conocimiento Pobre.
- FET:** Field Effect Transistor –Transistor de Efecto de Campo–.
- FT:** Fault Tolerance – Tolerancia a Fallos
- HTML:** Hyper Text Markup Language –Lenguaje de Marcas Hipertextuales–.
- HSW:** Hardware y Software.
- I+D:** Investigación y desarrollo.
- IEEE:** Institute of Electrical and Electronics Engineers –Instituto de Ingenieros Eléctricos y Electrónicos–.
- INTEL:** Integrated Electronics –Electrónica Integrada–.
- ISL:** Indexación Semántica Latente.
- K8L:** Microprocesador de AMD.
- KB:** Kilo Bytes.
- LSI:** Large Scale Integration –Integración a Gran Escala–.
- MB:** Mega Bytes.
- MHz:** Mega Hertz.
- MOSFET:** Metal-Oxide-Semiconductor Field Effect Transistor –Transistor de Efecto de Campo Metal-Oxido-Semiconductor–.
- MPI:** Message Passing Interface –Interface de Pase de Mensajes–.
- MSI:** Medium Scale Integration –Integración a Mediana Escala–.
- NP:** Negativo-Positivo.
- NPN:** Negativo-Positivo-Negativo.
- PC:** Personal Computer –Computadora Personal–.
- PCDS:** Paralelo, Concurrente, Distribuido y Simultáneo.
- PDF:** Portable Document Format –Formato de Documento Portátil–.
- PLN:** Procesamiento de Lenguajes Naturales.
- PN:** Positivo-Negativo.
- PNP:** Positivo-Negativo-Positivo.
- PRO:** Prototipo Rápido de Ontología.
- PVM:** Parallel Virtual Machine –Máquina Virtual Paralela–.
- RAM:** Random Access Memory –Memoria de Acceso Aleatorio–.
- SIMM:** Single In-Line Memory Module –Módulo de Memoria de Línea Única–.
- SSI:** Small Scale Integration –Integración a Pequeña Escala–.
- TTL:** Transistor-Transistor Logic –Lógica Transistor-Transistor–.
- URL:** Uniform Resource Locator –Localizador Uniforme de Recursos–.



VHCA: Visual Heuristic Cluster Analysis.

VLSI: Very Large Scale Integration –Integración a Muy Gran Escala–.

WEB o **WWW:** World Wide Web -Telaraña Mundial-.

ZSI: Zero Scale Integration –Integración a Escala Cero–.

Resumen

Para investigar la tendencia de evolución del *hardware* y *software* -HSW-, se implementó una metodología basada en el análisis del flujo de resúmenes. Esta metodología muestra visibles ventajas sobre los métodos de *cienciometría* y *bibliometría*, que simplemente ‘cuentan artículos’ sin preocuparse de los niveles de abstracción, y además requieren grandes cantidades de información. Para desarrollar ésta investigación, se recopilaron 2300 resúmenes de la librería digital del Institute of Electrical and Electronics Engineers -IEEE-, relacionados con los temas de computación *Paralela*, *Concurrente*, *Distribuida* y *Simultánea* -PCDS- para un período de desarrollo de 15 años. Debido a que los resúmenes se consideran ‘objetos de conocimiento pobre’, fue necesario agruparlos en base a atributos externos -como es su fecha de publicación- para convertirlos en ‘objetos de conocimiento enriquecido’. Se empleó la herramienta «Visual Heuristic Cluster Analysis» -VHCA- para crear un Diccionario Orientado a Dominio -DOD-, éste se utilizó para construir un Prototipo Rápido de Ontología -PRO-, utilizando la estructura jerárquica de los documentos sobre computación PCDS: *nombres de congresos, de libros y revista; índices de congresos, de libros y revistas; nombre de artículos y cuerpo de resúmenes*. Cada nivel superior de la jerarquía pertenece a conceptos más abstractos, hasta el nivel más alto donde se tienen los cuatro conceptos PCDS. Las Ontologías construidas muestran la dinámica de conceptos y palabras; lo cual manifiesta la tendencia de evolución. Los resultados obtenidos revelaron que dentro de la estructura jerárquica de los documentos, los temas «VLSI» y «tolerancia a fallos», tienen un desarrollo significativo durante todo el período de investigación. Basado en éstos resultados se obtiene una *curva J*, y a partir de la cual, se reconstruye el modelo de *ciclo de vida* -curva S- utilizando las curvas *cúbica* y *logística*. Asimismo, la evolución de VLSI se vincula inevitablemente con la cantidad de transistores dentro de un chip; por lo que fue necesario investigar la *ley de Moore*, y establecer una característica más universal para medir su desarrollo. El análisis de esta ley produjo un conjunto de *curvas S*, lo que obligó a implementar un modelo de *curva tangente* a éste conjunto de curvas. El resultado principal en un aspecto práctico es la metodología de construcción de ésta tangente al conjunto de curvas S, ya que proporciona un modelo para predecir el límite temporal de desarrollo de la electrónica basada en el *flujo de electrones*. La metodología de construcción de ontologías, el descubrimiento dinámico del dominio científico y, la metodología de construcción de *curvas S* y *curva tangente*, pueden ser utilizadas para analizar cualquier rama científica sustentada en ciencia normal.

Abstract

In order to investigate the tendency of evolution of hardware and software –HSW– an economic methodology based on the analysis of abstracts flows of was developed and implemented. This methodology has notable advantages over *scientometrics* and *bibliometrics* methods, which turn out to be too tedious, and furthermore, require a great amount of information. In order to carry out this analysis, 2300 abstracts were compiled from the IEEE Digital library, relating to the fields of Parallel, Concurrent, Distributed and Simultaneous computation –PCDS– over a period of 15 years development. Because abstracts are regarded as ‘poor knowledge objects’, it was necessary to group them on the basis of external attributes –such as the date of publication– to transform them into ‘objects of enriched knowledge’. The tool kit «Visual Heuristic Cluster Analysis for Texts» –VHCA– was used to create a Domain Oriented Dictionary –DOD–, which was used to construct a Rapid Prototype of Ontology –PRO– based on hierarchical text structures on PCDS computation: names of congresses, books and journals; indexes of congresses, books and magazines; article titles and collections of summaries. Each higher level of the hierarchy belongs to more abstract concepts up to the highest level where the principal concept is “PCDS calculation”. The ontologies constructed show to the dynamic aspects of concepts and words, which reflects the evolutionary tendency. The results obtained revealed that within the hierarchic structure of PCDS computation, «VLSI» and «Fault Tolerance» subjects show significant development throughout the research period. On the basis of these results a *J-curve* is obtained and from which the *life cycle* model – *S-curve* – is reconstructed using the cubical polynomial segment and logistic curve. Also, the development of VLSI becomes inevitably tied with the amount of transistors inside a chip, which involved having to investigate Moore’s law and to formulate a more universal characteristic to measure its development. The analysis of this law produced a set of *S-curves* which made it necessary to implement a tangent curve model to this set of *S-curves*. The result of the reconstruction of this *tangent curve* provides a model to predict the temporal limits of development of electronics based on *electrons flow*. The ontology construction methodology, the dynamic discovery of the scientific domain and the *S-curve* and *tangent curve* construction methodology can be used in the analysis of any scientific field.



Parte I

Capítulo 1

Introducción.

El enfoque de la tesis vincula dos direcciones principales: **el análisis de la evolución temporal del hardware y software y, el estudio de las leyes de desarrollo técnico para el área de electrónica.** En la primera dirección se realiza un «trabajo histórico», donde se busca que las respuestas obtenidas estén sustentadas no solamente en apreciaciones subjetivas, sino principalmente en técnicas de minería de textos. Para la segunda dirección se realizaron algunos intentos para aplicar los métodos de la **teoría para la resolución de problemas inventivos (TRIZ)** [57], y descubrir tendencias y problemas relacionados con el área de electrónica. Estos intentos no surtieron el efecto deseado, debido a que las **leyes de desarrollo técnico (LDT)**, cuyo núcleo es TRIZ, están basadas en principios de la mecánica clásica y fueron desarrolladas considerando aspectos de procesos macrofísicos. Por lo que no es posible aplicarlos a procesos de nanoelectrónica, debido a que éstos se encuentran regidos por leyes de la mecánica cuántica.

Estos intentos muestran que es importante aclarar al menos la pregunta ¿Por qué no existe progreso suficiente en esta dirección de investigación? En atención a la pregunta es necesario mencionar que existe interés por parte de organismos públicos [56] e investigadores nacionales e internacionales [60][105][106][107][108][109][110].

1.0. Formulación y justificación del problema.

La creciente demanda en el financiamiento del sistema de Investigación y Desarrollo (I+D); cuyos resultados se reflejan en el progreso económico y social de un país, ha producido la firme necesidad de cuantificar y evaluar la actividad científica, su productividad y resultados. Las Instituciones Públicas implicadas en la distribución de los escasos recursos económicos destinados a la investigación, demandan cada vez más instrumentos de análisis que les permita comprobar la eficacia de sus políticas, y les ayude a racionalizar y adecuar sus inversiones de acuerdo a la proyección planeada. Realmente no es posible predecir el tiempo o el lugar donde emergerá un nuevo descubrimiento, pero es posible predecir el desarrollo de la «**ciencia normal**» [3], y de esta manera garantizar que el capital invertido en la investigación sea distribuido en base a las exigencias de cada rama científica y sobre todo proporcionar más apoyo aquellas que tengan mayores perspectivas de desarrollo.

La incertidumbre en la ciencia normal es menor que para el avance de la ciencia; la ciencia normal se basa en resultados anteriores y el estudio de éstos antecedentes, da la posibilidad de conocer las tendencias de desarrollo en una rama científica. La evolución del ciclo de vida para la ciencia normal transcurre de acuerdo a la **curva-S** para los parámetros cuantitativos [3]. Cuando un cierto límite de evolución cuantitativa se alcanza, el sistema experimenta cambios cualitativos o es sustituido por otro. Durante este proceso, los cambios cuantitativos (ciencia normal) ocurren continuamente mientras que los cambios cualitativos (ciencia revolucionaria) ocurren en pasos discretos [3]. La ciencia normal crece en la etapa de desarrollo del sistema y disminuye en el momento de desuso. Para conocer el desarrollo de la ciencia normal se usan enfoques basados en la **cienciometría** y la **bibliometría** [2][25][URL-1][URL-2][URL-3][URL-4]. Estos métodos exigen grandes cantidades de información, por ello, son utilizados por Instituciones Científicas, organizaciones Gubernamentales o grandes compañías con capacidad financiera. La idea de este trabajo consiste en desarrollar una metodología que permita ser empleada por Instituciones Públicas, encargadas de la distribución de los recursos económicos destinados a la investigación científica; así como por investigadores que no tienen acceso a las grandes fuentes de información, debido a que carecen de recursos económicos, pero tienen acceso a información (resúmenes) disponible de forma gratuita en Internet; o a su vez por investigadores principiantes que necesitan estimar las perspectivas de las distintas ramas científicas existentes.

1.1. Objetivos y aportaciones del trabajo.

El trabajo se orienta a la creación de una metodología para el análisis del desarrollo de una rama específica del conocimiento. Los resultados arrojados permiten detectar tendencias, observar sus cambios y predecir futuras líneas de oportunidad. Conocer estas líneas de oportunidad es de gran relevancia, sobre todo en un país con fuertes limitantes económicas en el campo de la investigación científica como es el caso de México. Para alcanzar las metas propuestas se parte de los siguientes objetivos:

Objetivo General

El objetivo principal de esta investigación consiste en:

Conocer la tendencia del hardware y software utilizando técnicas de minería de textos y modelos basados en la evolución de los dispositivos electrónicos.

Objetivos Específicos:

- Proponer e implementar una metodología para construir un **prototipo rápido de ontología** (PRO) u **ontología**, manipulando distintos documentos en un dominio limitado a **computación paralela, concurrente, distribuida y simultánea** (PCDS) para conocer la evolución del hardware y software.
- Implementar una metodología para medir y conocer la tendencia de evolución temporal del hardware y software (HSW).
- Investigar la evolución de los dispositivos electrónicos y conocer las leyes que rigen su comportamiento.
- Proponer una generalización de la ley de Moore basado en la evolución de **la terminal de control de los dispositivos electrónicos**.

A continuación se presentan las principales aportaciones derivadas de este trabajo:

1. Desarrollo de una herramienta computacional de apoyo en la toma de decisiones para la adjudicación de financiamiento en proyectos de investigación.
2. Integración de una herramienta para el análisis y estudio de cualquier rama científica sustentada en «ciencia normal», para conocer las perspectivas de su desarrollo.

3. Construcción de ontologías (o al menos su prototipo rápido) en bases de datos de acceso libre, basadas en el nivel de abstracción o generalización de documentos o entes analizados, y que requieran estructuración sin la participación de expertos en el dominio investigado.
4. Construcción automática de ontologías que muestran la dinámica de evolución de la rama científica estudiada.
5. Contribuir a los métodos bibliométricos y cientiométricos mediante un enfoque que toma en cuenta los niveles de abstracción.
6. La reinterpretación de la ley de Moore y su reemplazo de su forma clásica (como ley de desarrollo del mercado) a una ley de desarrollo técnico con contenido más físico.
7. El modelado original de los ciclos de vida (curvas “J”, “S” y tangentes al conjunto de curvas S) pueden aplicarse a cualquier disciplina científica.

Asimismo se muestran algunas aplicaciones prácticas de este trabajo:

1. La importancia práctica de la primera parte de la tesis es la orientación que ofrece al investigador inexperto en el inicio de su investigación, es decir, en la elección del tema para su trabajo futuro con amplias perspectivas de desarrollo; y que además no cuente con los recursos necesarios para comprar grandes colecciones de textos y/o no tenga la posibilidad de consultar expertos en el área de su interés.
2. Este trabajo es aplicable al sector gubernamental en el sentido de que los órganos encargados de regular y guiar el desarrollo de la ciencia y tecnología (SEP, CONACYT, Consejo Consultivo de Ciencias, Secretaría de Educación Pública, entre otros) puedan contar con una alternativa para evaluar el desarrollo de ciertas áreas, o para encontrar rezagos y deficiencias.

1.2. Hipótesis y enfoque.

Se expone un enfoque basado en el análisis de resúmenes sobre computación PCDS; así como la construcción de ontologías que muestran la dinámica de conceptos y palabras, y manifiestan la tendencia de evolución del hardware y software. La hipótesis para esta parte se plantea a continuación:

- La construcción de **ontologías** o **prototipo rápido de ontología** basados en listas de frecuencias de textos científicos, reflejan el estado de la ciencia normal y proporcionan una perspectiva para investigar los modelos de desarrollo del hardware y software.

La siguiente hipótesis surge a raíz de reconocer que las **leyes de desarrollo técnico** no contemplan el progreso del hardware y software:

- La razón de la escasa productividad de las **leyes de desarrollo técnico** de la electrónica bajo el esquema de la **teoría para la resolución de problemas inventivos** (TRIZ); se debe al cambio del conjunto de leyes físicas.
- El desarrollo de computación PCDS basado en ciencia normal, sigue un comportamiento análogo al modelo de ciclo de vida, el cual corresponde al clásico ciclo de nacimiento, desarrollo, madurez y muerte.

Para comprobar la hipótesis anterior, fue necesario aplicar y desarrollar métodos y modelos cuantitativos, utilizando la función cúbica, logística y paralelamente un conjunto de curvas tangentes.

1.3. Restricciones.

- La información relacionada con el hardware y software es muy extensa, lo que impide recopilar un porcentaje relativamente grande de información para obtener un muestreo representativo. Debido a esto, la investigación se delimita al estudio de computación PCDS, el cual nos proporciona una aproximación de la tendencia del hardware y software.
- Una parte muy pequeña de la información clasificada para computación PCDS no pertenece a la «ciencia normal», más bien pertenece a información clasificada como «ciencia revolucionaria» [3], la cual es ignorada por que no produce cambios sustanciales en los resultados.
- Según la opinión de los grandes círculos científicos todo tema suficientemente grande es distribuido en una proporción 1:3 en **revistas cuyo título corresponde al nombre del tema; revistas cuyas secciones contienen el nombre del tema; y otras revistas que no contienen el nombre del tema ni en su título ni en sus secciones**. Basado en esta hipótesis y tomando en cuenta que es imposible coleccionar el 100% de la información existente para computación PCDS se construye un muestreo representativo, seleccionando toda la información relacionada con computación PCDS de la base de

datos de la librería digital del Institute of Electrical and Electronics Engineers (IEEE) [URL-6].

- El anhelo de los investigadores es introducir (crear) palabras nuevas o complejas, que bajo ciertas situaciones son ignoradas por el sistema. Para corregir esta situación se incluyen estas palabras junto a las palabras de las cuales se derivan en el **diccionario orientado a dominio** (DOD. Por ejemplo: hypercube (cube), multithreads (thread), supercomputer (computer), etc.

1.4. Estructura de la tesis.

La estructura de la tesis se desarrolla en dos partes:

La parte I se centra en un trabajo histórico, basado en el análisis de información científica sobre computación PCDS (resúmenes, artículos, libros etc.). Se establece el análisis de clústeres para determinar las palabras de mayor frecuencia y que bajo ciertos criterios, forman la base de conocimiento de la ontología. Esta primera parte se estructura de la siguiente forma:

Capítulo 1. Describe una introducción del contenido de la tesis, sus alcances y limitaciones; la formulación y justificación del problema central.

Capítulo 2. En este capítulo no sólo se realiza una búsqueda y recopilación de información, sino un análisis de los aspectos que muestran los diferentes puntos de vista, el nivel de transformación y avance, y el estado actual que conforma la estructura de esta investigación.

Capítulo 3. Expone un enfoque basado en el análisis de resúmenes, para conocer el desarrollo de computación PCDS desde el punto de vista de ciencia normal. Se construye un **prototipo rápido de ontología** para analizar el desarrollo de estos temas, utilizando el esquema jerárquico de conceptos y palabras; por último se implementa un método semiautomático para construir una **ontología** con la cual se mide la evolución temporal.

La parte II se enfoca al estudio de las palabras más relevantes de la ontología. Para ello se implementan modelos de **curva S** y **curva tangente**; estos modelos permiten conocer la tendencia de desarrollo del hardware. La estructura de esta segunda parte se muestra a continuación:

Capítulo 4. Se construyen modelos matemáticos para determinar el desarrollo de **computación PCDS**; **integración a muy gran escala** (very-large-scale integration - VLSI) y **tolerancia a fallos** (fault tolerance - FT). Para ello, se propone el modelo de **ciclo de vida**, tomando en cuenta que

el desarrollo de computación PCDS desde el punto de vista de ciencia normal sigue un comportamiento similar (tercera hipótesis). Este modelo se aproxima utilizando una aproximación cúbica y paralelamente la curva logística. Asimismo se propone un enfoque nuevo (basado en la **evolución de la terminal de control de los dispositivos electrónicos**) para generalizar la ley de Moore; se establece una metodología para construir la curva tangente al conjunto de curvas S. Esto permite conocer el momento de desuso de **la ciencia normal de electrónica**, así como el tiempo en que se deben implementar nuevos métodos de **ciencia revolucionaria**, para sustituir la electrónica de flujo de electrones por alguna otra tecnología.

Capítulo 5. Se resume el objeto de la investigación y sus resultados, a través de una breve discusión sobre los aspectos citados; además se añade una lista con las posibles líneas de investigación que emergen de este trabajo.

Capítulo 2

Estado del arte.

En esta primera parte se da un esbozo general del estado que guardan los métodos y estrategias de **rescate de información**; y que debido a la gran acumulación de información textual disponible electrónicamente se ha incrementado vertiginosamente en las últimas décadas. Estos métodos y estrategias fueron introducidos a finales de los años 60 [1], desde entonces, han surgido distintos métodos como la **bibliometría y cienciometría** con el fin de explotar esta información. La importancia que tienen estos datos ha ido en aumento. Sin embargo, sin una semántica asociada, no son útiles, ya que resultan ambiguos, es por ello que se tienen que documentar para dotarlos de un significado. Es decir, se deben usar datos para describir los propios datos, o lo que es lo mismo, **metadatos**. Los metadatos por lo tanto, van a permitir realizar una estructuración de los contenidos, esto es, permiten describir las partes de un recurso. Sin embargo, si lo que se quiere es dar una descripción formal de contenidos mediante

el modelado de conceptos y relaciones, hace falta algo más potente que los metadatos, es aquí donde surgen las ontologías definidas por Gruber [70], como se describe más adelante.

En la segunda parte se trata los modelos temporales de desarrollo tecnológico y científico; así como el desarrollo de los dispositivos electrónicos desde su concepción (efecto Edison) hasta la era de los **nanoprocesadores** (22nm). Posteriormente se presenta los fundamentos de la Teoría para la Resolución de Problemas Inventivos (TRIZ) en la que se deduce que las **leyes de desarrollo técnico** (LDT, cuyo núcleo es TRIZ) están basadas en principios de la mecánica clásica, lo que da soporte para concluir que no es posible aplicarlas a procesos de **nanoelectrónica**, debido a que éstos se encuentran regidos por **leyes de la mecánica cuántica**. Por último se realiza un análisis de los modelos auxiliares del **ciclo de vida** de un sistema y se concluye con la reinterpretación de la **ley de Moore**.

2.1. Ontologías.

En filosofía, el término ontología se define como **“la parte de la metafísica que trata del ser en general y de sus propiedades trascendentales”** [100]; en cambio en la representación del conocimiento, tiene diferentes connotaciones. Existen diferentes autores que ofrecen su propia interpretación, algunas recogidas y analizadas en 1995 por Guarino [83] y Giaretta [83]. La definición declarativa más consolidada es la propuesta por Gruber [70][82] y extendida por Studer [93], como **“una especificación formal y explícita sobre una conceptualización compartida”**. La interpretación de esta definición es que las ontologías definen sus conceptos, propiedades, relaciones, funciones, restricciones y axiomas de forma “explícita” en algún lenguaje de implementación capaz de contener este conocimiento. El término “conceptualización” es una vista simplificada y genérica del mundo que deseamos representar para algún propósito específico, definiendo un vocabulario controlado. El conocimiento de las ontologías es establecido para ser usado de forma compartida por diferentes sistemas, que deberán comprometerse con el vocabulario utilizado en la ontología. El término “formal” se refiere a que la ontología debe implementarse en algún lenguaje computable por la máquina.

Desde la perspectiva de las aplicaciones que las utilizan, existen tres definiciones diferentes de ontología, aunque complementarias:

- para Swartout [94] **es un conjunto de términos estructurados jerárquicamente que describen un dominio**. La ontología es el “esqueleto” sobre el cual se construye la base del conocimiento. La característica principal que se acentúa en esta definición es que las ontologías proporcionan la estructura taxonómica de un dominio, la cual será especializada con los conocimientos específicos necesitados por la aplicación.
- Para Bernaras [72] **una ontología proporciona el significado que describe explícitamente la conceptualización del conocimiento representado en una base de conocimientos**.

Mediante un proceso de abstracción, el modelo conceptual de la ontología se construye a partir de los conocimientos procedentes de bases de conocimiento existentes. De esta forma, los conocimientos especificados en la ontología pueden ser utilizados más fácilmente para construir otras bases de conocimientos en el mismo dominio, o en dominios similares.

- Para Mizoguchi [89][90] **es una conceptualización en la que se hacen explícitos conceptos, significados y las relaciones semánticas entre ellos.** En este sentido, el propósito en la construcción de ontologías es capturar los conocimientos para que puedan ser empleados en diferentes sistemas, independientemente de la tarea que pretendan resolver.

A modo de conclusión, estas tres definiciones tienen en común la idea de que las ontologías proporcionan la conceptualización de los términos de un dominio, que sirve como soporte para la implementación de bases de conocimientos, preparadas para ser utilizadas por aplicaciones, y que resuelven diferentes tareas. No obstante, además de ser la base para los sistemas basados en conocimientos, actualmente el ámbito de aplicación de las ontologías abarca también otras áreas: procesamiento de lenguaje natural, web semántica, comercio electrónico, gestión de conocimientos, etc.

2.1.1. Términos de las ontologías.

Las ontologías deben ser creadas conforme a determinados criterios de diseño [82] y, para que este campo sea considerado como una ingeniería “madura”, las ontologías deben ser desarrolladas con adecuadas metodologías de desarrollo [76]. Según Uschold y Grüninger [95], dependiendo del propósito para el que sea creada, será construida con distintos grados de formalidad, desde los altamente informales expresados en lenguaje natural (lo cual contradice la definición de Gruber [82] y extendida posteriormente por Studer [93] que indica que las ontologías deben ser específicas formalmente), hasta los rigurosamente formales expresados en lenguajes con lógica de primer orden. Así, la apariencia de una ontología altamente formal será muy distinta a otra cuyas definiciones estén expresadas en lenguaje natural. En la mayoría de las ontologías aparecerán los siguientes términos [81][78]:

- **conceptos** o **clases**: son colecciones de objetos del dominio.
- **relaciones**: representan interacciones entre conceptos del dominio. Suelen aparecer las relaciones del tipo: subclase-de, parte de, conectado-a, etc. Pueden ser binarias o n-arias dependiendo de los argumentos que tomen.
- **funciones**: son un tipo especial de relación en las que, para un conjunto de argumentos de entrada, se obtiene un único argumento de salida. Por ejemplo, pueden aparecer funciones como días-de-demora, precio-de-coche, etc.
- **instancias**: representan elementos determinados de una clase o concepto.

- **axiomas:** son fórmulas, en la mayoría de las ocasiones expresadas en lógica de primer orden, que se usan para modelar sentencias que son siempre verdad, y se utilizan para verificar la corrección de la información contenida en la ontología. Los axiomas también son usados para realizar inferencias sobre los conocimientos almacenados.

En función de estos términos que se utilizan para representar los conocimientos del dominio se distinguen dos tipos de ontologías: **las ontologías ligeras** (lightweight ontologies) y **las ontologías pesadas** (heavyweight ontologies) [80]. Las ontologías ligeras contienen todos los términos excepto los axiomas. Las ontologías pesadas contienen todos los términos y además axiomas; cuantos más axiomas contengan serán consideradas más “pesadas”.

2.1.2. Clasificación de ontologías.

Existen diferentes clasificaciones de ontologías dependiendo del punto de vista que se quiera considerar. Mizoguchi [90] clasifica las ontologías desde la perspectiva del tipo de problemas que resuelven. En el nivel superior de su clasificación se consideran cuatro categorías:

- **ontologías de contenido:** construidas para reutilizar su conocimiento, vistas como cajas blancas que permiten reutilizar su vocabulario.
- **ontologías de indexación:** permiten la recuperación de casos cuando los agentes comparten conocimientos a través de bases de casos.
- **ontologías de comunicación:** usadas por agentes para obtener respuestas a preguntas concretas, vistas como cajas negras de conocimientos.
- **meta-ontologías:** utilizadas para representar ontologías.

En función del grado de reutilización de cada ontología, Van Heist [96] clasifica las ontologías en:

- **ontologías de representación:** proporcionan conceptualizaciones subyacentes a los paradigmas o formalismos de representación del conocimiento, es decir, proporcionan el vocabulario necesario para modelar otras ontologías, utilizando un determinado paradigma de representación del conocimiento. El ejemplo más característico es la frame ontology [75] disponible en el servidor de ontolingua, que proporciona el vocabulario necesario para representar una ontología siguiendo el paradigma de marcos.
- **ontologías genéricas o meta-ontologías:** proporcionan términos genéricos reutilizables en diferentes dominios. Como, por ejemplo, los términos: estado, evento, acción, componentes, etc.
- **ontologías de dominio:** expresan conceptualizaciones que son específicas para dominios particulares. Los conceptos en las ontologías de dominio son usualmente definidos como especializaciones de conceptos en ontologías genéricas.

- **ontologías de aplicación:** contienen todas las definiciones necesarias para modelar los conocimientos requeridos por una aplicación particular, incluyendo conceptos tomados de ontologías de dominio y genéricas, a menudo definidas utilizando el vocabulario indicado en las ontologías de representación. Puede contener extensiones de métodos y tareas específicas. Las ontologías de aplicación tienden a ser menos reutilizables que las anteriores, ya que son especificaciones concretas del dominio que se necesita para realizar una tarea particular en ese dominio determinado.

En ocasiones, es difícil clasificar una ontología sólo en uno de estos tipos, ya que el límite entre ellos no está definido claramente.

2.1.3. Ontologías existentes y librerías de Ontologías

En la actualidad existe un auge en el desarrollo de ontologías, tanto en ámbitos científicos, de las ciencias sociales, así como de la salud. Existen varias iniciativas que impulsan el uso de ontologías, como ontology.org, [Library linkBank](http://LibraryLinkBank.org) para la promoción de aplicaciones basadas en ontologías, especialmente en el campo de comercio electrónico; y en los últimos años se están desarrollando muchas ideas relacionadas con la web semántica, [OntoWeb](http://OntoWeb.org) [73] y [DAML](http://DAML.org). Atendiendo la clasificación de Van Heist [96], en este apartado se presenta una breve muestra de las ontologías más conocidas:

- **ontologías de representación:** la más utilizada es la [frame ontology](http://FrameOntology.org) [75] del [ontolingua server](http://Ontolingua.org), que contiene primitivas de representación, usadas en lenguajes basados en marcos, está implementada en [KIF](http://KIF.org) [77].
- **ontologías genéricas:** ejemplo más típico de ontologías genéricas es [mereology](http://Mereology.org) [74] que incluye, entre otras, definiciones de la relación “parte_de” y de sus propiedades. Una de las ontologías más populares, dentro de esta clasificación, es [CYC](http://CYC.org) [87], que contiene una enorme cantidad de términos de sentido común.
- **ontologías de dominio:** continuamente se publican ontologías en muchos dominios; se señala aquí solo una muestra representativa bien conocida por la comunidad que trabaja en el área de las ontologías. En el dominio de la ingeniería se puede mencionar [EngMath](http://EngMath.org) [81].
- **ontologías de aplicación:** se pueden examinar las noticias publicadas en Ontology.Org, [Library LinkBank](http://LibraryLinkBank.org) para comprobar cómo se construyen y se adaptan ontologías para aplicaciones concretas.

2.1.4. Aplicaciones basadas en ontologías.

Las ontologías son usadas en diversos proyectos de investigación y, en áreas como el procesamiento de lenguaje natural, sistema de información cooperativos, integración de información inteligente, comercio electrónico, etc., ya que proporcionan una visión compartida y común de un dominio, entre las personas y las aplicaciones [75].

Las ontologías son usadas en diferentes sistemas y entornos, Uschold y Jasper [95] proponen una clasificación de las aplicaciones ontológicas con fines prácticos. Su clasificación se basa en un conjunto de escenarios, tomados de la industria o de entornos académicos, donde las ontologías pueden ser aplicadas para alcanzar uno o más propósitos. Cada escenario se caracteriza por: el objetivo que se pretende conseguir, el papel que desempeña la ontología, los actores que se necesitan para implementar el escenario, la tecnología utilizada y su nivel de madurez. Una misma aplicación basada en ontologías puede pertenecer a más de un escenario. Los autores categorizan estos escenarios en tres grandes áreas:

- **realización neutral:** una ontología sobre un dominio determinado es implementada en un único lenguaje. Diferentes aplicaciones usarán la ontología como fuente de conocimientos, pero traducida al lenguaje que necesite la aplicación. Los beneficios de esta aproximación son la reutilización del conocimiento y el almacenamiento de gran cantidad de términos.
- **acceso común a información:** en algunos casos, se necesita algún tipo de información para una o más personas y aplicaciones con un vocabulario no familiar para todos, o con formatos inaccesibles. La ontología se construye proporcionando un vocabulario común consensuado para las múltiples aplicaciones, permitiendo el acceso a una fuente de información homogénea y comprensible. La diferencia con el escenario anterior es que, en este caso, no se traduce a diferentes lenguajes o formatos, si no que la ontología se mantiene como estándar, que deben usar las diferentes personas y aplicaciones.
- **indexación:** aplicaciones que realizan búsquedas basadas en conceptos utilizan una ontología como índice para localizar términos. Se obtiene un acceso más rápido, consensuando y controlando los recursos de información almacenados.

Aunque el objetivo de crear ontologías es poder compartir y utilizar su conocimiento por otros sistemas, aún no se ha conseguido que estos procesos se realicen de forma sistemática. Esto es debido a los diversos problemas [71]: las ontologías están diseminadas en diferentes servidores, la representación difiere dependiendo del servidor en el que están almacenadas, suelen estar descritas con diferente nivel de detalle, y no existe un formato común para describir la información relevante, donde el usuario potencial pueda examinar las características de la ontología que pretende reutilizar. Estos problemas suelen obstaculizar la construcción de aplicaciones basadas en ontologías, aunque es previsible que, con el diseño de nuevos estándares

de lenguaje web, se incrementa notablemente la reutilización de ontologías en la llamada web semántica [73].

Existen algunos enfoques interesantes como **Clasitex** de Guzmán Arenas [104], que analiza documentos escritos en español, y presenta los temas principales de un documento; ésta herramienta hace el análisis por conceptos, no por palabras clave. Cuenta con una ontología de dos niveles “conceptos y palabras”. Asimismo el trabajo de Jiménez [URL-19] plantea desarrollar herramientas computacionales con el propósito de extraer términos; identificar conceptos, y reconocer relaciones léxico-semánticas (hiponimia y holonimia, entre otras), a partir de textos del dominio cultural.

De esta forma podemos enumerar mas ejemplos de trabajos interesantes, pero el común denominador es la demanda de trabajo manual.

2.2. Demandas prácticas de la bibliometría y la cienciometría.

El primer problema que se plantea para la definición conceptual de las ciencias de la documentación, estriba en la delimitación misma de sus campos [2]:

- **Cienciología:** es la aplicación de las técnicas científicas en el estudio de la propia ciencia; supone un cambio radical en el enfoque del estudio de la ciencia, hasta aquí realizado por vía de la filosofía y las ciencias humanas. En cuanto supone una técnica de medida no es sino cienciometría.
- **Infometría:** ciencia de la aplicación de los métodos matemáticos a los hechos y situaciones que se producen en el campo de la información, para describir y analizar sus fenómenos, descubrir sus leyes y servir de soporte a sus decisiones. Se ocupa de la aplicación de cualquier rama de las matemáticas a las ciencias de la información.
- **Bibliometría:** subdisciplina que se ocupa del empleo de una rama de las matemáticas (la estadística) a las ciencias de la información, así como de las leyes o modelos matemáticos derivados de dicha aplicación y que reflejan el comportamiento colectivo de los hechos, fenómenos o relaciones del mundo informativo-documental.

Pritchard, padre del término bibliometría la define como "la aplicación de las matemáticas y métodos estadísticos a libros y otros medios de comunicación para arrojar luz sobre los procesos de la comunicación escrita y de la naturaleza y el curso de desarrollo de una disciplina mediante el recuento y análisis de las diferentes facetas de esta comunicación"[2]. Raisig precisa que la Bibliometría es el establecimiento e interpretación de estadísticas relativas a libros y revistas para demostrar los movimientos históricos, determinar el uso nacional o universal de aquellos con fines de investigación y establecer en situaciones locales su uso general. En la actualidad se plantea el problema de adoptar una definición amplia para la bibliometría que incluya su

aspecto sociométrico o dejarla restringida a límites más estrictos. En el primer caso supone aceptar que bibliometría y sociometría documental no son más que dos vertientes de una misma ciencia. López Piñero, que se inclina por esta tendencia considera que la bibliometría se define por dos objetivos [URL-18]:

- Volumen, crecimiento, obsolescencia y difusión de la literatura científica.
- Estudio de la estructura social de los grupos que producen, transmiten y utilizan la Ciencia.

Potter[2], siguiendo también esta línea afirma que la bibliometría es el estudio y medida del espectro de publicaciones de todas las formas de comunicación escrita y de sus autores. Lara Guitard mantiene que la sociometría documental es una ciencia claramente interdisciplinar que puede ser reclamada indistintamente por las ciencias de la información y por la sociología [61].

2.2.1. Origen y evolución histórica de la bibliometría.

El primer trabajo bibliométrico fue presentado en 1917 por Coles y Eales en Science Progress, y en él analizaban publicaciones sobre anatomía comparativa entre 1543 a 1860. En 1923 Hulme introduce el término bibliografía estadística; en 1926 Lotka procedió a estudiar la productividad de los autores y enunció la **ley de Lotka**. En 1934, Bradford, bibliotecario del museo de la ciencia de Londres enunció el principio de la dispersión de la literatura científica en las publicaciones periódicas y en 1948 dedicó un capítulo, titulado **el caos documental** de su libro **documentation** para exponer sus ideas sobre las distribuciones temáticas, aspecto más tarde comentado por Vickery. El mismo año, Ranganathan propuso el término Librametry. Una año después Zipf enuncia su ley sobre la frecuencia de utilización de las palabras en un texto y finalmente en 1969, Pritchard propone por primera vez el término Bibliometría en un trabajo titulado **Statistical bibliography or bibliometrics**. La primera tesis doctoral sobre este campo aparece en 1970 y el primer libro fue publicado por Donohue [2][61][URL-18].

2.2.1.1. Leyes empíricas de la bibliometría.

Tras el desarrollo de la ciencia bibliométrica se han enunciado multitud de leyes relacionadas con ella, unas de mayor peso que otras. Sin embargo, las leyes bibliométricas fundamentales son las siguientes [61][65][66]:

Ley de Zipf: si las palabras que aparecen en un texto lo suficientemente largo se cuentan y clasifican por la frecuencia de aparición, esta frecuencia es inversamente proporcional al orden de clasificación. Dicho de otro modo: en todo documento elaborado en lenguaje natural existe un pequeño grupo de palabras cuya frecuencia es muy alta, mientras que existe un grupo muy

amplio de palabras de escasa utilización. Zipf atribuye este fenómeno a la ley del mínimo esfuerzo que presta un significado amplio a ciertas palabras por ejemplo (hacer en español, faire en francés, get en inglés), provocando así su uso continuo.

Ley del crecimiento de Price: la tasa de crecimiento documental en una materia es proporcional al volumen alcanzado (curva exponencial), hasta el momento en que se llega al grado de saturación (curva logística).

Ley de obsolescencia de Price: la curva de obsolescencia de la literatura científica es exponencial, aunque con una deformación inicial, ya que la literatura reciente se cita comparativamente más.

Ley de la dispersión de Bradford: si se considera un número suficientemente elevado de trabajos sobre un tema determinado; la mitad de los artículos relativos al tema de interés se encuentra concentrada en un número muy pequeño de revistas, mientras que a partir de esta tasa, aumentos muy pequeños del número de artículos determinan grandes aumentos del número de revistas.

Brookes: estudió el fenómeno de la dispersión, llegando a la conclusión de que sus principales aplicaciones son las siguientes:

- cálculo del costo de adquisición de las revistas científicas que publican una cantidad suficiente de trabajos sobre un tema de interés.
- determinación de la fracción de cobertura total que podría obtenerse para un determinado presupuesto.
- distribución óptima de revistas entre una biblioteca central y bibliotecas coordinadas.
- subdivisión de fondos documentales de acuerdo con la reiteración que se hace de su uso.
- cálculo de la cobertura documental facilitada por servicios de indización y resúmenes.
- cálculo de la cobertura documental necesaria con que debe contar un servicio de fotodocumentación.
- predicción de las tendencias de publicación de trabajos sobre temas específicos.

Ley de la productividad de Lotka: menciona que el mayor número de autores publican el menor número de trabajos, mientras que el menor número de autores publica un mayor número de trabajos, siendo éste el grupo más prolífico. Además junto con la ley de Zipf son útiles para formar muestreos.

Ley de Platz: la visibilidad de un autor es el logaritmo de las citas que produce.

Ley de Price: el 50% de las citas se concentra en un reducido número de trabajos y el otro 50% se reparte entre la totalidad de los artículos publicados.

2.2.2 Evaluación de la actividad científica a través de la bibliometría.

La explosión documental y el avance de la investigación científica han hecho necesaria la aplicación de técnicas bibliométricas en el estudio de la documentación. Por otra parte, al ser esta ciencia en buena parte de índole sociológica se ha visto sometida al mismo proceso cuantificador que han sufrido en los últimos años todas las ciencias humanas, aunque con un considerable retraso en comparación con otras tales como psicología o sociología [62][63][64]. No hay que olvidar que las leyes de Bradford, por ejemplo, fueron en un principio tomadas como simples curiosidades sin valor y fue necesario que nuevas generaciones de documentalistas con mentes más analíticas fueran llegando a los centros de trabajo, para que estas investigaciones se aplicaran al campo de la actividad cotidiana. Dichas aplicaciones cubren los siguientes campos:

- planificación de servicios de información y política de investigación.
- estudio y control de los fondos activos de las bibliotecas.
- determinación de los núcleos de revistas (selección de suscripciones).
- delimitación de tendencias en la investigación (colegios invisibles).

El uso de los indicadores bibliométricos para el estudio de la investigación de un área se basa en que las publicaciones científicas son un resultado esencial de dicha actividad. Un nuevo conocimiento adquiere valor cuando se da a conocer y se difunde, porque sólo así podrá contribuir al avance científico [62][63][64]. Por ello, la publicación científica se convierte en un resultado importante y tangible de la investigación, y los indicadores bibliométricos adquieren validez como medida de la actividad científica [63][64].

2.2.3. Indicadores bibliométricos

Los indicadores bibliométricos proporcionan información sobre los resultados del proceso investigador, su volumen, evolución, visibilidad y estructura. Así permiten valorar la actividad científica, y la influencia (o impacto) tanto del trabajo como de las fuentes. De acuerdo con ello, se pueden clasificar en dos grandes grupos: indicadores de actividad e indicadores de impacto [66].

Indicadores de actividad: los indicadores de actividad científica permiten visualizar el estado real de la ciencia. Dentro de ellos se encuentran:

- número y distribución de publicaciones: miden el número total de publicaciones por instituciones y su distribución.
- productividad: número de trabajos por autor, revista o institución.
- dispersión de las publicaciones: análisis de las publicaciones sobre un tema o área entre las diversas fuentes de información. Permite descubrir núcleos de autores o revistas.

- colaboración en las publicaciones: índice de firmas por trabajo empleado para determinar la actividad y cooperación científica entre grupos de científicos o instituciones.
- vida media de la citación o envejecimiento: número de años transcurridos desde la publicación, en el cual las citas disminuyen a 50% de su valor inicial.
- conexiones entre autores: estudio de referencias que un trabajo hace a otro, y citas que éste recibe de aquel. Los datos obtenidos son muy importantes porque a partir de ellos se pueden obtener otros indicadores; el índice de obsolescencia o envejecimiento, vida media y detección de colegios invisibles.

Cabe destacar que los estudios de productividad por autores han sido investigados por Lotka, y arrojan como resultado la existencia de un pequeño grupo de personas muy productivas, al lado de un gran número de personas que apenas publican [65][66]. La productividad por revistas fue estudiada por Bradford, quien determinó que una gran cantidad de publicaciones sobre un tema determinado están presentes en un núcleo pequeño de revistas [65][66].

Indicadores de impacto: con los indicadores de impacto se puede valorar el impacto de autores, trabajos o revistas.

- Documentos recientes muy citados (hot papers). Los documentos alcanzan su máximo de citas entre 2 y 4 años después de ser publicados, pero algunos muestran una actividad de citas inusualmente alta poco tiempo después de su publicación, con respecto a otros documentos del mismo campo de investigación y antigüedad similar.
- Impacto de las revistas. Se obtiene mediante los datos publicados periódicamente por el Institute for Scientific Information (ISI) en el Journal Citation Report (JCR), que presenta datos estadísticos cuantificables y proveen una vía para evaluar las revistas más importantes a nivel mundial, así como su impacto e influencia en la comunidad de investigación [67].

2.2.3.1. Limitaciones de los indicadores bibliométricos.

Existen dos tipos de limitaciones en los indicadores bibliométricos que son los de **actividad** y de **impacto**:

- **de actividad:** los indicadores bibliométricos son válidos en aquellas áreas en las que las publicaciones son un resultado esencial de la investigación. Por esta razón, su validez es de máxima relevancia en el estudio de las áreas básicas, donde predominan publicaciones científicas, y menor en las áreas tecnológicas o aplicadas [62]. Sólo aportan información sobre la cantidad de publicaciones, pero no sobre su calidad, por lo que tiene mucho interés combinar estos indicadores con los de impacto y/o juicio de expertos [64]. Además, no deben efectuarse comparaciones entre áreas temáticas, porque

los hábitos de publicación y productividad de los autores difieren. Estas diferencias son especialmente importantes entre las distintas áreas científicas (ciencias sociales, ingeniería y tecnología, ciencias naturales y experimentales), y también existen entre las diversas disciplinas que componen un área.

- **de impacto:** las limitaciones expuestas para el análisis de citas son también válidas para el uso del factor de impacto, ya que este indicador se calcula en función de las citas que reciben las revistas.

Un análisis bibliométrico es una valiosa herramienta, tanto para el estudio del estado de las distintas disciplinas científicas, como para el de producción científica de una determinada área, disciplina o tema. Su objeto de estudio son los productos del pensamiento representados físicamente en documentos. En otras palabras, el conocimiento intelectual apoyado en soporte material [62][70][71], con lo que se puede ayudar a valorar el estado actual de la ciencia y apoyar la toma de decisiones y dirección de la investigación. El inconveniente que presentan estos métodos es que solo “cuentan artículos” sin preocuparse de sus niveles de abstracción y además requieren grandes cantidades de información; por lo que se propone un enfoque y una metodología más económica para evaluaciones más rápidas.

2.3. Modelos temporales de desarrollo tecnológico y científico.

Es importante mostrar los distintos campos de la física y señalar la interrelación entre ellos, y en especial su posición en el marco de su evolución, mostrando las rupturas sustanciales que dieron origen a la ciencia electrónica. La figura 2.1 muestra la evolución de la física desde el modelo aristotélico, pasando por dos grandes rupturas que tuvieron un impacto significativo en el pensamiento y la cultura de la época, dado que cambiaron radicalmente la visión del universo: la ruptura causada por la revolución Copernicana y la causada por la teoría de la relatividad de Albert Einstein y la mecánica cuántica de Max Planck, en lo que constituye la física moderna desde el año 1905. Dichas rupturas no fueron sólo fracturas epistemológicas sino también culturales ya que sacudieron los cimientos de la física que, hasta esos momentos, permitía validar la representación característica de la naturaleza. Las teorías de Einstein y Planck constituyen, por consiguiente una revolución conceptual [44].

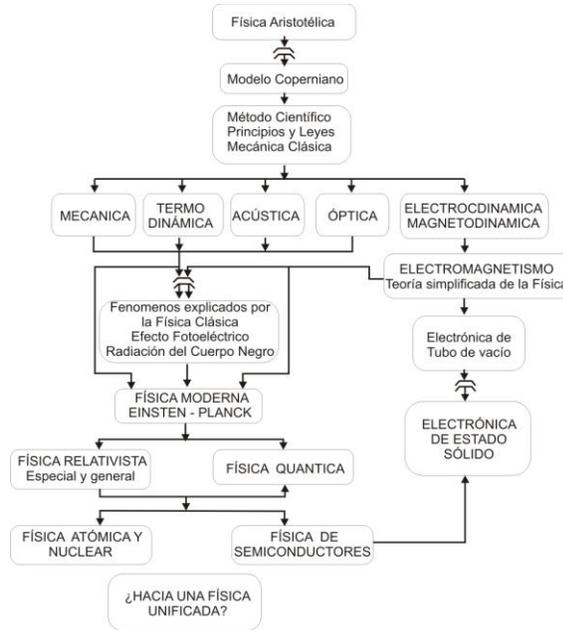


Figura 2.1. Revoluciones científicas en la evolución de la física. Publicado con permiso del autor [44].

Es importante resaltar la aparición de la **teoría unificada de la física**, en relación con los fenómenos eléctricos y magnéticos, los cuales están implícitos en la teoría electromagnética formulada por el físico James Clerk Maxwell; teoría que permitió integrar los fenómenos eléctricos y magnéticos, mediante un conjunto de ecuaciones generales que explicaban perfectamente la propagación del campo electromagnético, y la naturaleza de la luz. La teoría electromagnética permitió la explicación y la aplicación de fenómenos asociados, a la transmisión de ondas electromagnéticas, y al funcionamiento del tubo de vacío o bulbo, que dominó el escenario de la tecnología electrónica durante aproximadamente medio siglo, hasta la aparición del transistor en 1947. La aparición de las teorías que sustentaron la física moderna del siglo XX, sirvieron para crear un nuevo campo de investigación científica, que habría de denominarse **física de semiconductores**, lo que originó una ruptura de la electrónica de vacío con la electrónica de estado sólido; origen de una revolución tecnológica.

2.3.1 Revolución científica y tecnológica de la electrónica.

Para entender el concepto de revolución tecnológica, se realiza una aproximación a partir de los conceptos de **ciencia normal** y **ciencia revolucionaria** establecidos por Thomas Kuhn, en su trabajo sobre la estructura de las revoluciones científicas y los elementos conceptuales básicos [3]. En especial el concepto de paradigma, a través del cual se realiza una aproximación, hacia el conocimiento de la revolución tecnológica producida por el transistor; originada a su vez por la

revolución científica de la teoría de la relatividad y la mecánica cuántica como se aprecia en la figura 2.2. Kuhn ofrece una teoría del cambio científico, en la que proporciona los elementos para hacer un análisis teórico que permita reflexionar, entre otras cosas, sobre la naturaleza y la necesidad de las revoluciones científicas y que por extensión, puede aplicarse al cambio tecnológico.

Algunos de esos elementos son: **comunidad científica, paradigma, revolución teórica, revolución científica, ciencia normal, ciencia revolucionaria, crisis, cambio paradigmático, conocimiento acumulativo, inconmensurabilidad**, etc. Según Kuhn, los cambios de paradigma conducen a una revolución en la concepción que se tiene del mundo, con lo cual puede asumirse de manera análoga, que un cambio de paradigma tecnológico, o mejor aún, científico-tecnológico como lo constituye la electrónica de estado sólido, induce igualmente una revolución en la concepción del mundo, como efectivamente ha ocurrido, dadas las implicaciones, no sólo tecnológicas y científicas, sino sociales, culturales y políticas que provocó la revolución tecnológica del transistor.

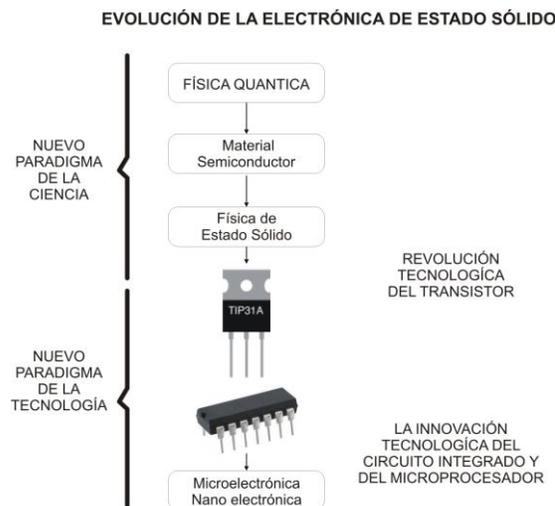


Figura 2.2. Evolución de la electrónica de estado sólido. Publicado con permiso del autor [44].

2.3.1.1. El paradigma tecnológico del transistor.

En la tecnología del transistor concebida como paradigma tecnológico, en el sentido de Kuhn, están presentes todos los elementos constitutivos de su teoría. Si combinamos la teoría física de los semiconductores con la electrónica de estado sólido, estaríamos ante un paradigma científico-tecnológico, retroalimentado en sí mismo con el desarrollo continuo de la ciencia y la tecnología; y consolidado como un todo, como un cuerpo holístico al que se aplican

perfectamente los términos asociados a cada uno de esos elementos: **comunidad científica** (la de los científicos de la física de semiconductores y, la de los ingenieros de desarrollo de tecnología electrónica de estado sólido); **paradigma** (el de la física y el de la electrónica); **revolución teórica** (en la conceptualización teórica que marca una ruptura en relación con la tecnología de vacío, soportada en la teoría electromagnética); **revolución científica** (en relación con el elemento anterior); **ciencia normal** (la que imperaba hasta el momento del desarrollo del transistor y, que fundamentaba la tecnología existente hasta entonces, el electromagnetismo); **ciencia revolucionaria** (la que entra a dominar el panorama a partir de la teoría cuántica y relativista y, que alcanza su esplendor con la teoría de semiconductores, que da lugar al nuevo paradigma tecnológico); **crisis** (cuando se genera la ruptura entre el transistor y el tubo de vacío); **cambio paradigmático** (el que sucede con la aparición y consolidación de la nueva tecnología).

2.3.1.2. Cambio de paradigma y revolución tecnológica.

El cambio de paradigma tecnológico se entiende como una transformación del patrón tecnológico, del modelo imperante; o dicho de otro modo, es un cambio de sentido común, y de la manera como se perciben las cosas, en lo que respecta a las prácticas más eficientes, en materia de producción de tecnología o del uso de la misma, para resolver problemas mediante determinadas aplicaciones.

El origen de ese cambio de paradigma es, en consecuencia, una revolución tecnológica, a la manera como Kuhn plantea su revolución científica. El gran problema que surge de los retos que implica un cambio de paradigma tecnológico, es que se hace necesario comprender fehacientemente, la naturaleza precisa del cambio para poder identificar claramente el nuevo escenario, y las nuevas reglas del juego. Un cambio de paradigma tecnológico como el que significó el transistor, es un cambio radical en las herramientas y en el modo (modelo) de hacer las cosas; es también un cambio en los patrones organizativos y en las posibilidades de la nueva tecnología; es enfrentarse a un nuevo y enorme potencial de generación de riqueza, cuyo aprovechamiento exige adoptar una nueva lógica y un nuevo modo de pensar. Una consecuencia natural de todo esto, es la dificultad en que la comunidad científica y académica se encuentra inmersa, una época de turbulencia y llena de incertidumbres. Es una postura conceptual que genera un nuevo problema, y ello sumerge a quien así piensa en un caos científico, en términos de no entender claramente lo que está ocurriendo con dicho cambio tecnológico.

Desde el punto de vista de las revoluciones tecnológicas, vale la pena mencionar también que a finales de la primera mitad del siglo pasado, es decir, la época en que hizo su aparición el transistor, ya se vislumbraba el agotamiento del sistema tecnológico electromecánico operante, con la aparición de tecnologías de "reemplazo". Este agotamiento se hacía evidente justamente con la aparición de los materiales sintéticos, la tecnología nuclear y espacial, la electrónica

avanzada (con el paradigma de estado sólido; el transistor y el circuito integrado como sus grandes exponentes).

El actual paradigma científico-tecnológico, está caracterizado por la preeminencia de las tecnologías de la información y las comunicaciones (los polimateriales, la biotecnología y las multienergías). En este paradigma, el elemento central es la electrónica, no sólo como núcleo del primer componente del paradigma, sino como elemento tecnológico articulador que permite el avance y aplicación de los demás componentes, ahora no con base en el transistor como unidad básica, sino del circuito integrado, el microprocesador y el nanoprocesador, entendido como un conjunto conformado por millones de unidades base. Esto indica, que el paradigma tecnológico del transistor, aún subsiste y permite la existencia y consolidación de otros paradigmas.

2.3.1.3. Revolución tecnológica del transistor y su retroalimentación con la ciencia.

Partiendo del electromagnetismo como teoría en la que sustenta el funcionamiento de los tubos de vacío (bulbos), no hay duda de la importante relación existente entre ciencia y tecnología en el estudio del electromagnetismo. La interacción, interdependencia y retroalimentación entre el conocimiento científico, asociado a dicho fenómeno y sus aplicaciones prácticas, es realmente evidente y notable, dado que una vez que se llevaron a cabo los descubrimientos científicos, tuvieron de inmediato una aplicación práctica y viceversa, tales aplicaciones prácticas, a su vez, impulsaron la investigación científica a niveles insospechados para resolver nuevos y complejos problemas, lo cual propuso nuevos horizontes científicos y tecnológicos, en una espiral ascendente que parece no tener fin. De la misma manera que con el electromagnetismo, la relación entre la física de semiconductores como teoría científica, está estrechamente ligada a la electrónica del estado sólido, basada en el transistor y los circuitos integrados; que la relación ciencia-tecnología, se estrecha y se cierra de tal forma, que la frontera entre ambas, hoy en día, no está claramente definida. Ni siquiera podríamos hablar de frontera, dado que las dos se hallan tan interrelacionadas que no es posible concebir el avance de una sin la ayuda de la otra. En la microelectrónica, por ejemplo, esta frontera se diluye hasta tal punto que la **integración a muy grande escala** (VLSI) (plasmado en un circuito integrado) y el desarrollo de la física de los semiconductores, se confunden como un único cuerpo de conocimiento. He aquí un excelente caso para ilustrar el concepto de **tecnociencia**, en donde no es posible hacer una separación entre los conceptos de ciencia y tecnología, es decir, entre la física de semiconductores y la electrónica del estado sólido. En definitiva, es el concepto de tecnociencia, el que mejor da cuenta de la complicada red de relaciones e interacciones entre la ciencia y la tecnología, implicadas en el paradigma de la electrónica de estado sólido, basada en el transistor. No es posible concebir el increíble y veloz desarrollo de la industria de los dispositivos de estado sólido, y su amplio espectro de aplicaciones, sin el desarrollo paralelo de la física de

semiconductores y, a su vez, la física de semiconductores no hubiera podido evolucionar hasta sus aportaciones actuales a la microelectrónica y la nanoelectrónica, sin la retroalimentación del desarrollo del transistor y los circuitos integrados [44].

2.3.2. El transistor como parte elemental de la integración a muy alta escala (VLSI).

Aún, con los grandes avances en la ciencia electrónica, su comienzo sigue en discusión; en ocasiones se atribuye a los esfuerzos de los primeros científicos, que emplearon corrientes eléctricas a través de envases de cristal al vacío. Sin embargo, muchos ubican el inicio con Thomas A. Edison, quien descubrió el flujo de electrones entre dos elementos en un tubo al vacío (el fenómeno fue patentado en 1883 y se le llamó **efecto Edison**) [32]. A partir de este momento el desarrollo de los dispositivos electrónicos ha seguido un camino vertiginoso como a continuación se describe. En 1904 el físico inglés John Ambrose Fleming, fue el primero en encontrar una aplicación práctica de este fenómeno al construir la **válvula termoiónica de vacío**. Se percató que si se colocaba una pila con el negativo en el filamento y el positivo en el electrodo interno (ánodo), fluía una corriente eléctrica; pero si este proceso se invertía, la corriente dejaba de fluir. Es decir, la válvula electrónica rectificaba la corriente. Este dispositivo se conoció con el nombre de **diodo rectificador** [32]. Dos años después el físico estadounidense Lee de Forest agrega un tercer electrodo reticulado al dispositivo de Fleming y surge el **tríodo**. Este dispositivo marcó el nacimiento definitivo de la tecnología electrónica. A partir de ahí, la electrónica de tubo de vacío garantizó un gran desarrollo en las telecomunicaciones y la computación. Sucesivamente el 8 de octubre de 1926 Julius Edgar Lilienfeld, científico de origen Alemán exiliado en EEUU, patenta el esquema de un «dispositivo para controlar el flujo de una corriente eléctrica entre dos terminales de un sólido eléctricamente conductor, mediante el establecimiento de un tercer potencial entre las terminales anteriores» (figura 2.3) [45].

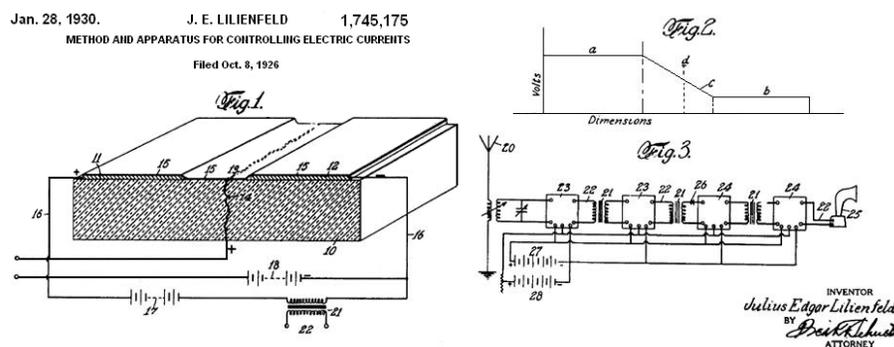


Figura 2.3. Dibujo de la patente de Lilienfeld, en la que describe por primera vez un FET.

En aquel momento Lilienfeld no estaba consciente del alcance que a largo plazo iba a tener su descubrimiento, que no era otro que el Transistor de Efecto de Campo (Field Effect Transistor, FET). Tres años después perfeccionó su invento incluyendo una capa de aislante entre el contacto modulador y el material conductor [45]. Había nacido lo que posteriormente sería el MOSFET (Metal-Oxide-Semiconductor FET). Sin embargo, las dificultades técnicas que suponía la fabricación de este dispositivo, estaban por encima, tanto de la tecnología de la época como de los conocimientos de la física de semiconductores que en ese entonces se tenían. Lo cual provocaron que el MOSFET debiera esperar más de treinta años hasta verse físicamente realizado de manera operativa.

La electrónica estuvo basada en dispositivos de vacío hasta la segunda guerra Mundial pero debido a las exigencias de la tecnología, el tubo de vacío presentó problemas tales como: tamaño, peso, calentamiento y el alto consumo de energía, lo cual era un conjunto de problemas que había que resolver y, para ello se contaba con grandes conocimientos científicos sobre los materiales semiconductores, como el silicio y el germanio adquiridos durante el desarrollo del radar. Después de la interrupción debido a la segunda guerra Mundial, sucedieron precipitadamente una serie de descubrimientos en el campo de los dispositivos electrónicos de estado sólido, que iban a revolucionar el mundo tal y como en ese entonces era conocido. En 1947, un equipo conformado por Stanley Morgan, William Shockley, John Bardeen y Walter Brattain, descubren el **efecto transistor**, mediante el cual inventaron el **dispositivo transistor**, (un acrónimo de transfer resistor) (figura 2.4) [32] [47]. El anuncio público se hizo en 1948, y con él se conocieron las primeras grandes aplicaciones: como oscilador, amplificador y conmutador. De inmediato, exhibió sus ventajas sobre el dispositivo electrónico que le antecedió, el tubo de vacío: calentamiento mínimo, menor consumo de energía, mayor velocidad, menor costo, menor tamaño y facilidad de producción en serie, todo lo cual contribuyó a crear la era de la miniaturización de los voluminosos dispositivos electrónicos de tubos de vacío que funcionaban en la época.

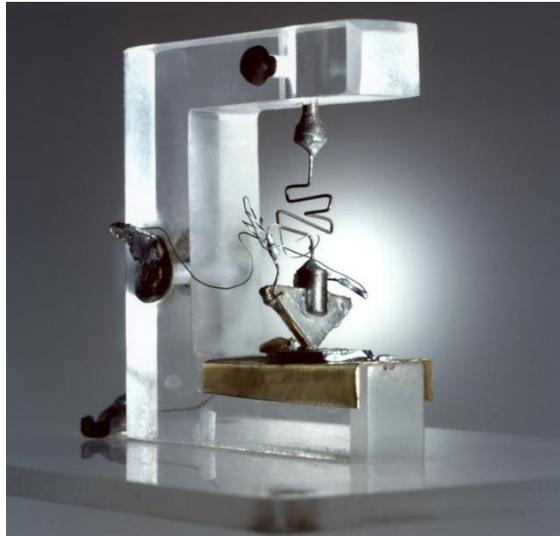


Figura 2.4. Fotografía del primer transistor. Los cristales de germanio eran pequeñas piezas en forma de cilindro (emisor) conectada al triángulo (base) y a la placa inferior (colector). Los demás elementos son de soporte.

La figura anterior es un transistor tipo **pn** de germanio en el cual las uniones **pn** son reemplazadas por diodos tipo Schottky, la patente de este descubrimiento se muestra en la figura 2.5.

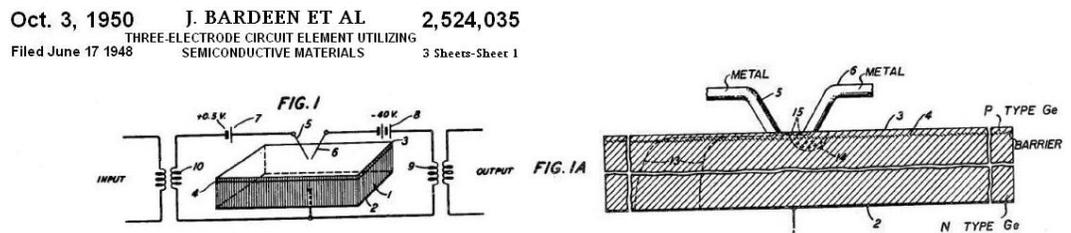


Figura 2.5. Dibujo de la patente de Bardeen y Brattain del transistor de punta de contacto.

Poco después, Schokley completó y perfeccionó la idea de Bardeen y Brattain [48]: había nacido el transistor bipolar de unión (Bipolar Junction Transistor, BJT). En 1954, Texas Instruments presenta el primer transistor comercial de silicio. Ese mismo año, en los laboratorios Bell se trabajaba en el desarrollo de procesos tecnológicos básicos, para el desarrollo de los circuitos integrados, como son la oxidación, el grabado, la difusión, etc. En 1958 Jack Kilby inventa el circuito integrado el cual se muestra en la figura 2.6.

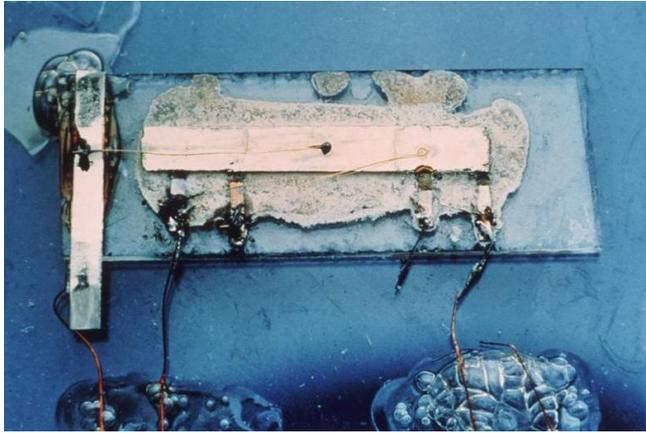


Figura 2.6. Fotografía del primer circuito integrado fabricado por Kilby, el cual contenía cinco componentes integrados. Publicado con permisos de Texas Instruments.

Este circuito integrado contaba de 2 transistores tipo PNP, 8 resistencias y 2 capacitores como se muestra en el diagrama eléctrico de la figura 2.7.

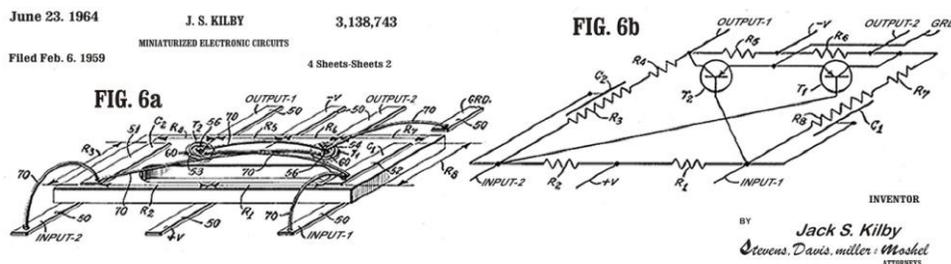


Figura 2.7. Patente del circuito integrado fabricado por Jack Kilby.

Un año después, Noyce sienta las bases de la tecnología planar, que abre el camino al desarrollo de circuitos complejos. Una vez iniciado el proceso de integración, la atención de la industria se orientó en la dirección de los circuitos integrados, en los que el transistor es la parte fundamental. La integración de grandes cantidades de diminutos transistores en pequeños chips, fue una enorme ventaja sobre el ensamblaje manual de los bulbos y circuitos utilizando componentes discretos. La confiabilidad y la capacidad de producción masiva de circuitos integrados (CI) impuso la estandarización de estos, sobre los diseños basados en transistores, que pronto dejaron obsoletos a los tubos de vacío (bulbos). Este paso de la electrónica a la microelectrónica puede definirse como la transición de los componentes individuales al CI, compuesto por un chip o pequeño cristal de silicio, que contiene gran cantidad de elementos individuales en forma miniaturizada. Con esta técnica de integración se logró una mayor complejidad de circuitos, menores dimensiones, mayor fiabilidad, a un coste de fabricación más reducido.

La historia de la microelectrónica se ha escrito por la sucesión de diversas tecnologías, según el número de transistores integrados en el mismo chip. La primera familia de circuitos digitales identificada con las siglas SSI (Small Scale Integration), fue puesta al mercado en el año de 1965. Consistió principalmente de compuertas lógicas, flip flops y funciones lógicas básicas (and, or, not, nand, etc). Los circuitos de la serie 74 de Texas Instruments son exclusivos de esta familia, que integró varias decenas de transistores por paquete. Sólo tres años después estuvo disponible la segunda familia: la MSI (Medium Scale Integration), en la que ya se había conseguido empotrar varios cientos de transistores en un solo paquete, dando lugar a circuitos con funciones mucho más complejas: contadores, codificadores y de selección; así como memorias de baja y mediana capacidad, fueron comunes en esta familia. Los circuitos integrados fueron incrementando, año con año, y la densidad de integración disminuyendo el tamaño del transistor. La siguiente familia: la LSI (Large Scale Integration), apareció en 1972; contenía varios miles de transistores por circuito y los elementos más característicos consistían en memorias de alta capacidad y microprocesadores (unidades de control) de 4 y 8 bits. Un ejemplo típico fue el procesador 4004 de Intel. El éxito obtenido con este procesador produjo un nuevo mercado y un panorama distinto en la tecnología de integración. Esta tendencia a la miniaturización, exigió la aplicación de técnicas más sofisticadas, con el objeto de lograr una resolución cada vez más fina para el grabado de los circuitos. Bajo este contexto nace la integración a muy grande escala (VLSI, Very Large Scale Integration) en 1980. Rápidamente aparecieron los microprocesadores VLSI; en definitiva, el nacimiento de estos dio paso a la fabricación masiva de microcomputadoras personales (PC, Personal Computer), a partir de la década de los 80's. En el año de 1991 fue lanzado el **procesador Intel Pentium** ocupaba una superficie de 0.8μ y contaba con 3 millones de transistores. Ocho años después la cantidad de transistores pasó a ser tres veces mayor (9.5 millones) con el Procesador Intel Pentium III. Al comienzo del nuevo milenio, Intel se encontraba desarrollando procesadores de 90nm que brindaban mayor desempeño y menor consumo energético. Fueron los primeros procesadores fabricados en silicio. En 2005 hizo su aparición en el mercado el **primer procesador multinúcleo**: el **Intel Pentium D**. Fue el comienzo de la tecnología **Dual-Core** en microprocesadores, que un año después llevó al desarrollo del procesador Intel **Core 2 Duo**. Su proceso de fabricación de 65nm le permitió llegar a disponer de 290 millones de transistores. Los avances en tecnología multicore siguieron su curso y los procesadores de cuatro núcleos Intel **Core 2 Quad** hicieron su aparición en 2007. De esta manera, el liderazgo de Intel una vez más se tradujo en mayor desempeño para el usuario, capaz ahora de disfrutar al máximo de la experiencia multimedia. Por fin, llega la era de los 45nm en 2008, dado a conocer bajo el nombre de **Nehalem**. Los procesadores más pequeños del mundo, que utilizan Hafnio como componente, se implementaron en los nuevos desarrollos de Intel Core 2 Duo, Intel Core 2 Quad, Intel **Xeon** y la línea de reciente aparición, los microprocesadores **Ivy Bridge** e **Intel Atom SoC** de 22 nm y mas de 2300 millones de transistores, estos procesadores tienen 80,000 veces más transistores que el 8086 lanzado hace mas 31 años [URL-12]. Las innovaciones en arquitectura informática han logrado que un poder mucho mayor al del gran coloso ENIAC habite en un microchip más

pequeño que la punta de un dedo. Así, los Procesadores Intel de 22nm son el más reciente eslabón del avance tecnológico en el que Intel continúa haciendo historia [102][URL-20].

| Escala de Integración | Abreviatura | Numero de Transistores | Año |
|-----------------------|-------------|------------------------|------|
| Cero | ZSI | 1 | 1947 |
| Pequeña | SSI | < 12 | 1965 |
| Mediana | MSI | 13 - 99 | 1968 |
| Grande | LSI | 100-9999 | 1972 |
| Muy Grande | VLSI | 10000-99999 | 1980 |
| Ultra Grande | VLSI | >100000 | 2009 |

Tabla 2.1. Escalas de integración para circuitos integrados

En la tabla 2.1 se resumen las distintas tecnologías de integración para CI y en la figura 2.8 se muestra el desarrollo de los microprocesadores.

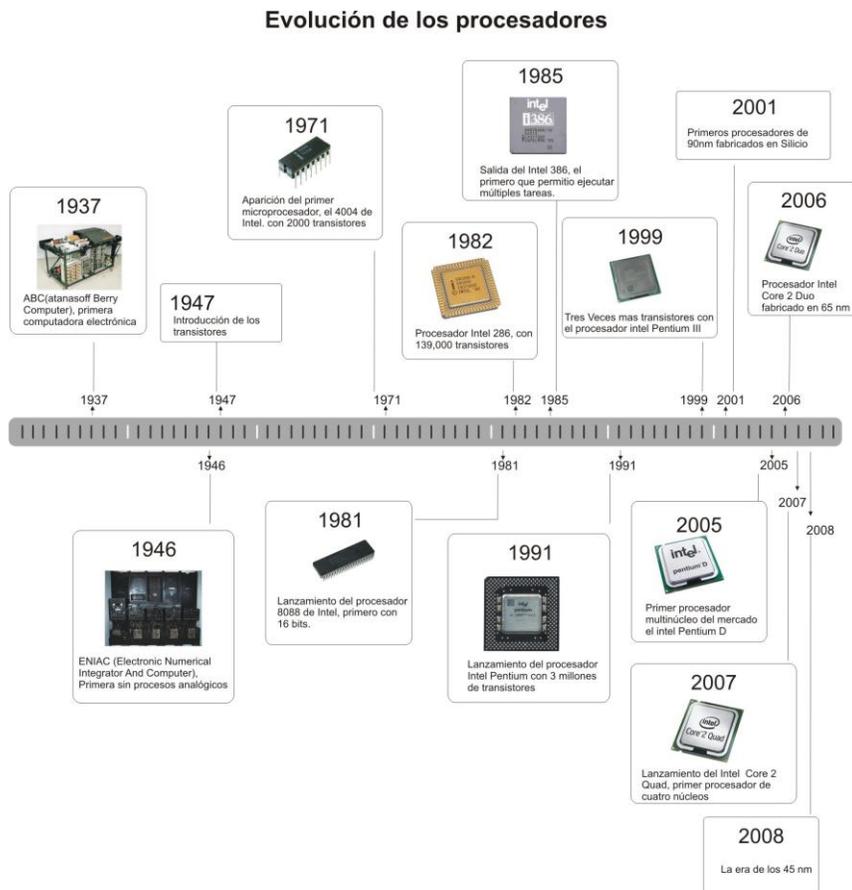


Figura 2.8. Evolución de los procesadores [102].

Estas tecnologías de integración han originado el desarrollo del microprocesador desde el 4004 de INTEL hasta el K8L de AMD y el Kentwfield de INTEL, como se muestra en la figura 2.9.

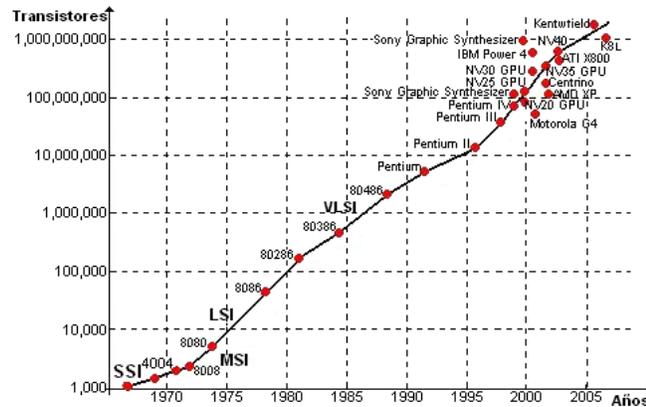


Figura 2.9. Desarrollo del microprocesador basado en las distintas tecnologías de integración. Publicado con permiso del autor [59].

2.3.3. Modelos temporales de desarrollo tecnológico y científico.

La Teoría para la Resolución de Problemas Inventivos (TRIZ) fue desarrollada en Rusia en un espacio de 40 años, por el científico Ruso Genrich Altshuller en la que analizó cerca de 1.5 millones de patentes de diferentes áreas de la tecnología. Estas patentes presentaban soluciones a difíciles contradicciones tecnológicas y fueron usadas para definir y clasificar la naturaleza de los problemas de inventiva. La base de TRIZ se funda en tres ideas:

- Todos los sistemas técnicos se desarrollan de acuerdo a leyes determinadas. La meta de este desarrollo es: la aspiración de encontrar la solución ideal de las funciones requeridas.
- Las soluciones creativas que están unidas a la eliminación de las contradicciones y que envían ideas para la realización de nuevos sistemas, se llaman inventos. Observado el origen de estos, desde el punto de vista de su sucesión cronológica, refleja la evolución en un sistema técnico. Por eso el desarrollo de los inventos está sujeto a determinadas regularidades, estas fueron descubiertas y acumuladas en base a la investigación de cientos de miles de inventos. El proceso de ésta investigación aun continúa.
- Una persona que se encuentra ante un problema, que tiene fuertes contradicciones físico-técnicas, derrocha mucho tiempo y energía en la búsqueda de ideas casuales, en estas circunstancias se encuentra en un camino de estereotipo psicológico no efectivo. En la teoría de TRIZ se desarrollaron algoritmos lógicos para eliminar los estereotipos ineficientes del pensamiento; además pone a disposición los principios fundamentales y cientos de procesos para la solución de problemas con fuertes contradicciones físico-técnicas.

El conocimiento representado por esas innovativas patentes fue la base del desarrollo de un método para resolver problemas tecnológicos y desarrollar productos y procesos. TRIZ comprende un grupo de regularidades, algoritmos y herramientas. Con esta metodología un individuo puede desarrollar sistemas tecnológicos en una forma sistemática a través de un proceso que comienza con la identificación del problema, su categorización, su formulación y finalmente el uso de las herramientas para generar conceptos de soluciones creativas. El desarrollo de TRIZ se fundamenta principalmente en que es un método basado en tecnología, no en psicología, como lo son el método de **tormenta de ideas** y el **método de prueba y error**. El método psicológico de prueba y error consiste en proponer alguna solución y después comprobar su validez; tiene sus raíces en épocas antiguas. En esencia, este método es tan antiguo como el hombre. Altshuller desarrolló la metodología con el objetivo de que la creatividad en tecnología no dependa totalmente de factores psicológicos. El método de prueba y error no garantiza que el problema sea resuelto en corto tiempo. Tal vez la solución aparezca hoy, o quizá no aparezca en toda la vida. Se requiere un método diferente para desarrollar productos y procesos, un método basado en el uso de los **patrones de evolución de los sistemas tecnológicos**. La base teórica del TRIZ clásico son los patrones de la evolución tecnológica. La evolución tecnológica tiene sus leyes y características, es por eso que personas de diferentes países, trabajando en el mismo problema independientemente, obtienen la misma solución. El descubrimiento fundamental de Altshuller fue que identificó regularidades recurrentes o patrones en el desarrollo y evolución de los sistemas tecnológicos. Conociendo esas regularidades, la solución de problemas técnicos es mucho más fácil. Altshuller, basado en este descubrimiento, desarrolló un grupo de patrones (regularidades) que describen la evolución de la tecnología, tal como se presenta dentro de la colección de la literatura de patentes mundiales. Altshuller pensó que si se pudiera identificar el estado actual de un sistema dado, con la aplicación de esos patrones, se podría acelerar la evolución de ese sistema a su próxima generación. A continuación se presentan algunos conceptos básicos: el tamaño de las máquinas se ha incrementado muy rápido en la historia, sin embargo, el crecimiento no es ilimitado. Las máquinas han incrementado su tamaño dos, diez, cien veces. Pero llega un momento en el que un crecimiento posterior deja de ser económico y benéfico. Sin embargo, si dos máquinas se unen, aparece un nuevo sistema y este sistema empieza a crecer. Cada vez que un sistema X se combina con un sistema Y, un nuevo sistema XY emerge. Este nuevo sistema XY tiene, en principio, nuevas características y una nueva calidad que ni X ni Y tenían antes. Para formar un nuevo sistema, se deben unir objetos de tal manera que aparezca una nueva característica. El crecimiento de los sistemas mediante el desarrollo y la complejidad, para luego pasar a la simplicidad, es una ley universal. En el mundo técnico el desarrollo va de la célula al sistema. Cuando una célula sencilla se incorpora a un sistema, se desenvuelve más eficientemente y se desarrolla más rápido. La célula depende del sistema y no puede existir sin él. La tecnología contemporánea es una tecnología de sistemas. Sus células son los diferentes accesorios, máquinas, equipos que funcionan dentro del sistema. Existen cuatro periodos en la evolución de

un sistema técnico que lo dirigen hacia un incremento en idealidad, y cada periodo tiene sus propios problemas y métodos para encontrar soluciones. Estos periodos son los siguientes:

- 1er. Periodo: selección de partes para el sistema. Surgimiento del sistema.
- 2do. Periodo: mejoramiento de las partes. Corrección de partes malas.
- 3er. Periodo: dinamización del sistema.
- 4o. Periodo: auto-desarrollo del sistema. El paso hacia el autocontrol.

Este desarrollo se lleva a cabo a través de las siguientes leyes:

Leyes estáticas

- **ley de la unión complementaria de las partes de un sistema:** un sistema técnico surge como resultado de una síntesis de partes previamente separadas. Para que este sistema pueda funcionar, debe contener cuatro partes básicas: un motor o fuente de energía; un órgano de trabajo que lleve a cabo la función del sistema; una transmisión que conduzca la energía del motor al órgano de trabajo y un órgano de control a través del cual se controle el sistema.
- **ley de la conducción de energía en un sistema:** un sistema tecnológico evoluciona incrementando la eficiencia en la transmisión de energía de la máquina al órgano de trabajo. Esta transferencia puede tomar lugar a través de una sustancia (flecha, engranes, etc.); a través de un campo (magnético, térmico, etc.); o a través de un campo-sustancia (flujo de aire con partículas cargadas, etc.). La selección de la forma de llevar a cabo esta transferencia está en el corazón de muchos problemas de inventiva.
- **ley de la armonización de ritmos:** un sistema evoluciona hacia el incremento de la armonía de los ritmos y la frecuencia natural de sus partes. Como un ejemplo de este patrón Alshuller describe una patente para mejorar la extracción de carbón barrenando un hoyo en el yacimiento, llenándolo con agua y transmitiendo impulsos de presión a través de él para romper el carbón. Siete años después, otra patente mejoraba el proceso aplicando los impulsos a una frecuencia igual a la frecuencia natural de la masa de carbón. Alshuller señala que si se hubiera conocido la ley de armonización de ritmos, el retraso de siete años se hubiera evitado.

Leyes cinemáticas

- **ley del incremento de idealidad:** un sistema tecnológico evoluciona incrementando su grado de idealidad. La idealidad es definido como el cociente de la suma de los efectos útiles del sistema (U_i) dividido entre la suma de sus efectos perjudiciales (H_j). **idealidad** = $i = U_i/H_j$. Otros conceptos se derivan de esta ley, como lo son la máquina ideal y el resultado final ideal.
- **ley del desarrollo disparejo de los elementos del sistema:** esta ley establece que aunque el sistema mejore como un todo, las partes individuales del sistema no mejoran sincronizadamente. Cada sistema técnico tiene un sistema superior (supersistema), y al

menos un sistema subordinado (subsistema). Cualquier cambio en el sistema tiene efecto en ambos. Las contradicciones técnicas surgen porque los subsistemas no evolucionan uniformemente. Por lo tanto, es necesario considerar el interés no solo del sistema que necesita ser mejorado, sino también el interés del subsistema y el supersistema. Altshuller menciona el ejemplo de los barcos de carga, cuya capacidad de carga y su potencia se han incrementado rápidamente, pero el sistema de frenos no lo ha hecho, es por eso que un buque petrolero tarda varios kilómetros en detenerse.

- **ley de la transición a un supersistema:** cuando un sistema ha alcanzado el límite de su propio desarrollo, puede continuar creciendo convirtiéndose en un subsistema de un sistema más general. De esta manera el sistema original se eleva a un nuevo nivel. Por ejemplo, los sistemas de comunicación como el radio, el teléfono y la televisión pueden seguir evolucionando como subsistemas de un sistema integrado en una computadora.

Leyes dinámicas

- **ley de la transición del macronivel al micronivel:** ésta ley refleja la tendencia de evolución de los sistemas tecnológicos hacia la miniaturización de los componentes del sistema. Por lo tanto, un método para resolver muchos problemas es la transición de la macroestructura a la microestructura del sistema, tal como es el caso de la microelectrónica.
- **ley del incremento del dinamismo:** ésta ley indica la tendencia de la tecnología moderna a evolucionar desde estructuras rígidas, en las cuales sus partes están fijas, a estructuras más adaptables y flexibles en las que sus partes se vuelven móviles o ajustables.

Estas leyes indican la forma correcta de pensar cuando se quiere mejorar la conductividad de energía, la armonización de los ritmos, la transición al micronivel, etc., de los sistemas tecnológicos. Por ejemplo: Cuando el sistema alcanza sus propios límites se integra con otro sistema y emerge un sistema nuevo más complejo. A veces la integración entre sistemas se vuelve muy difícil. Para que el sistema siga evolucionando es necesario integrarlo con otro sistema y al mismo tiempo es imposible hacerlo. Este tipo de barrera puede ser removida descomponiendo el sistema existente y recombinando sus partes en un nuevo sistema. Otro problema que puede surgir es que el sistema no pueda ser cambiado. Entonces ¿Cómo se hace la transición a un nuevo sistema? Hay una solución: el nuevo sistema está "escondido" dentro del antiguo. En conclusión: hay dos direcciones para el desarrollo de sistemas que parecen usar todos los recursos de desarrollo:

1. la primera dirección es la consolidación del sistema existente con otros sistemas o descomponerse y recombinarse en otro sistema.
2. la segunda dirección es la transición de la macroestructura (macronivel) a la microestructura (micronivel) donde el mundo interno del sistema (partículas, moléculas, átomos y campos) se involucran en el juego.

La transición de los sistemas técnicos al micronivel es una ley. Pero la ley de evolución en el desarrollo de sistemas técnicos establece que el sistema debe agotar sus recursos antes de moverse al micronivel. La situación más ventajosa es cuando el sistema agota su crecimiento y necesita ser reemplazado por uno nuevo, basado en otros principios. Esas leyes son guías muy poderosas para pronosticar y acelerar la evolución de la tecnología. Inicialmente, cada patrón incluía una definición primaria y un medio de aplicación. Posteriormente se desarrolló una descripción más detallada llamada líneas de evolución. En la sección 2.3.4 se tratará más a detalle este tema.

El primer gran paso en la creación de TRIZ fue cuando Altshuller observó que todos los problemas de inventiva involucraban lo que él llamó una contradicción técnica. Una contradicción técnica existe cuando se quiere mejorar un parámetro X de un sistema tecnológico y esto deteriora un segundo parámetro Y. En otras palabras, los sistemas técnicos son similares a los organismos vivos, consisten de partes interrelacionadas. Si se cambia una parte del sistema, se tendrá un efecto negativo en las demás partes y se producirá lo que se llama una contradicción. Un problema de inventiva tiene siempre dos exigencias:

- se requiere que mejore una de las partes (o características) del sistema,
- sin deteriorar las otras partes (características) del sistema o todos los sistemas relacionados.

Esto es una **contradicción técnica**, y hacer un invento implica remover las contradicciones técnicas. El enfoque de ingeniería tradicional al tratar con tales contradicciones es el compromiso, los dos parámetros ceden y sacrifican su máximo nivel.

Con los conflictos y contradicciones que presentan los sistemas tecnológicos se pueden lograr ideas creativas para resolver problemas y generar nuevos productos y servicios. Hay tres niveles de conflictos y contradicciones: **administrativos**, **técnicos** y **físicos**.

Conflictos administrativos: son los que se presentan en la descripción general de los sistemas y no son de carácter técnico. Generalmente se relacionan con costos, tiempos de producción y entrega, administración del personal y similares; para resolverlos se pueden emplear técnicas de carácter psicológico como las propuestas por De Bono, VanGundy [101].

Conflictos técnicos: estos aparecen cuando al tratar de mejorar una característica del sistema se perjudica otra de las características del mismo sistema. Así que un conflicto técnico con frecuencia está relacionado con dos características del sistema; por ejemplo, si aumento el tamaño de una ventana aumenta el peso del vidrio.

Contradicciones físicas: estas son las causas de los conflictos técnicos. Aparecen cuando se hacen exigencias contradictorias sobre un elemento o un objeto del sistema tecnológico. Por ejemplo,

en el diseño de las latas para envasar cerveza se requiere que las paredes del recipiente sean gruesas y delgadas a la vez: gruesas para soportar el peso de otras durante el almacenamiento y para soportar la presión interior, y delgadas para disminuir su peso. El desafío del nuevo diseño es eliminar la contradicción física.

2.3.4. Modelos auxiliares del ciclo de vida de un sistema.

Thomas S. Kuhn en su libro «The Structure of Scientific Revolutions» [3], señala que las actividades científicas actúan en un área de observación estable y durante largos períodos, dentro de lo que se puede denominar períodos de ciencia normal. Períodos en los cuales no se pone en cuestión las bases estructurales de una ciencia, tampoco sus definiciones substanciales. Durante esos períodos pueden suceder cambios menores en la establecida superficie de toda investigación, pero esos cambios no afectan los aspectos que hacen «respetable» una teoría o un procedimiento científico.

Cuando en el desarrollo de la ciencia normal empiezan a suceder crecientes anomalías tanto en el nivel cuantitativo como en el cualitativo, y ésta ya no puede explicar el origen de esas anomalías, aumenta el estado de crisis dentro del mundo conceptual de la ciencia normal. Al final reina una situación incontrolable y ahí, ya no se puede poner en discusión las posibles soluciones que dejaron de serlo. Es entonces cuando la ciencia normal pierde sentido y da lugar a la «ciencia revolucionaria» [3]. Antes de una revolución, acontece una sorprendente ruptura, algo totalmente nuevo parece dirigir los cambios dentro de una especie de desastre inicial que resquebraja el mundo conceptual científico que gobernaba antes. Ese cambio radical se convertirá finalmente en otro período de ciencia normal, pero basado en un **nuevo paradigma**. Un nuevo paradigma dominante no cede hasta que las evidencias incompatibles se acumulan; al grado de que ya no es posible sustentar al antiguo paradigma. Por iniciativa y juicio de un grupo de científicos innovadores se da la transformación o cambio del paradigma, pero siempre con gran resistencia por parte de la comunidad académica y científica. Esta nueva teoría, no ofrece necesariamente información adicional a la existente, sólo es más justificable ante las nuevas evidencias, y permite un número mayor de predicciones acertadas dentro del campo de estudio; es decir, simplemente predice con mayor exactitud los hechos que ahora han cobrado mayor relevancia, aunque no necesariamente aquellos que eran relevantes con el paradigma anterior.

El desarrollo de la ciencia normal tiene un comportamiento similar al modelo de «ciclo de vida», el cual corresponde al clásico ciclo de **nacimiento, desarrollo, madurez y muerte**. De esta forma, se tiene que este modelo como cualquier otro paradigma, es solo un convencionalismo que sustenta el desarrollo de la ciencia normal. En algunas ocasiones este modelo ha sido desafiado, porque aparentemente existen sistemas que resultan ser contradictorios con este ciclo, mismas que se han corregido con el apoyo de hipótesis auxiliares [30]. La figura 2.10 muestra

algunas hipótesis auxiliares, que tratan de corregir las anomalías detectadas alrededor de este modelo. La figura 2.10a corresponde al ciclo de vida que siguen sistemas pequeños como electrodomésticos; la figura 2.10b representa el ciclo de vida para algunos medicamentos; y la gráfica 2.10c se refiere a las sucesiones de los ciclos de vida para explicar el comportamiento de la tecnología de fabricación de los dispositivos semiconductores.

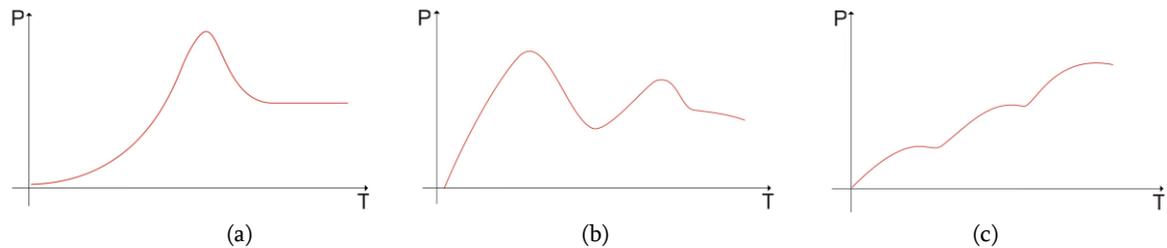


Figura 2.10. Ciclo de vida de un sistema. (a) crecimiento-caída-madurez, (b) cíclico decreciente, (c) escalonado.

Del modelo mostrado en la figura 2.10c, se derivan dos categorías especiales para representar el ciclo de vida de un sistema como se observa en la figura 2.11.

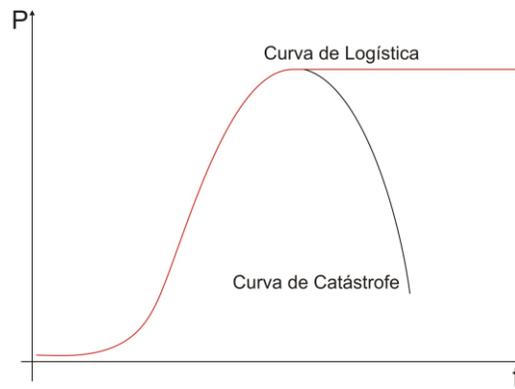


Figura 2.11. Modelos del ciclo de vida derivados de la figura 2.10c.

Este modelo en ocasiones se asocia con la **curva logística**. Cuando este es el caso, el lado derecho de esta curva tiene una asíntota horizontal. Otras veces se representa como la **curva de catástrofe** que describe el fin de un desarrollo, que ha tenido un comienzo gradual, suave, pero que sin embargo se prevé que irá al final, hacia una completa extinción, tal vez abrupta. La figura 2.12 muestra las cuatro etapas de este modelo. Al principio, el desarrollo del sistema es lento (etapa I), en cierto punto hay un incremento (etapa II), después de esto el desarrollo se reduce antes de que el crecimiento del parámetro se termine (etapa III); esto significa la aparición de contradicciones. A veces los parámetros empiezan a decrecer; lo que manifiesta la muerte del sistema (etapa IV). Esta curva es por lo general llamada «curva S» [30].

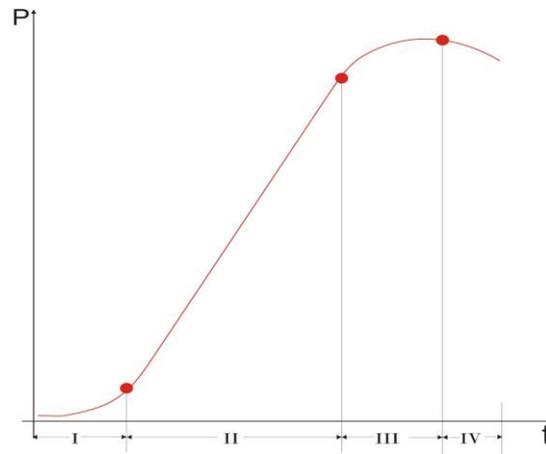


Figura 2.12. Etapas del ciclo de vida de un sistema.

El concepto de curva S, fue introducido por Richard Foster en el libro titulado «Innovation» en el año de 1986, para describir el funcionamiento de una tecnología que varía en un cierto plazo [31]. Desde entonces este modelo ha tenido gran aceptación dentro de la comunidad científica, tecnológica y empresarial para predecir el comportamiento de un sistema.

Para el análisis realizado en esta investigación, se acepta el modelo de curva S como una **función no decreciente** como se analizará en el capítulo 4. Donde las características esenciales son: **punto de inflexión** y **restricciones horizontales**. Para el caso del modelo logístico, estas restricciones horizontales son sus asíntotas. Como muestra la figura 2.13.

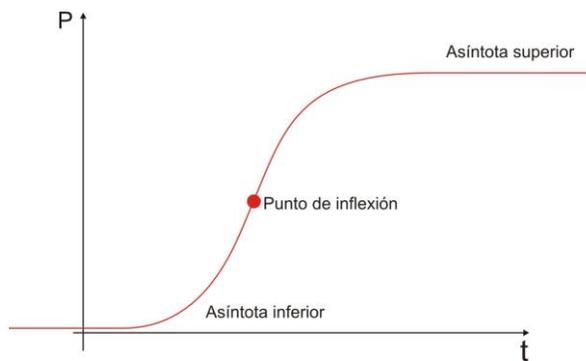


Figura 2.13. Curva S. Asociado a un polinomio cúbico o a la función logística.

Cuando un cierto límite de evolución cuantitativa se alcanza, el sistema experimenta cambios cualitativos o es sustituido por otro sistema. La nueva calidad aumenta el índice de crecimiento. Durante este proceso, los cambios cuantitativos (ciencia normal) ocurren continuamente, mientras que los cambios cualitativos (ciencia revolucionaria) [3] ocurren en pasos discretos. La ciencia normal crece en la etapa de desarrollo del sistema y disminuye en el momento de

desuso. El final del crecimiento de un sistema no significa el final de la evolución en este campo. Los nuevos sistemas más avanzados cobran vida, un salto en la evolución ocurre. La evolución adicional del sistema comienza según una nueva curva S. Tal proceso se observa en la figura 2.14.

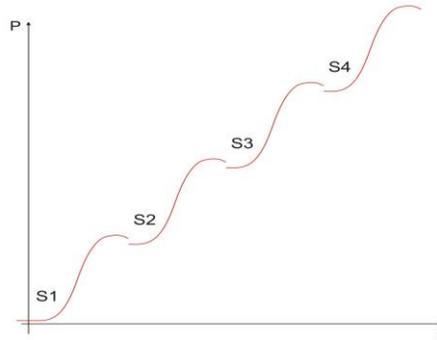


Figura 2.14. Evolución de un sistema a salto.

El sistema S2 aparece en vez del sistema S1. La evolución continúa, del mismo modo aparece el sistema S3, después el sistema S4 etc.; el avance entero en el campo puede ser visto como una curva tangente a las curvas S. Un ejemplo de éste comportamiento se observa en la figura 2.15, en la que se muestran los diferentes dispositivos electrónicos descritos en la sección 2.3.2. En esta figura se aprecia que los cambios cuantitativos (ciencia normal) suceden en pasos continuos, mientras que los cambios cualitativos (ciencia revolucionaria) suceden en pasos discretos, y la evolución de la ciencia electrónica puede verse como una curva tangente, al conjunto de curvas S.

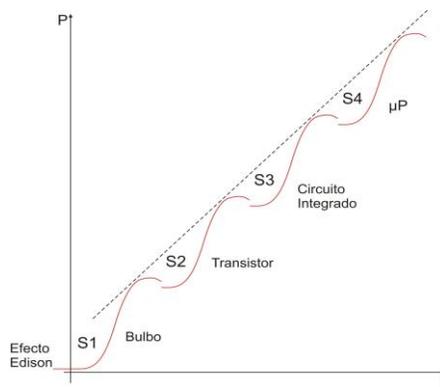


Figura 2.15. Curva tangente al conjunto de pequeñas curvas S que muestran la evolución de los dispositivos electrónicos.

2.4. Formulación original de la ley Moore.

La evolución tecnológica de los circuitos integrados, está resumida en la denominada **ley de Moore**. Esta ley es un razonamiento formulado el 19 de abril de 1965 por el Dr. Gordon E. Moore, en su artículo titulado, «**Cramming more components onto Integrated Circuits**» publicado en el 35 aniversario de la revista «Electronics» [49]. En este artículo Moore realiza dos argumentos distintos, que sin embargo están enlazados. El primero está basado en el costo de producción de los CI con distinta complejidad. Para fundamentar este argumento, Moore graficó las curvas de costos de producción durante dos años; 1962 y 1965, de las cuales concluyó que el costo por componente era alto si había muy pocos o también si el número de componentes en el chip era extremadamente alto, como se muestra en la figura 2.16. En 1962 el número óptimo de componentes era aproximadamente 10, en tanto que para 1965 era alrededor de 50. Extrapolando esta tendencia, Moore notó que para 1970, el costo de fabricación por componente sería de sólo una décima parte del costo actual, y que un chip con un costo mínimo podría tener hasta 1,000 componentes.

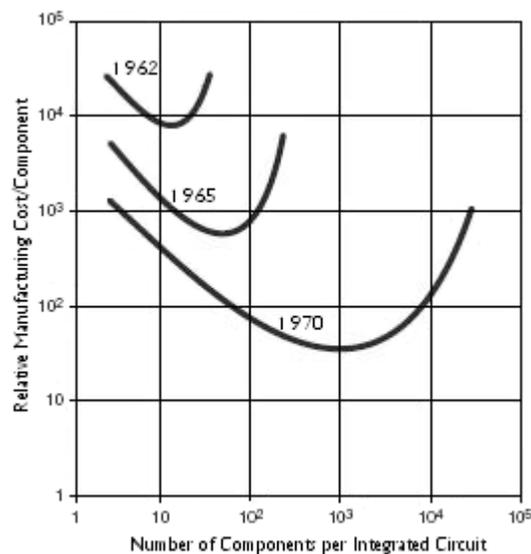


Figura 2.16. Grafica costo-componente del artículo original de Moore: Cramming more components onto Integrated Circuits.

En el segundo argumento, Moore analiza los chips existentes; comenzando desde el CI de Jack Kilby en 1959. Moore observó que la complejidad para los costos mínimos por componente aumenta a un factor de dos por año como muestra la figura 2.17.

A corto plazo se espera que esta tasa se mantenga, en tanto que a largo plazo la tasa de incremento resulta ser incierta, aunque no hay razón para creer que no se mantendrá casi constante, cuando menos por 10 años. Eso significa que para 1975 el número de componentes por circuito integrado será 65,000 para lograr el costo mínimo [49].

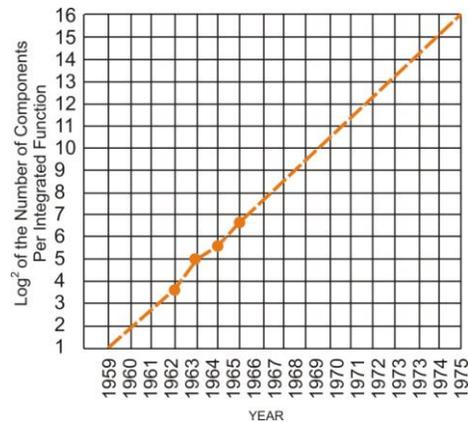


Figura 2.17. Gráfica del número de componentes por función integrada.

Es útil observar que la estimación de Moore fue basada en su creencia que no había embotellamientos importantes que retrasarían este proceso. De hecho, conjeturó que toda la tecnología requerida estuviera ya disponible:

“En la oblea de silicio que se usa actualmente, por lo general con un diámetro de una pulgada o más, hay amplio espacio para una estructura de este tipo, si los componentes se pueden empaquetar uno cerca del otro sin que se desperdicie espacio en los patrones de interconexión. Esto es realista, porque ya se realizan esfuerzos por lograr un nivel de complejidad por arriba de los circuitos integrados disponibles en la actualidad, utilizando patrones de metalización de múltiples capas separados por películas dieléctricas. Tal densidad de componentes se puede lograr con las técnicas ópticas actuales y no requiere las técnicas más exóticas” [49].

Similarmente, Moore argumentó que no había razones fundamentales por las que la producción no podría ser mejorada. Según Moore, no era necesario hacer alguna investigación elemental o sustituir los procesos actuales. **“Solamente el esfuerzo de ingeniería es necesario.”** Moore concluyó: **“Claramente, podremos construir tal equipo atestado de componentes”**. En su forma original la ley de Moore estableció que: **el número de elementos que la industria sería capaz de colocar en un chip se duplicaría cada 12 meses.**

Más adelante Moore notó que las tecnologías existentes podrían mantener esta tendencia para la próxima década, si los esfuerzos de ingeniería condujeran a producciones mejoradas en el

proceso industrial. La discusión de Moore no contempló explícitamente gastos de inversión. Asumió implícitamente que los gastos de infraestructura y financiamiento de la producción industrial del semiconductor eran fijos e insignificantes.

Desde el punto de vista económico, la ley de Moore es una ley bastante fascinante. Esto implica que el desarrollo de los circuitos integrados, fue completamente determinado por los costes de fabricación. La ley de Moore, por lo tanto, define completamente una nueva economía. En esta economía, la demanda es infinita. En el capítulo 4 se propone una generalización de la ley de Moore tomando en cuenta el tamaño del transistor y la tendencia de miniaturización.

Capítulo 3

Metodología de construcción y algoritmo de la Ontología.

En la sección 2.2 se presentaron algunos métodos para el rescate de información, entre ellos se encuentra la bibliometría y la cienciometría. Estos métodos son laboriosos y requieren que se disponga de una gran cantidad de artículos, es por ello que lo utilizan grandes instituciones científicas, organizaciones gubernamentales o compañías con capacidad financiera. A continuación se expone un enfoque basado en el análisis de resúmenes y para demostrar su utilidad se aplica a un cuerpo de documentos sobre computación PCDS. El trabajo se orienta a la creación de una metodología para investigar el desarrollo de una rama específica del conocimiento.

3.1. Análisis de la evolución de computación PCDS analizando textos cortos.

Para este objetivo se coleccionaron 2300 resúmenes relacionados con computación paralela, concurrente, distribuida y simultánea durante un período de 15 años. Cada resumen contiene un promedio de 200 palabras, por lo que son considerados «**objetos de conocimiento pobre**» [4]. Para convertirlos en «**objetos de conocimiento enriquecido**», se realizaron acciones tales como agruparlos en base a atributos externos (fecha de publicación), y de ésta forma se obtuvieron documentos con mayor cantidad de palabras, eliminando el problema expuesto anteriormente.

3.1.1. Muestreo para modelos con conocimiento científico.

Para conocer el estado futuro del hardware y software (HSW) se propone evaluar la dinámica de las publicaciones en distintas direcciones; el inconveniente es que la información relacionada es muy extensa, lo que impide recopilar un porcentaje relativamente grande de información, para obtener un muestreo representativo, que permita realizar una investigación más real. Es por ello que el estudio se delimita a computación PCDS, ya que en este dominio se manifiesta la relación intrínseca entre el HSW y se observa un hecho adicional: que el desarrollo de ambas áreas es necesariamente paralelo, lo que implica que son sistemas que trabajan de modo cooperativo.

Cada rama tiene un período característico para producir un cambio sustancial. Para computación PCDS, es de 6 años [15]; además tiene una historia de desarrollo alrededor de 30 años y en los últimos 10 se ha desarrollado intensamente. La información relacionada con esta área se encuentra disponible gratuitamente en Internet, en sitios como ACM [URL-5], CiteSeer [URL-8], IEEE [URL-6], Research [URL-9], Springer [URL-7], etc. Tomando en cuenta la disponibilidad de esta información, se coleccionaron 2300 resúmenes de la Librería Digital de IEEE [URL-6] para un período de desarrollo de 15 años (1990 al 2004). La diversidad de resúmenes disponibles y la velocidad de acceso, fueron los factores determinantes para seleccionar el cuerpo de textos de la Librería Digital de IEEE.

Para conocer el porcentaje de resúmenes disponibles en Internet, se propone la siguiente estimación: la exploración de información se realizó con la ayuda de **buscadores**, también conocidos como **motores** o **robots de búsqueda**, los cuales tienen por objeto detectar la información existente en Internet, utilizando palabras clave. Sin embargo, los resultados obtenidos difieren ampliamente según el buscador utilizado, ya que cada uno de ellos tiene su propio sistema de indexación. Entre los buscadores más conocidos se encuentran:

- Google
- Yahoo
- HotBot
- AltaVista

Durante mucho tiempo Yahoo y Altavista se disputaron el honor de tener el mejor sistema de búsqueda; Altavista contaba con un motor de búsqueda bastante eficiente, mientras que Yahoo acometía vanamente, la monumental e imposible tarea de ordenar en categorías los heterogéneos sitios. Naturalmente el propósito de Yahoo fracasó, en tanto que Altavista fue reemplazado gradualmente por Google. Este último funciona con una combinación única de hardware y software avanzado. La velocidad que se experimenta puede ser atribuida en parte a la eficiencia del algoritmo de búsqueda, y en parte a la gran cantidad de PC's que se conectan en red, para crear una máquina de búsqueda super rápida. Al mismo tiempo incorpora una serie de técnicas para mejorar la calidad de búsqueda; incluyendo el «PageRank», corazón del software. La razón principal de utilizar Google se debe a que proporciona los resultados más relevantes en un **tiempo de acceso** mucho menor que cualquier otro motor de búsqueda.

El motor de búsqueda Google muestra cerca de 1,050,000 registros al introducir las palabras «**artículos sobre computación PCDS**», entre los cuales aparecen publicaciones científicas, anuncios de congresos, invitaciones, directorios, centros de investigación, talleres, portadas, donaciones, proyectos de investigación, librerías virtuales, productos para computación, etc. Si ahora, la búsqueda se delimita y se emplean solamente las cuatro palabras que conforman el dominio PCDS; se obtiene un total de 400,000 registros que contienen información relacionada con el tema. De estos 400,000 registros se exploraron los primeros 1000, de los cuales 50 resultaron ser artículos científicos sobre computación PCDS (esto representa el 5%). Extrapolando se tiene que de los 400,000 registros, 20,000 corresponden a resúmenes sobre computación PCDS. En la figura 3.1 se muestra esta estimación.

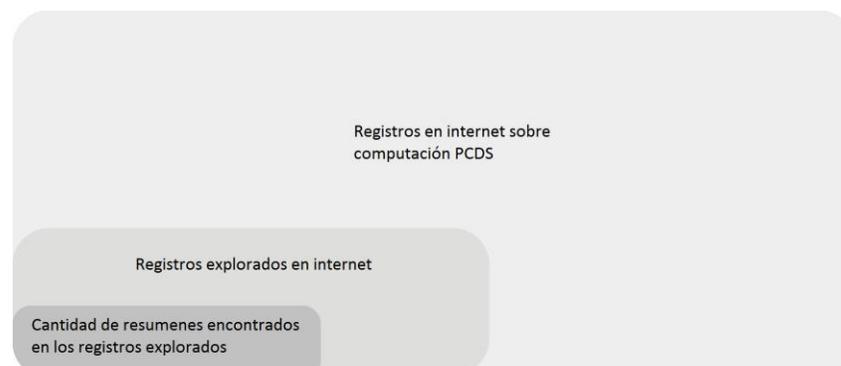


Figura 3.1. Estimación para explorar la información sobre computación PCDS existente en Internet.

De esto se deduce que la información coleccionada en la librería digital de IEEE, representa el 11.5 % de la información en inglés existente en Internet. Esto no significa que el muestreo representa el 11.5 % de toda la información publicada para computación PCDS, debido a que existen publicaciones en diferentes idiomas (ruso, chino etc). Además, es difícil detectar lo que está sucediendo con relación a una gran parte de la información relevante, ya que ésta circula a través de los llamados **colegios invisibles**, esto es, entre grupos de expertos, profesionales o académicos de diferentes países que se comunican entre sí, mediante el correo electrónico, o también se encuentra en forma de **literatura gris**, es decir, en documentos de difícil acceso que no se distribuyen a través de los canales de difusión convencionales, tales como tesis doctorales, actas de congresos, documentos de trabajo, etc. En términos generales y para fines prácticos se puede decir que el muestreo obtenido, es representativo.

En la selección de resúmenes de la librería digital de IEEE, el criterio empleado fue que apareciera una o más palabras clave (paralelo, concurrente, distribuido o simultáneo), o información relacionada con ellas (multihilos, exclusión mutua, PVM, MPI, hipercubo, etc.) ya sea en el nombre del artículo o en el cuerpo del resumen, sin tomar en cuenta, si pertenecían al área del hardware o software. Fue necesario depurar y ordenar la información debido a que presentaba detalles como por ejemplo palabras unidas o mal escritas que originaba cambios en los resultados.

La figura 3.2 muestra el archivo D69-2002 que corresponde al resumen número 69 de computación Distribuida, publicado en el año 2002. Se observa en ésta figura que el texto presenta errores ortográficos (marcadas en color amarillo) y palabras unidas (marcadas en color rojo). En la figura 3.3 se muestra cada una de las correcciones realizadas al archivo D69-2002, las cuales se resaltan en color verde. Estos detalles se realizaron a cada uno de los 2300 archivos.

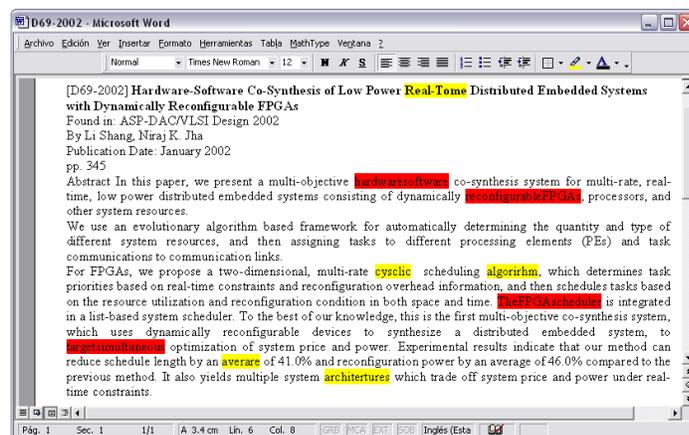


Figura 3.2. Errores comunes en archivos de texto plano. El archivo D69-2002 muestra errores ortográficos (resaltadas en color amarillo) y palabras unidas (resaltadas en color rojo).

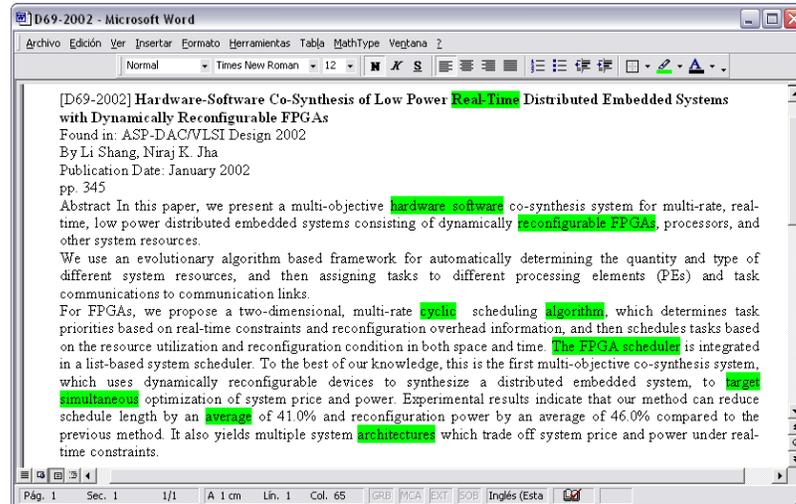


Figura 3.3. Corrección de los errores del archivo D69-2002.

Los resúmenes disponibles en la Librería Digital de IEEE están en formato PDF, por lo que es necesario pasarlos a una representación de texto plano (formato txt). El «block de notas» de Windows es la herramienta adecuada para realizar este trabajo. Después de convertirlos a texto plano, se agrupan como:

- archivos individuales,
- por año,
- en dos período,
- todos los resúmenes en un solo archivo.

El primer caso consiste en nombrar cada uno de los 2300 resúmenes de forma individual como se mencionó en la figura 3.2. El tamaño promedio de cada archivo es 1.2 KB.

Para el segundo caso, los resúmenes se agrupan en base a atributos externos como es su año de publicación. En la figura 3.4 se observan los 15 archivos integrados.

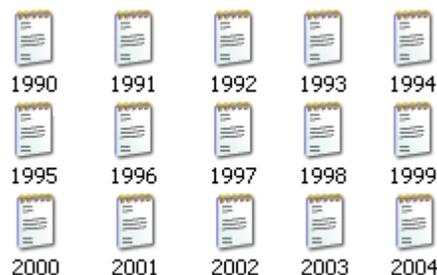


Figura 3.4. Archivos agrupados por año (segundo criterio).

El tamaño de cada archivo varía en función del número y tamaño de cada resumen. En la tabla 3.1 se muestran las características de cada archivo.

| Año | Resúmenes | Tamaño (KB) | Año | Resúmenes | Tamaño (KB) | Año | Resúmenes | Tamaño (KB) |
|------|-----------|-------------|------|-----------|-------------|------|-----------|-------------|
| 1990 | 50 | 57.0 | 1995 | 142 | 180.0 | 2000 | 215 | 271 |
| 1991 | 53 | 65.1 | 1996 | 163 | 200.0 | 2001 | 217 | 262 |
| 1992 | 58 | 72.5 | 1997 | 212 | 257.0 | 2002 | 221 | 261 |
| 1993 | 62 | 78.5 | 1998 | 118 | 145.0 | 2003 | 225 | 260 |
| 1994 | 67 | 90.7 | 1999 | 187 | 226.0 | 2004 | 229 | 252 |

Tabla 3.1. Características de los resúmenes de computación PCDS agrupados por año de publicación.

En el tercer caso, los resúmenes se agrupan en dos períodos, el primero corresponde a los resúmenes publicados de 1990 a 1997, mientras que el segundo de 1998 al 2004. En la tabla 3.2 se muestran las características.

| Período | Artículos | Tamaño en KB |
|-----------|-----------|--------------|
| 1990-1997 | 807 | 1000 |
| 1998-2004 | 1412 | 1677 |

Tabla 3.2. Características de los resúmenes agrupados por períodos

Para el último caso, se agrupan los 2300 resúmenes en un archivo. El tamaño del archivo obtenido es de 2.677 MB.

3.1.2. Visual Heuristic Cluster Analysis for Texts (VHCA).

Para este análisis se utilizó la herramienta «**Visual Heuristic Cluster Analysis for Texts**» (VHCA para textos) [17] [apéndice C], la cual proporciona un Diccionario Orientado a Dominio (DOD) [17]. Un DOD representa un tema específico, en este caso representa a computación PCDS; y se construye en base al análisis de la colección de textos. La herramienta VHCA facilita el desarrollo de un sistema jerárquico de DOD's; utilizando un conjunto de criterios heurísticos para la selección de las palabras clave. Esta herramienta nos proporciona una imagen de cada texto como un vector de cantidades de las palabras de DOD. Estas Imágenes se utilizan para formar tres matrices:

- **Matriz «texto / palabra»** contiene la cantidad de palabras de DOD (columna) en cada texto (fila);
- **Matriz «palabra / palabra»** contiene la cantidad de textos que contienen los pares de palabras de columna y fila;
- **Matriz «texto / texto»** contiene las mismas palabras de DOD en cada par de textos.

Para la construcción de DOD con VHCA se toma la frecuencia de las palabras del cuerpo de textos (resúmenes), y se forma un diccionario. La frecuencia de cada palabra del diccionario se compara con la frecuencia que tiene cada palabra del léxico común, si es tres veces mayor se agrega a DOD. Estas palabras deben satisfacer los siguientes criterios:

- La frecuencia relativa de una palabra en el cuerpo de textos, debe ser K_0 veces mayor que la frecuencia relativa de la misma palabra, en la lista de frecuencia del léxico común usado (para este caso $K_0 = 400\%$).
- El Criterio para K_1 define el límite superior para la mínima cantidad de textos que contienen una palabra dada al menos una vez.
- El Criterio para K_2 define el límite inferior para la máxima cantidad de textos que contienen una palabra dada (el objetivo es excluir palabras científicas comunes para todos los subtemas).

De acuerdo al principio de simetría se tiene que

$$K_2 = 100\% - K_1 \quad (3.1)$$

3.1.3. Método para análisis de datos con conocimiento pobre.

El resumen (abstract) es una de las partes más importantes de un artículo científico, ya que sintetiza el propósito del trabajo (introducción), los métodos principales (metodología), los resultados más importantes (resultados) y las conclusiones principales (discusión). Es la representación abreviada y precisa del contenido de un artículo científico. El número de palabras que contiene varía de 150 a 200, y se estima que no debe exceder 250, independientemente de la extensión del artículo. Las opciones reales de K_0 , K_1 y K_2 reducen estos números de 2 a 5 veces. En condiciones muy pobres cuando $K_0 = 100\%$, solamente un promedio de 5 palabras son candidatos a DOD. Esto se debe a que los resúmenes son considerados «objetos de conocimiento pobre» [4]. Para obtener «objetos de conocimiento enriquecido» se unen los resúmenes publicados durante un año en un texto anual (segundo criterio de la sección 3.1.1), y de esta forma se obtiene un archivo que contiene aproximadamente 100 palabras como nuevos candidatos de DOD. La figura 3.5 muestra la forma idealizada para convertir documentos con conocimiento pobre a documentos con conocimiento enriquecido.

| Palabras de DOD | | | | | | | |
|-------------------------|-------|-------|-------|-----|-----------|-----------|-------|
| Resúmenes | w_1 | w_2 | w_3 | ... | w_{n-2} | w_{n-1} | w_n |
| r_1 1990 | * | * | | | | | |
| r_2 1990 | | * | * | | | | |
| r_3 1990 | | | * | | * | | |
| $\sum r_1 r_2 r_3$ 1990 | * | * | * | | * | | |
| r_1 1991 | | * | | | * | | |
| r_2 1991 | | | * | | | * | |
| r_3 1991 | | | * | | * | | |
| $\sum r_1 r_2 r_3$ 1991 | | * | * | | * | * | |
| r_1 1992 | | | * | | | * | |
| r_2 1992 | | | * | | * | | |
| r_3 1992 | | | | | | * | * |
| $\sum r_1 r_2 r_3$ 1992 | | | * | | * | * | * |

Figura 3.5. Técnicas para convertir un documento de conocimiento pobre a un documento de conocimiento enriquecido.

La columna de la izquierda contiene los resúmenes relacionados con el tema de computación PCDS para los años 1990, 1991 y 1992; la fila superior muestra las palabras que contiene DOD, mientras que las filas en color amarillo muestran la suma anual de estos resúmenes. Por ejemplo, para 1990 se tienen tres resúmenes: r_1 contiene las palabras w_1 y w_2 ; r_2 las palabras w_2 y w_3 ; r_3 las palabras w_3 y w_{n-2} , estas seis palabras pertenecen a DOD. La primera fila amarilla se compone de cuatro palabras diferentes (w_1, w_2, w_3 y w_{n-2}) que pertenecen a la suma anual de resúmenes para 1990. En la suma anual para 1991, la palabra w_1 desaparece y es sustituida por w_{n-1} ; en 1992 desaparece w_2 y es sustituida por w_n . Siguiendo esta tendencia se observa que las palabras más viejas desaparecen (dando lugar a palabras nuevas), mientras que las palabras más recientes permanecen.

En la figura 3.6 se observa que las palabras w_3 y w_{n-2} son comunes a los resúmenes r_3 , r_3 , r_2 para los años 1990, 1991 y 1992 respectivamente; pero debido a la disposición individual de estos resúmenes (primer criterio de la sección 3.1.1) no es posible descubrir la tendencia temporal, ya que los resúmenes se encuentran dispersos, y la cantidad de palabras es muy pequeña (como se mencionó anteriormente) para hallar alguna relación entre ellas y DOD.

| Palabras de DOD | | | | | | | |
|-----------------|-------|-------|-------|-----|-----------|-----------|-------|
| Resúmenes | w_1 | w_2 | w_3 | ... | w_{n-2} | w_{n-1} | w_n |
| r_1 1990 | * | * | | | | | |
| r_2 1990 | | * | * | | | | |
| r_3 1990 | | | * | | * | | |
| r_1 1991 | | * | | | * | | |
| r_2 1991 | | | * | | | * | |
| r_3 1991 | | | * | | * | | |
| r_1 1992 | | | * | | | * | |
| r_2 1992 | | | * | | * | | |
| r_3 1992 | | | | | | * | * |

Figura 3.6. Palabras comunes para los años 1990, 1991 y 1992 de la distribución de resúmenes sin agrupar.

Agrupando los resúmenes en un texto anual, es posible observar la tendencia temporal del desarrollo de palabras sobre computación PCDS. Esta tendencia se observa claramente en la figura 3.7, donde las palabras establecidas en los resúmenes de los primeros años, van cediendo lugar a palabras que surgen en los últimos años de publicación. Sin esta acción, los resúmenes son autoorganizados en clústeres alrededor de una cantidad pobre de palabras para la cual es difícil revelar la tendencia temporal.

| Palabras de DOD | | | | | | | |
|-------------------------|-------|-------|-------|-----|-----------|-----------|-------|
| Resúmenes | w_1 | w_2 | w_3 | ... | w_{n-2} | w_{n-1} | w_n |
| $\sum r_1 r_2 r_3$ 1990 | * | * | * | | * | | |
| $\sum r_1 r_2 r_3$ 1991 | | * | * | | * | * | |
| $\sum r_1 r_2 r_3$ 1992 | | | * | | * | * | * |

Figura 3.7. Tendencia de los resúmenes agrupados por año de publicación.

En la figura 3.8 se muestra la matriz «texto / palabra» construida con la herramienta VHCA, donde cada texto es la suma anual de resúmenes durante un año. Cada elemento distinto de cero en la matriz, es substituido por un rectángulo que contiene el nivel de gris correspondiente, según la escala localizada en la parte izquierda de la figura.

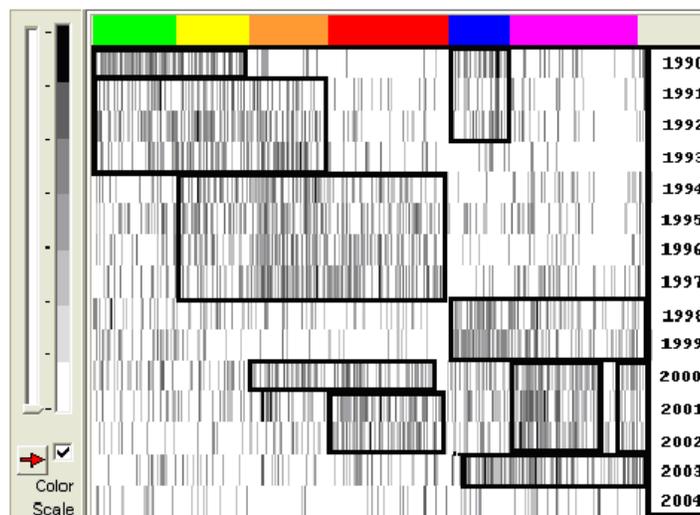


Figura 3.8. Matriz texto / palabra para la suma anual de resúmenes.

Esta matriz se obtiene del DOD en base a las siguientes condiciones: $K1 = 14\%$ y $K2 = 86\%$ de resúmenes. Es posible ver los grupos firmes (clústeres) de palabras para algunos grupos de textos anuales. Construyendo la matriz del tipo «texto / palabra» directamente de los resúmenes, se obtiene una matriz muy rara o un objeto de conocimiento pobre. La matriz de la figura 3.8 es usada para calcular la matriz «texto / texto» presentado en la figura 3.9.

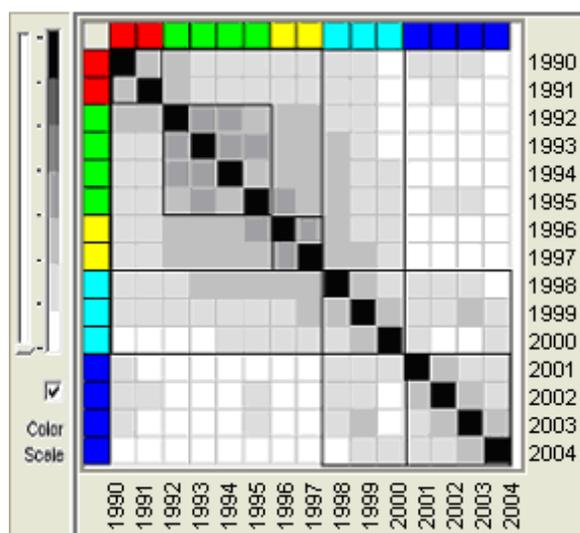


Figura 3.9. Matriz «texto / texto» para la suma anual de resúmenes.

Para un conjunto uniforme de textos la investigación es completa, pero los textos anuales son una mezcla de distintos temas (según el método de preparación) y se puede asumir una mayor diversidad de temas para los mismos años.

Para una investigación más detallada de un conjunto no homogéneo de textos es posible obtener mayor detalle. Para este objetivo se preparó la matriz «texto / texto» para los 2300 resúmenes. En la figura 3.10a y 3.10b se muestra la matriz antes y después del clustering. En la figura 3.10b se observa claramente cuatro clústeres grandes de resúmenes que están relacionados con más temas parciales.

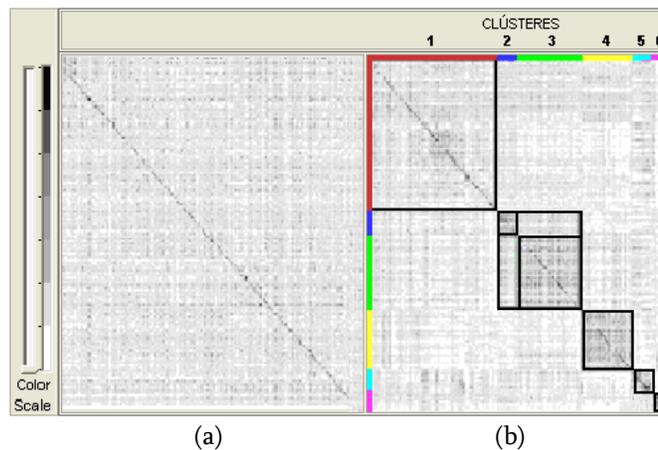


Figura 3.10. Matriz texto / texto para los 2300 resúmenes de diferentes años a) antes del clustering y b) después del clustering.

La cantidad de resúmenes se representa en tres clústeres grandes como se muestra en la figura 3.11. Se observa que el cluster 3 tiene un pequeño comportamiento oscilante hasta 1997, en 1998 decrece y nuevamente reanuda a su estado anterior. El cluster 1 sigue un comportamiento similar al cluster 3 hasta 1997, y a partir de 1998 presenta una tendencia creciente hasta el 2004.

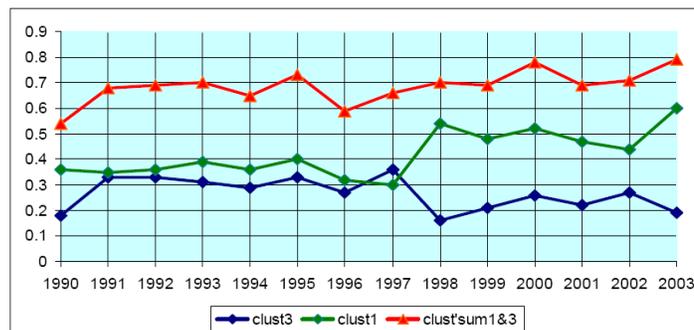


Figura 3.11. Evolución temporal de los clústeres 1 y 3.

La desviación del número de resúmenes en los clústeres, proporciona información sobre el cambio de interés para los temas que están relacionados con cada cluster. El problema es llegar a conocer el contenido de estos temas.

3.1.4. Análisis de resultados.

Para esta finalidad se utilizó el conjunto de textos anuales para los clústeres numerosos. Cada cluster contiene una suma de resúmenes del mismo año de publicación. De lo contrario se obtiene textos con conocimiento pobre. Este nuevo conjunto de textos anuales pueden ser analizados por el mismo método que se utilizó para el cuerpo de textos inicial. La diferencia es que los nuevos textos anuales son más uniformes (homogéneos). Por esta razón es posible revelar cada cluster nuevo para la investigación anual. De hecho, con la ayuda de VHCA se obtiene DOD's parciales de clústeres nuevos. Y estos DOD's son diccionarios de subtemas que un experto puede revelar (dar a conocer) analizando su contenido.

Ahora la pregunta es: ¿Qué hacer si no se cuenta con un investigador experto? En caso de no contar con un experto en el área, es posible usar la misma herramienta (u otra) para el reconocimiento de patrones. Para este objetivo se combinó un nuevo cuerpo de textos que contienen:

- Un conjunto de capítulos de manuales, artículos con diferentes temas parciales (y diferentes conjuntos de palabras especiales);
- Un conjunto de resúmenes agrupados por año de publicación (como se obtuvo anteriormente).

De la figura 3.12 se observa que las palabras de los textos anuales tienen correspondencia con los libros, y un título donde se pueden encontrar estas palabras. Los clústeres de los textos anuales, se enlazan con el clúster de palabras que contiene los títulos de libros. Por eso es productivo usar libros como «expertos», mediante los cuales se reconocen a los clústeres de palabras como parte de los temas más abstractos (nombre de libros o títulos de capítulos de libros y manuales). Los vínculos visibles (a) entre los textos anuales de clústeres diferentes, pueden ser explicados usando palabras científicas que son más comunes en los artículos que en los manuales. Los textos del **cluster 3** tienen menos vínculos con los manuales que el **cluster 1**. Posiblemente la razón es que no se tiene patrones correspondientes en el conjunto o bien, el tema correspondiente a este cluster es nuevo o, es una discusión de problemas muy generales, sin palabras especiales de temas parciales. El cluster 1 por el contrario tiene vínculos bastante fuertes (b) con el tema de Threads (hilos); en años posteriores con el tema de MPI (Message Passing Interface) y PVM (Parallel Virtual Machine) en casi todos los años (c) a excepción de 1998 y 1999, y con los temas de Monitor y Monitoring (g) y (h). El **cluster 3** tiene solamente vínculos pobres (e); con Java Threads, Sockets en 1999 (d), y con PVM (f). Parte de las palabras de los manuales, son

distribuidas entre los clústeres de los textos anuales. En este caso, es posible contar el porcentaje de la representación de cada tema en los manuales, para cada clúster en los textos anuales.

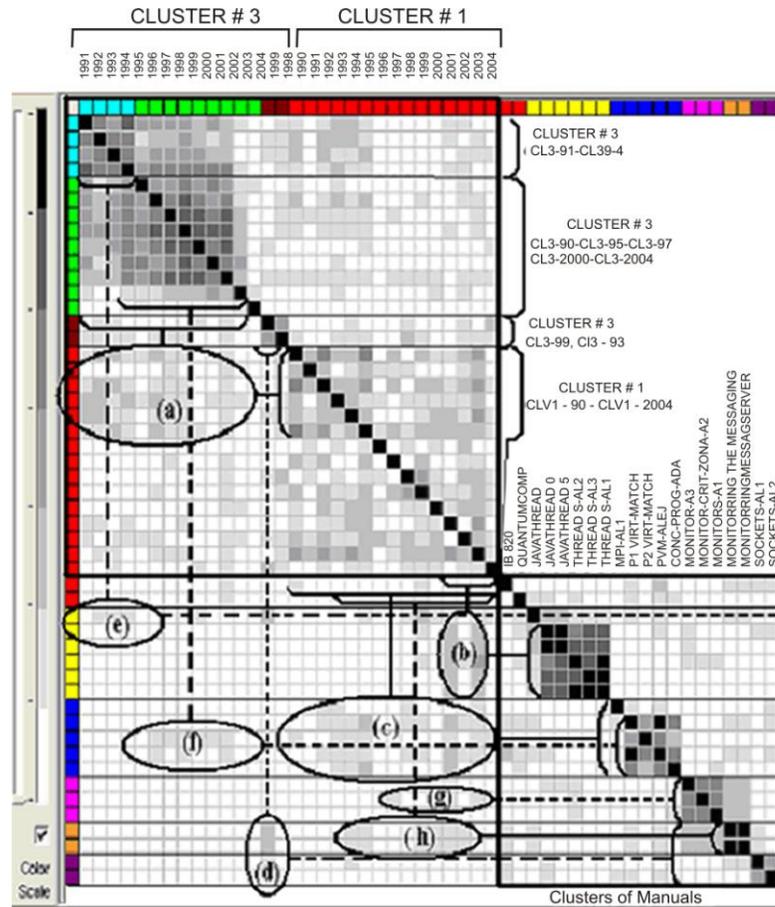


Figura 3.12. Matriz texto / texto para patrones y textos anuales del cluster 1 y el cluster 3.

3.2. Método de construcción y algoritmo del Prototipo Rápido de Ontología (PRO).

Cualquier documento científico o técnico es organizado jerárquicamente: algunas secciones del texto (tales como el resumen o las conclusiones) simplifican el contenido del texto principal; los títulos de las secciones describen el contenido en palabras generales; el título de cada capítulo describe el contenido de un conjunto de secciones; el título del libro describe el contextos de todos los capítulos, etc. Por otra parte, los documentos científicos por lo general se organizan jerárquicamente: por ejemplo, los artículos son organizados en Journals, Conferencias, etc., los

cuales a su vez tienen sus propios títulos. Aprovechando esta estructura jerárquica se construye un PRO [16] de este dominio, basado en una colección suficientemente grande de distintos documentos relacionados con computación PCDS. Para ello se admite que ontología es el esquema jerárquico de conceptos y palabras, donde los niveles más bajos de la «pirámide» son conjuntos de conceptos más detallados de un área o rama del conocimiento científico. Cada nivel superior contiene conceptos de nivel más abstracto hasta el nivel superior donde se tienen los cuatro conceptos: **paralelo**, **concurrente**, **distribuido** y **simultáneo**. La creación de una ontología usualmente implica trabajo manual que involucra a expertos de una rama específica, y a ingenieros del conocimiento. Existen una variedad de métodos para extraer automáticamente ontologías, de textos. Por ejemplo [18] busca en textos, patrones tales como **Francia y otros países Europeos** para extraer el hecho de que **Francia es un país europeo**. [19] publica un experimento sobre la aplicación de un método similar, para una recopilación de 500 millones de páginas Web. Sin embargo, la cantidad de textos disponibles en un dominio técnico limitado no es suficiente para estos métodos. Este método por el contrario, puede hacer frente a una cantidad relativamente pequeña de datos. Por otra parte, los métodos existentes para extraer **ontologías de dominio limitado**, requieren de esfuerzo manual o solo pueden extraer un número muy pequeño de palabras. La observación respecto a la utilidad de la estructura jerárquica de textos, proporciona una fuente alternativa de información sobre relaciones entre palabras.

3.2.1. Estructura del muestreo para creación y análisis de PRO.

Para conocer el estado de un área científica durante cierto período, es posible usar su ontología: un informe de las palabras que describen sus ramas de modo general, y las palabras que describen los conceptos o sub-ramas dentro de cada rama. Este trabajo se basa en que las ramas más grandes de un campo corresponden a los títulos de revistas (Journals), libros, o conferencias; las sub-ramas corresponden a títulos de artículos publicados en estos Journals o conferencias etc.

Para la creación del PRO se emplea un cuerpo de textos que contiene conceptos sobre computación PCDS en cuatro niveles diferentes de abstracción:

- **Fuente del primer nivel de abstracción:** títulos de congresos, de libros, de monografías, de manuales y de revistas;
- **Fuente de segundo nivel de abstracción:** capítulos de manuales y / o monografías, títulos de las secciones de congresos, nombre de secciones de congresos, áreas de revistas;
- **Fuente de tercer nivel de abstracción:** títulos de artículos (pero no sus partes), de congresos y nombre de párrafos de manuales;
- **Fuente de cuarto nivel de abstracción:** artículo, resúmenes y palabras clave.

El cuerpo de textos contiene 16 libros sobre Computación PCDS y 2300 resúmenes. Los resúmenes pertenecen a 834 congresos sobre computación PCDS. Estas cuatro palabras son los

conceptos de más alto nivel de abstracción de la ontología. De esta colección se extrajo el material textual de los niveles de abstracción. El material textual se agrupó en archivos T_i . Cada archivo contiene material textual del mismo tipo (títulos de congresos, de artículos sin datos del autor, contenido de libros, etc.) y el mismo período de tiempo, es decir unido por año o años vecinos.

3.2.2. Método de construcción de PRO.

Cada parte del material textual (a excepción del cuerpo de artículos) es muy pequeño y no contiene información suficiente para efectuar el clustering. Por lo que fue necesario unirlos para enriquecer los vínculos entre las unidades textuales. En las secciones 3.1.1 y 3.1.3 se expuso una metodología para enriquecer textos con conocimiento pobre. Para unir textos cortos se usan los atributos externos de cada texto como es la fecha de publicación. El enlace del contenido de libros es posible por la coincidencia de tema (lo que realmente se ha estado haciendo), también por años de publicación. Sin esta acción, los textos cortos son autoorganizados en clústeres alrededor de una cantidad pobre de palabras, para las cuales es difícil revelar tendencias temporales. Los archivos T_i formados de esta manera son usados para las siguientes etapas de nuestro estudio:

Primera etapa. Se usa la herramienta VHCA como paso elemental del método propuesto. Con VHCA se obtiene un DOD para el cuerpo de textos, y una imagen de cada texto como vector de la cantidad de palabras de DOD. Estas imágenes textuales se usan para formar tres matrices: matriz «texto / palabra», matriz «palabra / palabra», matriz «texto / texto» (ver sección 3.1.2).

Segunda etapa. Se produce el clustering de las palabras en la matriz «texto / palabra» para el conjuntos de textos con la herramienta VHCA. Cada conjunto contiene textos de distinto nivel de abstracción. La primera matriz contiene “títulos de congresos, títulos de artículos, y los índices de los libros”. La segunda matriz contiene “cuerpos de resúmenes y títulos de textos”. Y por último la tercera matriz contiene los “resúmenes y los artículos”.

En la primera matriz los clústeres de palabras que contienen títulos de congresos se separan de los clústeres de palabras que no representan títulos de congresos. En la parte izquierda de la matriz de la figura 3.13 se muestra el **cluster A**. Cada elemento distinto de cero en la matriz es substituido por un rectángulo que contiene el nivel de gris correspondiente según la escala localizada en la parte izquierda de la figura.

Los clústeres de palabras que no están vinculados con los títulos de congresos, se pueden usar para los próximos niveles (bajos) de abstracción. Se excluyen los títulos de congresos de los procesos de construcción del próximo nivel de PRO.

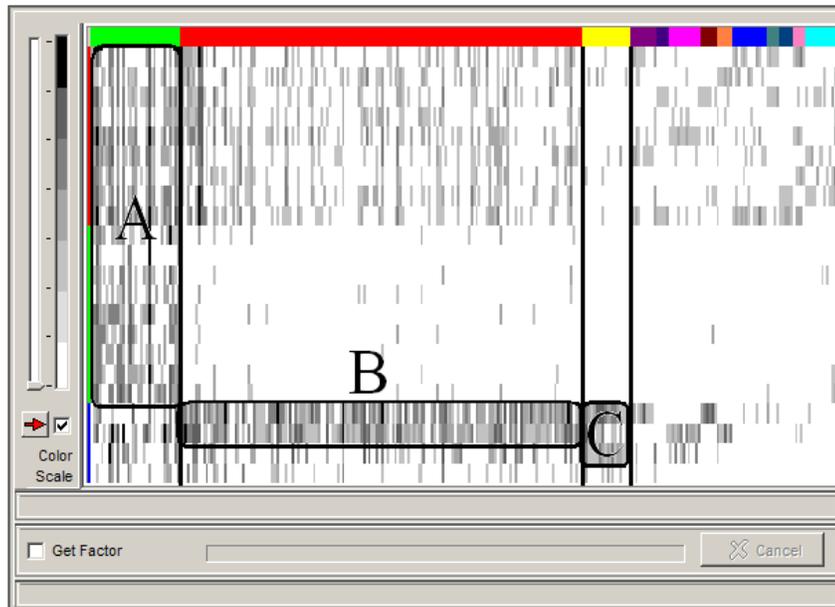


Figura 3.13. Distribución de clústeres de palabras por tipo de documentos (títulos de resúmenes, títulos de congresos e índices de libros).

Para el último paso de esta etapa se utiliza la matriz de textos de resúmenes.

El segundo nivel de la jerarquía de PRO representada por los **clústeres B y C** de la figura 3.13, tiene una representación rica de clústeres de **índices de libros** y pobre en **títulos de resúmenes**. Es necesario observar los otros clústeres en el nivel más bajo de la jerarquía. Ya que existen palabras del **cluster A** en los **títulos de resúmenes** y en los **títulos de congresos**. La ontología más detallada se prepara con un muestreo mayor. Su análisis detallado se muestra en la figura 3.22.

Tercera etapa. Usando la matriz «texto / palabra» se analiza la distribución de palabras y la evolución temporal de cada cluster. Para el ejemplo de la figura 3.14, los temas relacionados con el **cluster A** decaen temporalmente, los temas del **cluster B** son más numerosos en 1997 con un levantamiento gradual y luego una disminución. Los temas, relacionados con el cluster C, crecen en 1997 y se desarrollan enormemente en el año 2003, mientras que los temas relacionados con el clúster D son locales.

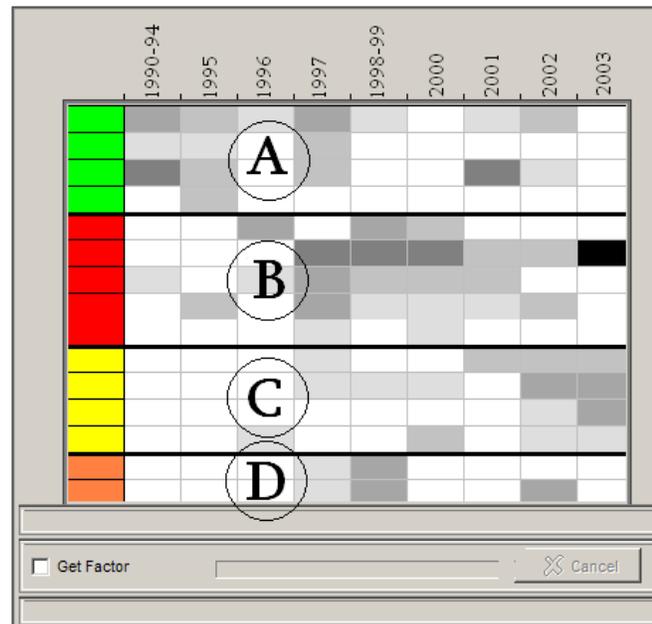


Figura 3.14. Distribución anual de clústeres de palabras.

3.2.3. Base de elementos de la Ontología.

El objetivo de esta sección es construir una jerarquía de temas y sub-temas basado en el dominio de computación PCDS. En particular, los niveles de la ontología están en una correspondencia uno a uno con los niveles de la jerarquía de textos. Aunque la idea parezca bastante simple, hay varios detalles que deben considerarse para que esta idea trabaje en la práctica. A continuación se definen las nociones principales usadas en el algoritmo.

3.2.3.1. Estructura de entrada de datos: cuerpo de textos.

La mayoría de los textos existentes son estructurados jerárquicamente de tal modo que algunos de sus segmentos (que pueden ser muy cortos, por ejemplo, los títulos) son marcados como «más generales» que otros. Usando tales textos y algunas fuentes de información estadística se construye la ontología jerárquica.

3.2.3.1.1. Textos estructurados jerárquicamente.

Es cualquier tipo de texto dividido en segmentos y ordenados en un árbol (o más generalmente, en un grafo dirigido acíclico) de modo que los segmentos de nivel superior contienen la metainformación de los texto subordinados. Por simplicidad nos referimos a tales segmentos de texto, como textos. Por ejemplo, en un libro, el nodo raíz es el **título de libro** (que es un texto corto); los niveles intermedios son formados por los **títulos de las partes**, **títulos de capítulos**, **títulos de secciones**, etc.; finalmente, el texto completo de cada sección es una hoja inferior del título de la sección correspondiente.

En los experimentos obtenidos se utilizó la colección de resúmenes de artículos presentados en conferencias sobre computación PCDS. Así la colección tiene una estructura tipo árbol de tres niveles, el nivel 0 (raíz) es agregado por conveniencia. En la figura 3.15, la descripción del dominio para este caso es: **paralela, concurrente, distribuida y simultánea**.

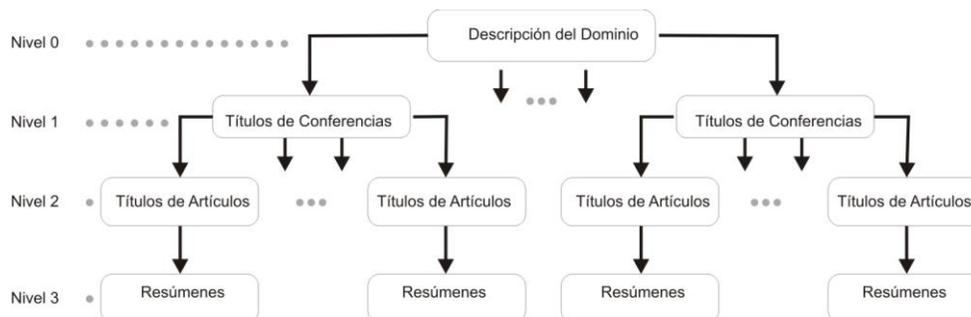


Figura 3.15. Jerarquía de los segmentos de texto.

3.2.3.1.2 Dominio.

Es un tema técnico limitado (computación PCDS). Los textos de dominio limitado incluyen tres tipos de palabras:

- Palabras de léxico general que aparecen en cualquier texto. Tales palabras incluyen artículos, preposiciones, o palabras como **have**, **give**, **see**, **number**, etc. Cualquier información sobre tales palabras puede ser fácilmente encontradas en los diccionarios existentes.
- Términos técnicos usados en un área más amplia del conocimiento, y no solamente dados en el dominio limitado. Para el dominio limitado de computación PCDS, son palabras usadas en textos en otras áreas de computación, tales como **program** o **execution**.

- Términos específicos para el dominio seleccionado, tales como **clock**, **monitor**, o **semaphore** para el dominio de computación PCDS.

En el dominio de la ontología solamente interesa el último tipo de palabras: términos de dominio específico. Estos términos tienen características especiales, comparados con las palabras o términos generales del dominio más general:

- No son homónimos (polisemous) es decir, son inequívocos. Mientras en el texto general, la palabra **paralelo** puede referirse a dos líneas rectas o dos ideas similares, en el dominio limitado, esto solamente puede significar **simultáneo**.
- Por lo general no son sinónimos. Mientras que en el texto general, **clock** puede ser substituido con **watch**, **chronometer**, o **timer**, en el dominio elegido ninguna palabra con excepción de **clock** puede ser usada para referirse a este dispositivo. Sin embargo, algunos sinónimos todavía existen en los textos de dominio limitado, por ejemplo computación **paralela**, **concurrente**, **distribuida** y **simultánea**.

Este método puede usarse fácilmente para construir una ontología de temas generales. Las propiedades del texto de dominio específico, solucionan enormemente problemas lingüísticos en su análisis estadístico automático, y nos permite omitir muchas complicaciones en la implementación de la discusión presentada aquí. En particular, evitan detectar y manejar sinónimos (los dos principales problemas impiden obtener la estadística exacta del uso de palabra en textos generales). Sin embargo, hay que ocuparse de otros fenómenos lingüísticos tales como morfología y expresiones multi-palabras, según se describe en la sección 3.2.3.2.

3.2.3.1.3. Datos.

Las fuentes usadas en el algoritmo son: una recopilación del dominio estructurado jerárquicamente, un cuerpo de un área muy amplia, y un diccionario de frecuencias para el lenguaje dado. Las últimas dos fuentes son usadas solamente para proporcionar ejemplos negativos para la clasificación de palabras en el dominio específico y no específico de términos.

Como cuerpo del dominio, se tiene un **cuerpo de artículos de conferencias** que pertenecen a un dominio limitado a computación PCDS, ver la figura 3.15. Como una fuente de ejemplos negativos para la selección de palabras del dominio, se usa un cuerpo sobre **programación suave**. El diccionario de frecuencias es necesario porque el cuerpo del dominio puede no contener la estadística confiable para algunas palabras.

3.2.3.2. Estructura de salida: Ontología.

La ontología consiste de conceptos (aproximadamente, palabras) de un dominio elegido, interrelacionado para formar una jerarquía (aproximadamente, un árbol). Más adelante se detallan estas nociones.

3.2.3.2.1. Relación.

Informalmente es una subordinación entre temas y sub-temas en un dominio dado. Ya que tales temas y sub-temas son descritos por palabras. Finalmente lo que se arregla en una jerarquía son palabras. La relación que resulta entre palabras difiere de la relación «**is-a**» (Sócrates is a man) a menudo considerado en ontologías donde las palabras no corresponden a ningún subtítulo en un contexto (**Sócrates died** \Rightarrow **a man died**) ni hereda propiedades de otros (**men are mortal** \Rightarrow Sócrates is mortal). En cambio, la relación en la que estamos interesados aquí puede ser llamada **implicación**. Esto se parece a la relación **holonym / meronym**: las palabras de más alto nivel describen situaciones que implican las palabras de más bajo nivel. Por ejemplo, **evaluation** implica **measurement** (que alternadamente implica **value** y **calculation**) como una parte del proceso; **processor** implica **memoria** como parte de su funcionamiento, e implica **register** como parte física.

3.2.3.2.2. Palabra.

Es un símbolo semánticamente significativo en el texto. Los fenómenos lingüísticos más relevantes relacionados con palabras, son expresiones morfológicas y expresiones de multi-palabra. En cuanto a la variación morfológica (por ejemplo, **do**, **did**, **does**, **done**, **doing**), se consideran todas las variantes como representativas de uno y de la misma unidad (lexema); de aquí en adelante se refiere simplemente como palabra. Así, la lista anterior menciona simplemente la palabra “**do**” cinco veces, incluso en diversas variantes morfológicas.

Las expresiones multi-palabra son definidas como combinaciones de palabras representando a una sola entidad que no puede ser representada por una sola palabra, por ejemplo: **mutual exclusion**, **hot dog**, **Nueva York**. A diferencia de expresiones multi-palabra, otro tipo de combinaciones de palabra se refieren a **hyponyms** de una de las palabras: **mutual love is love** que es **mutual**; **litttle dog** es un **dog** que es **litttle**. Los términos multi-palabra son mas frecuentes en textos de dominio limitado que en temas generales. En consecuencia, en el algoritmo tales expresiones son identificadas y tratadas del mismo modo como palabras solas. Por simplicidad, nos referiremos a tales expresiones como **palabras** (compuestas).

3.2.3.2.3. Concepto.

Es una entrada elemental en la ontología. Es una palabra individual o un grupo de palabras (cluster) que se utilizan con mayor frecuencia en el mismo texto, por ejemplo, {**clock**, **mutual exclusion**, **monitor**} se recuerda que estas palabras son inequívocas dentro del dominio limitado y las expresiones multi-palabra tales como **mutual exclusion** son consideradas como una sola palabras. Por definición, cada palabra pertenece solamente a un concepto (grupo); esto es posible porque se consideran todas las palabras inequívocas. De hecho, las propiedades estadísticas de tales palabras que frecuentemente **coocurren** son prácticamente idénticas, así no existe razón para distinguir las en una ontología construida automáticamente. Por otra parte, agrupar palabras que coocurren reduce la dimensionalidad de la tarea y mejora el significado estadístico de los resultados, aunque a costa de su **granularidad**.

Existen otros métodos de reducción de dimensionalidad, tales como la Indexación Semántica Latente (ISL). Sin embargo, en la ISL las unidades resultantes son combinaciones lineales difíciles de entender por los usuarios de la ontología, y con un sentido lingüístico confuso. Esta es la razón por la cual optamos por simples grupos de palabras.

La representación de una entrada elemental en una ontología (como un conjunto de palabras) se parece a los **synsets** familiares de WordNet. Sin embargo, hay una diferencia crucial entre los **synsets** y los clústeres de coocurrencia. Las palabras combinadas en un **synset** son distintas variantes para referirse a la misma entidad o idea (intercambiable en cada contexto). Así, un texto que menciona **n₁** veces la primera palabra en el **synset**, ..., **n_k** veces la última palabra en el **synset**, se considera señalando la entidad correspondiente **n = n₁ + ... + n_k veces**, es decir, la frecuencia de entrada (**synset**) en un texto es la suma de las frecuencias de sus palabras .

Un texto mencionando a uno de ellos implica todos los otros: por ejemplo, un texto sobre computación paralela hablando sobre **mutual exclusion** implica necesariamente la presencia de **monitor** para su dirección y de **clock** usado por este último. Así, todas las **k**-palabras en el cluster se refieren a la situación correspondiente solamente una vez, cada uno prediciendo la presencia del resto. En el caso ideal, la frecuencia **n** del concepto (cluster), es la frecuencia **n₁** de uno de los representados, ignorando las demás palabras. Dado que en la práctica algunas palabras pueden estar implícitas, se define la frecuencia del cluster como la frecuencia promedio de sus palabras miembros:

$$\frac{n_1 + \dots + n_k}{k} \quad (3.2)$$

Se observa que la demanda de palabras en el cluster es válida solamente para textos con dominio limitado, donde las palabras se refieren a una situación estándar inequívocamente, y así claramente predicen la presencia de las otras. Es más, puesto que la igualdad no existe en los

textos de dominio limitado, se ignora su presencia. Para manejarlas es necesario utilizar **synsets** en vez de palabras:

$$\{(\text{clock, watch}), (\text{mutual exclusion, avoidance}), (\text{monitor, supervisor})\} \quad (3.3)$$

Resumiendo las frecuencias de las palabras dentro del **synsets** y promediando los resultados sobre todos los **synsets**.

Aquellos sinónimos raros que existen en el texto de dominio limitado, resultan ser mapeados al mismo concepto. La razón de esto es que coocurren con las mismas palabras en los textos. Ya que los conceptos son clústeres de palabras que coocurren frecuentemente, los sinónimos son mapeados al mismo cluster, porque todos tienen las mismas palabras con las cuales coocurren; incluso si no coocurren con cada uno, ver la figura 3.16. Se observa que a diferencia de la ecuación 3.3, la estructura del cluster es plana: **{clock, watch, mutual exclusion, avoidance, monitor, supervisor}**, de modo que no se sepa que palabra es sinónimo de cual. Esto conduce al cálculo incorrecto de frecuencias: las frecuencias de todas las palabras en el cluster son promediadas, en vez de resumir las frecuencias de los sinónimos. Sin embargo, esto no se considera un problema serio, desde (a) existen pocos sinónimos en los textos de dominio limitado y (b) casi no aparecen juntos en el mismo texto ya que se usa solamente una variante del término.

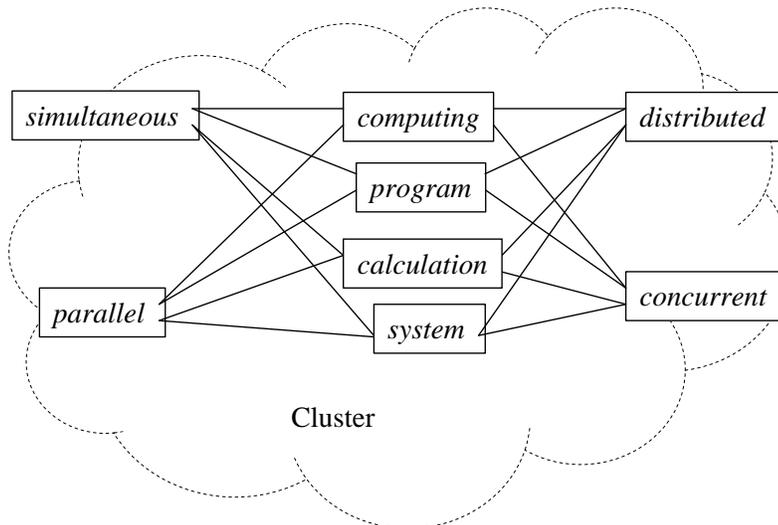


Figura 3.16. Estructura plana de clústeres.

3.2.3.2.4. Tema.

Es un cluster de conceptos, agrupado según su coocurrencia: por ejemplo para **{{clock, mutual exclusion, monitor}}**, **{thread, semaphore}}**, el umbral de la frecuencia de coocurrencia para el clustering, es inferior que para el caso de clustering de las palabras en conceptos. Tales clústeres son más difusos que los conceptos, es decir corresponden a ideas más generales que se refieren a una situación específica. Esto permite una estadística confiable incluso sobre textos muy cortos.

Tales temas no son variantes de conceptos «difusos». En efecto, el clustering de conceptos (cluster de palabras) no es lo mismo que clustering de palabras individuales. Se debe recordar que en un texto, la frecuencia de ocurrencia de un concepto es el promedio de las frecuencias de sus palabras miembros: un texto mencionando **clock** y **monitor** contiene solamente una ocurrencia del concepto **{clock, mutual exclusion, monitor}**.

Así, para encontrar tales temas se tiene un proceso de dos pasos: primero, las palabras individuales son agrupadas en conceptos (con un umbral de concurrencia alto), y luego las unidades resultantes son agrupadas otra vez, pero con un umbral bajo. La práctica muestra que el grupo resultante difiere. Los clústeres del segundo paso son identificados mas confiablemente, debido a la **reducción de la dimensionalidad** que resulta del primer paso. Por las razones discutidas arriba, el resultado de un procedimiento de clustering de dos pasos es lingüísticamente más plausible. Una vez más, por definición cada concepto (y así cada palabra) pertenece solamente a un tema.

3.2.3.2.5. Ontología.

Es un grafo dirigido acíclico con una raíz, en la práctica «casi» un árbol enraizado, con nodos raros que tienen múltiples padres. Se mantiene una estructura de niveles en el grafo: si un nodo tiene múltiples padres, todos ellos pertenecen al mismo nivel (es decir, la misma distancia de la raíz). Los nodos de las hojas del grafo son conceptos descritos arriba. Sin embargo, los nodos no terminales, son temas. Esto se debe a dos razones. La primera, puesto que la ontología es «casi» un árbol, para su utilidad es necesario pocos nodos en niveles más altos del árbol. La segunda razón es puramente técnica: la confiabilidad estadística de las relaciones entre N nodos es:

$$O \frac{1}{N^2} \quad (3.4)$$

así, con muchos nodos en los niveles más bajos y en los más altos, es difícil agrupar la estadística confiable de sus relaciones. Considerando pocos clústeres (aunque difusos) en los niveles más altos, conduce a la reducción detallada en los niveles inferiores. La relación de subordinación entre nodos es interpretada como la relación de implicación (tema / sub-tema) descrito arriba;

mientras que las capas de los nodos, corresponden a niveles percibido intuitivamente de los temas abstractos. La última intuición, se expresa en la estructura de capas de la jerarquía, como se observa en la figura 3.15.

3.2.4. Algoritmo para construir la Ontología.

Para cada nodo de la jerarquía ontológica, el algoritmo construye recurrentemente los nodos subordinados de las palabras usadas en el siguiente nivel, que frecuentemente coocurren con las palabras del nodo dado. En la figura 3.17 se muestra el algoritmo para construcción de la ontología.

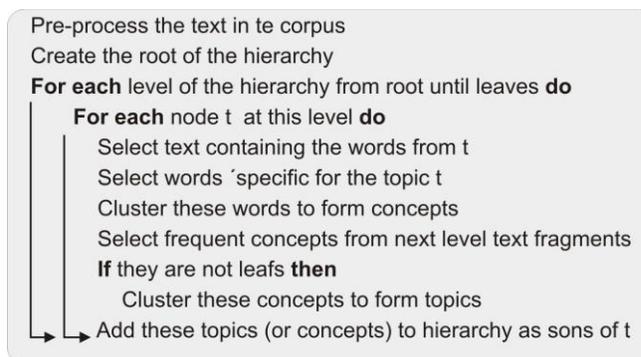


Figura 3.17. Algoritmo para construcción de la ontología.

3.2.4.1. Pre-procesamiento del cuerpo.

Usando las palabras clave asignadas a la raíz, se busca en Internet textos que contengan estas palabras o información relacionada con ellas; ya sea en su título o en el cuerpo del resumen. Posteriormente se descarga los resúmenes de los artículos presentes en cada conferencia encontrada. Los textos obtenidos es la raíz descrito por [21]. En lo que sigue el término **palabra** se refiere a palabras raíz o expresiones de multi-palabra.

3.2.4.2. Formación de la raíz.

Las palabras que identifican el dominio de interés del modo más abstracto (en este caso **paralela**, **concurrente**, **distribuida** y **simultánea**) son asignadas a la raíz de la ontología como el tema más general del dominio. La raíz está en el nivel $L = 0$ de la jerarquía.

3.2.4.3. Recursión por nodos.

Los siguientes pasos son realizados para cada nodo no terminal. El proceso comienza con el nodo raíz, y se realiza para cada nodo recién construido, excepto para los nodos del nivel de hoja. El objetivo de estos pasos es, dado un tema (nodo t) en un nivel $L - 1$ en la ontología; se construye sus nodos subordinados (de **nivel L**). La recursión debe ser organizada nivel por nivel (todos los nodos de nivel L se construyen antes de que se continúe con la construcción de los nodos de nivel $L + 1$), puesto que en cada paso necesitamos saber si un concepto pertenece a un tema de nivel superior.

3.2.4.4. Selección del sub-cuerpo.

Dado un tema t de nivel $L - 1$, todos los textos del siguiente nivel que contienen este tema son seleccionados. Un texto de nivel L , significa un cuadro en la jerarquía de la figura 3.15, enraizado por un fragmento de nivel L (un cuadro enraizado por un **nodo x** es el conjunto de nodos subordinados directamente o indirectamente a x , incluso a x). Por ejemplo, si t pertenece al nivel 1 (los títulos de conferencia en la figura 3.15), los textos se consideran artículos individuales, es decir, una concatenación de un resumen con su título. Decimos que un texto contiene un tema si contiene cualquier palabra que pertenece a este tema.

3.2.4.5. Selección de vocabulario.

Solamente son seleccionadas las palabras del sub-cuerpo, cuya frecuencia es $k \geq 3$ veces su frecuencia en el lenguaje común usado. Todas las demás palabras son ignoradas y no participan en ningunas otra operación para el **tema t** dado. También se excluyen las palabras ya asignadas a los nodos de niveles más altos (inferiores a L).

3.2.4.6. Agregado de documentos.

Para calcular la coocurrencia de palabras, se considera cada fragmento de texto en cada nivel de la jerarquía como un documento individual (figura 3.15). Sin embargo, en los niveles más altos de la jerarquía, tales "documentos" son muy cortos; es por ello que se concatenan todos los textos cortos hermanos. Por ejemplo, los títulos de conferencias fueron combinados en un solo "documento"; del mismo modo, los títulos de artículos dentro de cada conferencia fueron unidos. Cada resumen fue considerado como un documento individual, lo cual no es problema porque los resúmenes son más grandes que los títulos.

3.2.4.7. Formación de conceptos.

Dado un sub-cuerpo que contiene N documentos y W palabras distintas (tipos), se considera una matriz de frecuencia de palabras $N \times W$ con $F = |f_{dw}|$, donde f_{dw} es el número de ocurrencias de la palabra w en el documento d , y una matriz $W \times W$ de coocurrencia de palabras $C = |c_{ij}|$, donde c_{ij} es una medida del coseno entre las filas correspondientes de F (la información correspondiente también puede ser usada como una medida de coocurrencia). Para formar conceptos, las palabras se agrupan según la medida c_{ij} , utilizando un algoritmo que permita un umbral $\alpha = 0.9$, en el cluster interno relacionado; ver detalles en [22]. La mayoría de los clústeres tienen solamente una palabra, lo cual no es problema.

3.2.4.8. Selección de conceptos.

Se consideran los conceptos presentes en los fragmentos de más alto nivel en el sub-cuerpo (en los ejemplos anteriores estos son los títulos de los artículos, pero no sus resúmenes). Nuevamente, un concepto está presente en un texto, si al menos una de sus palabras está presente en este texto. Para reducir el ruido de estos conceptos se seleccionan los que ocurren más que

$$\lambda_t = \ln N_t \quad \text{veces} \quad (3.5)$$

3.2.4.9. Formación de temas.

A menos que L sea de nivel hoja, los conceptos seleccionados son clústeres de temas. Esto se hace de la misma forma como a las palabras que se agruparon en conceptos, pero con un umbral bajo, $\beta = 0.5$ (se puede ajustar manualmente). Los clústeres triviales (aquellos que tienen solamente un elemento) no están permitidos aquí. Para medir la coocurrencia entre conceptos, se utiliza el mismo cuerpo que se utilizó anteriormente para la creación de conceptos (y no el sub-cuerpo usado en el paso anterior). Cada tema recién construido (o concepto en el caso del nivel hoja) se convierte en un nuevo nodo subordinado al tema t .

3.3. Análisis dinámico de la Ontología.

La medida y el análisis de la evolución de una rama científica son importantes, ya que de esta forma se puede conocer su estado futuro y las tendencias en su desarrollo (al menos en los períodos de estabilidad). Los períodos de desarrollo de la ciencia normal, se basan en resultados

anteriores (aun cuando la dirección es imprevisible). Aumenta en los períodos de desarrollo activo de la infraestructura científica, y decrece algún tiempo antes de ser obsoleto y entrar en un período de estancamiento. Lo último es un indicio de que un cambio en el sistema de conocimiento está cerca (por ejemplo, un nuevo descubrimiento), y que además resuelve las dificultades acumuladas en el subsistema. En esta sección se propone un método **semi-automático** para construir una ontología que mida la evolución temporal de computación PCDS. El método se basa en la colección de documentos utilizado en la sección 3.2 para construir el PRO de PCDS. La estructura jerárquica de los documentos permite determinar la relación **padre-hijo** entre palabras: por ejemplo, una palabra que aparece en el título de un artículo, es candidato a ser padre de las palabras que aparecen en el cuerpo de éste; si tal relación se repite varias veces, se coloca el par **padre-hijo** en la ontología. Estas se construyen para cada año independientemente y se comparan usando el algoritmo **distancia de edición en árboles**.

La metodología se ilustra construyendo una ontología en un dominio temático específico: **paralela, concurrente, distribuida y simultánea**. En esta parte, se introduce un nuevo punto de vista sobre ontología: no como una colección estática de hechos acerca del lenguaje y el mundo; sino como una «**foto instantánea**» del estado del lenguaje y del mundo en un momento específico; la cual evoluciona mientras el mundo se desarrolla. Esto conduce a una nueva aplicación de la ontología como una herramienta para describir y estudiar la evolución del dominio: razonando a cerca de la pregunta: **¿Qué ha cambiado en el mundo (dominio)?** La podemos formular más exactamente como: **¿Que ha cambiado en su ontología?** Esto da una herramienta para estudiar cuantitativa y cualitativamente los cambios en un área específica [25].

Desde tal «**foto temporal**» es casi imposible construir manualmente una ontología, es por ello que se ha desarrollado un nuevo método para construir automáticamente una ontología, siempre y cuando deba construirse a partir de un dominio, en el cual la estructura jerárquica de documentos está disponible [24]. Este método tiene sus limitaciones cuando describe los cambios en el dominio, a través de los cambios en su ontología. Los cambios en el dominio se reflejan en tres procesos distintos:

- aparición o desaparición de términos;
- aparición o desaparición de relaciones entre términos;
- reconceptualización del significado de términos.

Se discute principalmente el primer efecto, como el más simple y más fácil de revelar con este método. El segundo efecto también puede ser estudiado, considerando cambios estructurales en la ontología (cualitativos en vez de cuantitativo). No se aborda directamente el tercer efecto. Sin embargo, en la sección sobre trabajos futuros se retoma.

3.3.1. Estructura de la Ontología de textos científicos.

Cuando se habla de una ontología se razona como una estructura jerárquica de conceptos y palabras, que pueden ser representadas por un grafo. Como se mencionó en la sección anterior, su nodo raíz (nivel cero de abstracción) identifica el dominio: **paralela, concurrente, distribuida** y **simultánea**. Cada siguiente nivel, contiene conceptos del nivel inferior. Las hojas de la jerarquía son términos más específicos. En todos los algoritmos se asume que los textos están enraizados [21]; en lo que sigue, por simplicidad el término palabras significa raíz.

Para detectar **conceptos** y **palabras** de un cierto dominio, se utiliza una lista de frecuencias de palabras comunes, y un cuerpo de textos con la misma lista de palabras de un nivel superior en la jerarquía: el cuerpo de textos relacionado con ciencias de la computación pero no computación paralela. Esto permite que se obtenga una ontología pequeña pero detallada.

Se experimentó con los 2300 resúmenes sobre computación PCDS. Los textos en esta colección se organizan jerárquicamente: cada artículo se encuentra por debajo del título de la conferencia o revista correspondiente, mientras que el cuerpo del artículo está por debajo de su título, etc.; específicamente, la jerarquía es como sigue:

- Descripción del dominio: **computación paralela, concurrente, distribuida y simultánea.**
- **Título de conferencia o journal.**
- **Título de artículos.**
- **Resúmenes, introducciones, conclusiones y referencias.**
- **Cuerpo del artículo.**

Los títulos de las conferencias y de los artículos son segmentos de texto muy cortos, así que es necesario enriquecer cada segmento, con palabras de los niveles inferiores de abstracción que ocurren debajo de él. Los segmentos de texto de cada nivel, proporcionan las palabras de los niveles correspondientes de abstracción. El cuerpo de artículos es la fuente de palabras del nivel más bajo. Se consideran tres niveles de abstracción:

- títulos de conferencias o journals;
- títulos de artículos;
- resúmenes (sin títulos).

Se unen los textos correspondientes a cada año de publicación (o a veces varios años, cuando se tiene pocos textos por año): por ejemplo, los títulos de los artículos para 1997, dio un grupo relativamente pequeño de archivos **Ti** (tres archivos por año). Para construir la ontología se utiliza la representación del espacio del vector de estos archivos.

Se comienza construyendo la lista de palabras de DOD; esta lista incluye los términos del dominio junto con sus **pesos** correspondientes para el dominio dado. El DOD incluye las

palabras cuya frecuencia es tres veces superior a las palabras de uso común. Usando DOD, se obtiene una representación del espacio del vector (imagen) de cada texto T_i , como un vector de frecuencias de las palabras de DOD en el texto. Estas imágenes son utilizadas para formar tres matrices (como se mencionó en la sección 3.1.2).

La figura 3.18 muestra la matriz «texto / palabra» (histograma) de títulos. El eje horizontal corresponde a las palabras ordenadas de izquierda a derecha en orden cronológico, de su primera aparición en un título de conferencia (o si no en un título de artículos). El eje vertical corresponde a los documentos (conferencias concatenadas a títulos de artículos para años específicos) en orden cronológico de arriba abajo; primero títulos de conferencias y después títulos de artículos. La frecuencia de cada palabra en cada documento es representada por la intensidad del color. Las palabras nuevas que aparecen en un documento dado (concatenando conferencia o títulos de artículos para un año) están claramente visibles como rectángulos grises, y la cantidad de palabras nuevas está caracterizada por el cuadrado de este rectángulo. Se observa que las palabras que aparecen en los títulos de conferencias se han considerado antes en los títulos de artículos: las conferencias se organizan basadas en las discusiones realizadas en los artículos.

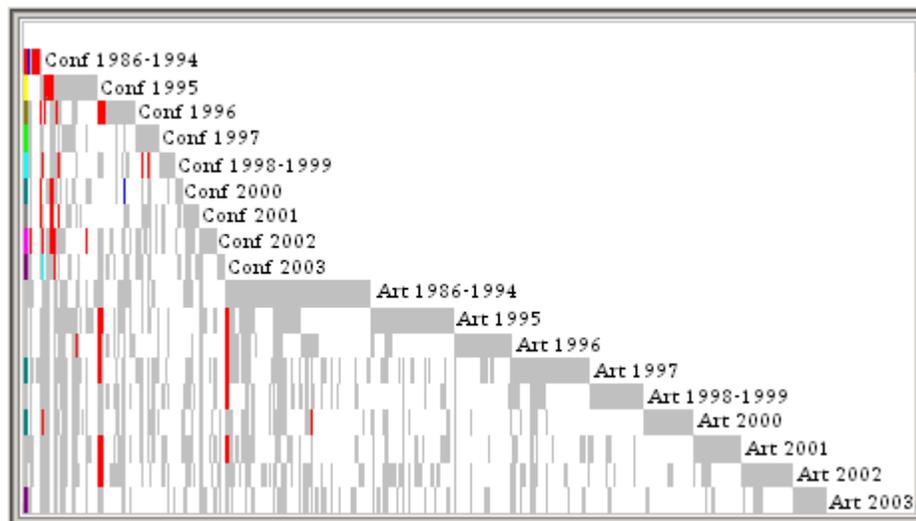


Figura 3.18. «Foto temporal» de la matriz «texto / palabras» de los títulos de conferencias y títulos de artículos.

En la figura 3.18. Se observa dos características importantes de la distribución de palabras:

- Los títulos de conferencias contienen el 23% de DOD; 13% de estas palabras (el 3% de la cantidad total de palabras en DOD) representan los títulos de artículos, y el 77% pertenece a los títulos de conferencias. Esto apoya la idea que el 23% del DOD pertenece a los títulos de conferencia y son de hecho palabras de un nivel superior de abstracción.

- El número de nuevas palabras en **títulos de conferencias** de 1990 a 1997 disminuye (los rectángulos grises se hacen cada vez más pequeños desde las conferencias de 1995 y alcanzan su mínimo en 2000), y una ola de nuevas palabras ocurre del 2001 al 2002 (los rectángulos grises crecen más para las conferencias del 2001 al 2002). Un efecto similar se nota levemente para los títulos de artículos. Esto indica un cierto cambio de interés en las principales áreas de investigación; demostrado por el cambio de temática en las conferencias.

Esta última observación permite bajo investigación subdividir el período en dos etapas: de 1990 a 1997 y 1998 al 2004. La comparación de estos dos estados permite determinar las tendencias de desarrollo de la ciencia. Es más, un análisis similar para un período más grande, permite detectar un mayor número de etapas, y así, mas detalles.

Para comparar los dos sub-períodos, se construyen dos matrices de «texto / palabras» (con los mismos atributos de la figura 3.18), uno para cada conjunto de textos, correspondiente a cada período (1990-1997 y 1998-2004). Cada conjunto contiene textos de distinto nivel de abstracción (títulos de conferencias, títulos de artículos, y resúmenes). Usando los datos correspondientes a cada período se construyen dos ontologías. Cada ontología está basada en tres matrices «texto / palabras» (títulos de conferencias, títulos de artículos, y resúmenes), según se describe en la sección siguiente.

3.3.2 Construcción de la Ontología.

En la figura 3.19 se muestra la matriz «texto / palabra» para el período 1998-2004 (esta matriz es trasladada con respecto a la figura 3.18, es decir, sus filas corresponden a los términos y las columnas a los textos). En esta matriz se distingue un **cluster A** de palabras. Obviamente estas palabras también se representan en los textos de nivel inferior de abstracción, pero con una intensidad más baja y reducida.

Estas palabras pertenecen a dos niveles distintos de abstracción:

- La raíz, el mayor nivel de abstracción de la ontología **LA#1** (que es “**level of abstractness**” numero 1) representada por las palabras sobre computación PCDS y sus sinónimos;
- Incluso un nivel de abstracción mayor: puesto que la ontología es una parte de una súper ontología, por decir, de ciencias de la computación; estos clústeres contienen también las palabras de este supersistema, es decir, términos generales de ciencias de la computación.

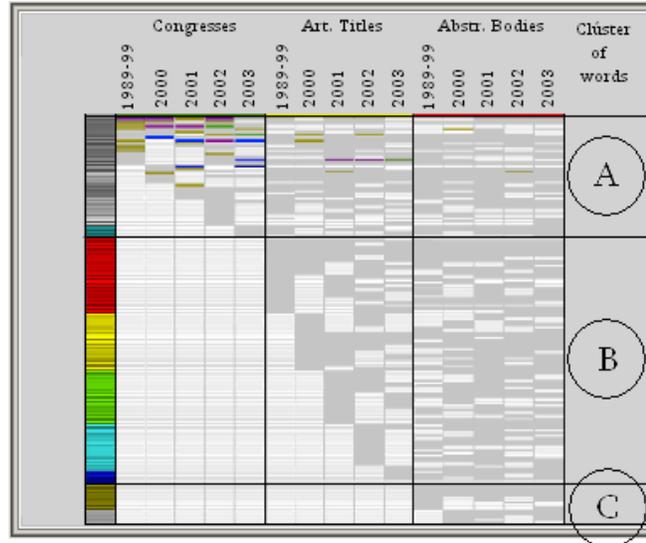


Figura 3.19. Distribución de cluster de palabras por tipos de documentos (títulos de conferencia, títulos de artículos y resúmenes).

Para excluir las palabras mencionadas en el último apartado se construye el DOD, utilizando alguna otra rama de ciencias de la computación (se selecciona el dominio de computación suave) y se eliminan las palabras comunes de esta rama y en DOD. Para alcanzar una estadística más confiable, se agrupan las palabras que coocurren y que forman conceptos. En el experimento, el 25% de todas las palabras, producen el 80% de conceptos de una sola palabra. Los «clústeres complejos» no triviales restantes, contienen el 75% del DOD; el número de conceptos es aproximadamente el 20% de esas palabras, como se muestra en la figura 3.20.

La subdivisión de conceptos (palabras) en los clústeres: A, B y C es la forma más simple de ontología (llamada preontología). De hecho, es importante determinar la distribución de palabras en los clústeres B y C para cada sub-ontología, es decir, establecer las relaciones **padre-hijo** entre éstas palabras y palabras de LA#1. Para cada vértice (concepto) del nivel dado LA#1, se seleccionan los textos de los resúmenes por debajo de él. Con esto se construye una sub-ontología (creando su propio DOD, matrices del tipo «texto / palabra», y revelando clústeres de palabras y conceptos) para cada nuevo sub-dominio, según lo discutido anteriormente. Así, el proceso de construcción de la ontología es recursivo: los niveles inferiores son construidos por una simple subdivisión de conceptos, bajo un nodo de nivel superior.

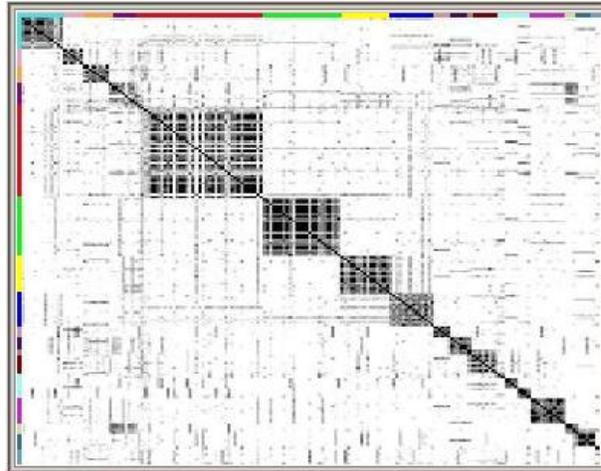


Figura 3.20. Cluster de la matriz «palabra / palabra» de la sub-base de la palabra VLSI para 1998-2004.

3.3.3. Resultados experimentales.

La motivación inicial era detectar tendencias temporales en el desarrollo de las ciencias de la computación. Por consiguiente, se experimentó con dos sistemas de resúmenes sobre conferencias, para los períodos de 1990-1997 y 1998-2004. A continuación se muestran ejemplos de las dos ontologías construidas. Las palabras son representadas por raíces (no por la forma normal), por ejemplo, **toleran** corresponde a **tolerance** y **tolerant**. Los conceptos no triviales (synset, formados por varias palabras) se muestran con llaves «{}», en tanto que los temas no triviales (grupos de conceptos relacionados) se muestran con corchetes «[]». Tales grupos de synset se obtuvieron agrupando las palabras con coocurrencias muy similares [24].

En la figura 3.21 se muestra la parte más importante de la ontología: el nivel raíz (PCDS) es el concepto más abstracto, realmente es el tema de análisis. El primer nivel, es el que se encuentra abajo de la raíz; para este nivel existen siete conceptos que se encuentran únicamente en el primer período (1990-1997). En el segundo período (1998-2004) existen cuatro conceptos, y para ambos períodos (1990-2004) seis conceptos. Con las cifras anteriores se puede decir que el núcleo de este tema conserva el 35.3% de información. Dos de los cuatro conceptos del segundo período están formados por varias palabras. El primer período consta de once palabras; el segundo de veinte y para ambos períodos se tiene ocho palabras.

- Root parallel, concurrent, distributed, simultaneous (computing).
- 1st level, below the root:
- Both periods: {analysis, network, vlsi}, fault, orient, transaction, volum, toleran;
 - 1990–1997: {graphic, model, securit, communicat, test}, frontier, massive, optic, real, reliabilit, reliabl;
 - 1998–2004: {autonom, defect, discret, event, foundation, generat, grid, integrat, interact, storag, technologi, tool}, {circuit, date, evolvabl, interconnect, languag, requirement}, knowledg, object.
- 2nd level, (example only for one concept) below toleran:
- Both periods {fault_toleranc}, inject, object, system;
 - 1990–1997: {allocat, spar}, barri, board, network, critic, efficient, execut, orient, reliabl;
 - 1998–2004: adapt, adaptat, alternat, amplifi, analog, communicat, control, controller, cost, determin, enhanc, etern, evolut, filter, flexibl, immun, java, motor, path, platform, schedul, test, tool, trigger, upgrad.

Figura 3.21. Parte más importante de la ontología.

En el análisis por palabras, los subtemas están más desarrollados, aunque el número de estos disminuye; esto se puede explicar como la transición del período de búsqueda en anchura a búsqueda en profundidad. Por ejemplo el tema de Sockets (figura 3.12) muestra interés a finales del siglo XX y casi desaparece a comienzos del siglo XXI; PVM tiene una tendencia decreciente. Los temas conectados con: **grid**, **integrat** **interconnect** sustituyen los temas conectados con problemas de **optic**, **grafic**.

Más interesantes son los conceptos: **analysis**, **network** y **vlsi** donde la palabra clave es VLSI y los conceptos Fault y Tolerance. En los niveles más bajos de la ontología, estas palabras se encuentran casi siempre juntas. En la figura 3.22 se observa que en el segundo nivel (menos abstracto) Fault y Tolerance forman el concepto **Fault-Tolerance**. Se puede decir que los conceptos VLSI y Fault-Tolerance se encuentran enlazados debido a que VLSI está conectado directamente con el tema de hardware y, Fault-Tolerance con los problemas más actuales de VLSI. Para Fault-Tolerance, se tienen nueve conceptos para el primer período; veinticinco para el segundo y cuatro para ambos períodos. Estas cifras muestran la tendencia de Fault-Tolerance en la segunda etapa. Sumando los conceptos del primer, segundo y ambos períodos da un total de treinta y ocho conceptos. El núcleo de palabras que se encuentran en todo el período son solamente cuatro, lo que implica que el núcleo conserva solamente el 10.5% de información. La parte que no cambia demuestra la evidencia de que la dinámica en esta temática es muy alta.

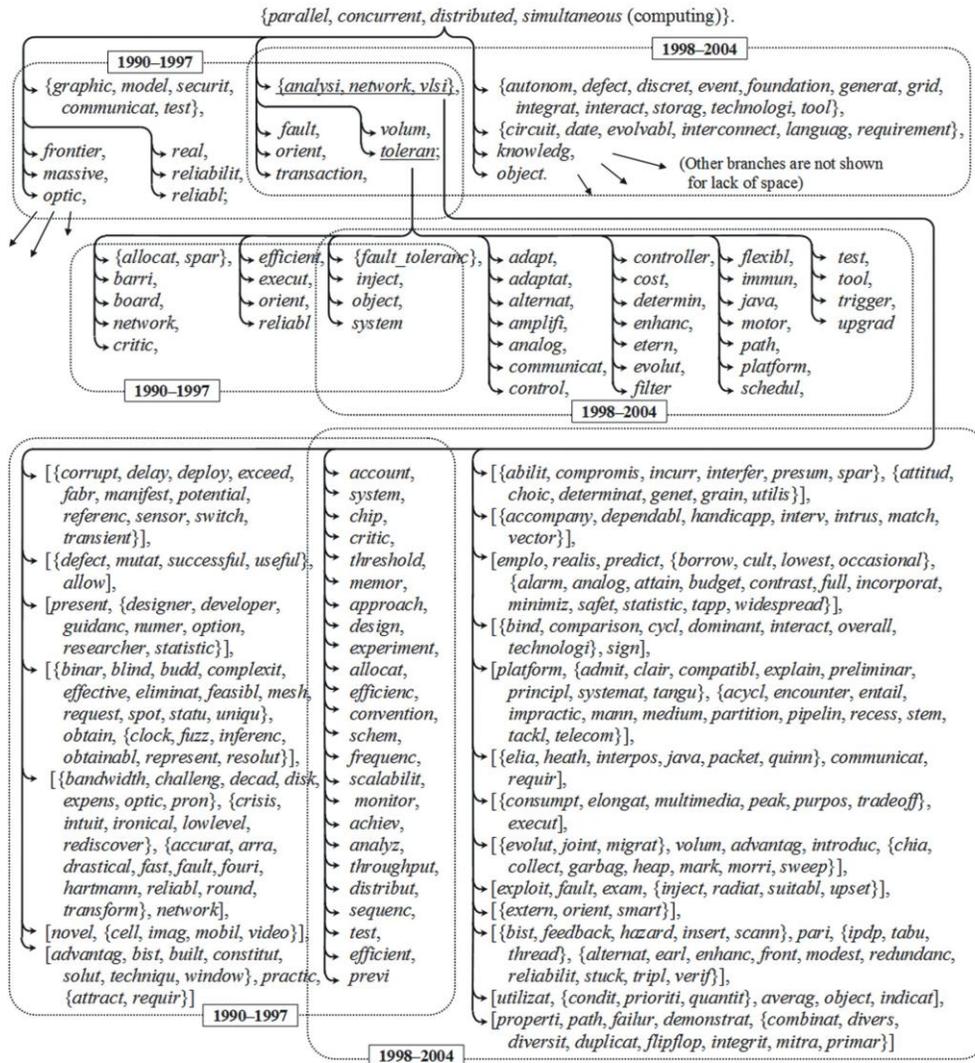


Figura 3.22. Ontología obtenida para los períodos 1990 a 1997 y 1998 a 2004.

Considerando el mismo nivel para VLSI, en donde se tienen siete conceptos construidos de ochenta y un palabras para el primer período; trece conceptos construidos de ciento cuarenta y tres palabras para el segundo período y, veinticuatro conceptos en ambos períodos. Se tienen que el núcleo conserva 54.5% de información. Esta temática es aun más estable, ya que cuenta con un diccionario más desarrollado, tanto en el primer como en el segundo período; en este último se tiene el doble de conceptos nuevos que muestran la tendencia de VLSI. El análisis de la estructura y la dinámica de ontologías para estos casos parciales, proporciona la dirección de mayor interés para este período.

La cantidad de resúmenes relacionados con VLSI y Fault-tolerance (604 resúmenes), obliga a investigar este subtema más detalladamente. Con VLSI se puede investigar la perspectiva de desarrollo de computación PCDS, ya que cuenta con datos más numerosos. En otras palabras tiene un muestreo más representativo. Es el tema de investigación para el próximo capítulo, en los cuales se implementan modelos de desarrollo para conocer la tendencia tanto de VLSI como de computación PCDS.

3.3.4. Estabilidad de la Ontología como medida de evolución.

Para determinar la estabilidad del procedimiento de construcción de la ontología, se comparan tres muestras de textos: dos muestras (de tamaño similar) sobre conferencias y artículos para el período 1990-1997, y la unión de ambas. Por consiguiente, se obtiene tres preontologías distintas. Para medir la distancia entre ellas, se utilizó el **coseno métrico**. Antes de medir esta distancia, se excluyen algunas palabras del sistema y supersistema: números, topónimos y nombres de eventos o de documentos (tales como talleres, conferencia, etc).

Para las primeras y las terceras muestras (comunes) se obtienen sesenta y dos palabras como vectores de su distribución anual; con el coseno 0.943, con la desviación estándar 0.149, y diez palabras que no son representadas en ninguna de las dos muestras (el coseno es cero). Sin embargo, una característica importante de la estabilidad es el año de aparición. Por ejemplo, de ochenta y nueve pares de palabras existentes en las dos primeras muestras, trece tienen diferencias en los primeros años de aparición. Considerando la diferencia anual de la primera aparición (como la distancia entre pares de palabras) se obtiene una distancia de veintitrés entre estas muestras. Lo cual se puede normalizar por la distancia máxima (7 años) y por la cantidad de palabras. Este algoritmo es similar al algoritmo de **distancia de edición en árboles**.

Una ontología es una jerarquía no un árbol, entonces algunas palabras pueden tener más de un padre. Sin embargo, se puede medir por separado los cambios en ambos períodos. En este caso se utiliza el criterio de evolución (discrepancia total entre las ontologías para los dos períodos: 1990-1997 y 1998-2004), el cual consiste de tres parámetros:

- aparición o desaparición de palabras;
- cambios de los niveles de las palabras;
- cambio de las palabras padres.

Para el primer caso, se puede decir que funciona bastante bien; la velocidad de cambio y la frecuencia de palabras para el primer y segundo período, muestran muy buena dinámica de desarrollo del tema.

Para el segundo caso no es muy frecuente, ya que existen algunos ejemplos con palabras **fault**, no solamente en la rama Fault-Tolerance sino también en VLSI (como conceptos distintos en el primer y segundo período). Estos parámetros proporcionan ejemplos de cambio de nivel y cambio de padre en el mismo nivel, con excepción de las palabras **fault**, **evolut** y **alternat** que tienen dos padres e incluyen conceptos distintos.

Conclusiones parciales.

El análisis de textos agrupados por año de publicación es opuesto al análisis de resúmenes individuales (sin agrupar); el primero da una visión más clara de la dinámica de desarrollo de los temas, es decir que la suma anual de textos se utiliza como un nuevo objeto de investigación, para luchar contra el efecto de datos con conocimiento pobre. El análisis de datos numerosos, facilita la posibilidad de descubrir el cambio de interés, en las direcciones más importantes de computación PCDS, durante el período 1997-1998. Se encontró que dos clústeres (cluster 1 y cluster 3) cambian su valor en este período y tienen una productividad casi constante antes y después de este período. El cluster 1 obtiene mayor importancia que el cluster 3, el cual representa el cambio de época en el desarrollo de computación PCDS, como muestra el escalón en el período 1997-1998 de la figura 3.11.

La idea de ontología basada en cadena (tema, concepto y palabra) es productiva, porque da la posibilidad de estructurar las palabras en el diccionario orientado a dominio de manera jerárquica, y al mismo tiempo proporciona claridad sobre el nivel de abstracción de cada palabra en el diccionario orientado a dominio. Realmente la ontología es una estructura jerárquica del diccionario orientado a dominio, que permite analizar el cambio temporal de ésta. De tal manera es posible ver el cambio de valor en la palabra, cuando ésta baja o sube de nivel jerárquico, o cuando las palabras de menor abstracción toman posición en los conceptos o perdieron su valor; pasando de un nivel de conceptos, a un nivel de palabras. Lo cual muestra pérdida de interés en los subtemas. Con esto, no solamente se tiene un método para construcción de ontologías, sino también la posibilidad de investigar y medir la dinámica de desarrollo del tema, a través de la dinámica de la ontología, la cual es la diferencia entre las estructuras del diccionario orientado a dominio para los dos lapsos de tiempo.

En el próximo capítulo se acepta este límite o frontera, como un hecho para agrupar los resúmenes antes y después de este escalón (tercer criterio de la sección 3.1.1). Los resultados de este capítulo proporcionan información relevante, para que un «experto» en computación PCDS construya una ontología. Si no se cuenta con un «experto» o conjunto de expertos en esta área, entonces se muestra que es posible sustituir a estos, por información contenida en libros, revistas y congresos. Por ejemplo, si se tienen palabras clave de DOD en algunos textos anuales y en algunos libros, el nombre de estos libros o sus capítulos, tienen el nombre del subtema que

enlaza a este conjunto de textos anuales; y con esto, se puede revelar lapsos de tiempo importante desde el punto de vista de desarrollo de este subtema (nombre del libro).

Los resultados experimentales muestran que la ontología construida es significativa. Específicamente, puede ser utilizada para el análisis comparativo (del estado y desarrollo) de una rama científica. Se puede aplicar a cualquier jerarquía de textos estructurados, por ejemplo, páginas web HTML. Sin embargo, el uso más probable del método es el PRO, para resultados de más alta calidad.



Parte II

Capítulo 4

Modelos de desarrollo del HSW basado en el estudio de computación PCDS.

Los resultados experimentales de la parte I demuestran que dentro de la estructura jerárquica de computación PCDS, los temas «VLSI» y «tolerancia a fallos» tienen un desarrollo significativo durante todo el período de investigación. Los clústeres relacionados con estos temas son numerosos en ambos períodos (1990-1997 y 1998-2004). En la ontología se manifiesta una relación intrínseca entre VLSI, tolerancia a fallos y computación PCDS (una razón por la que surge éste vínculo es la disminución del tamaño del transistor). Esto hace necesario desarrollar modelos más finos para conocer la tendencia de éstas áreas, y tener la posibilidad de controlar su desarrollo. Para ello se propone el modelo de ciclo de vida (visto en la sección 2.3.4), partiendo del supuesto de que el desarrollo de computación PCDS desde el punto de vista de ciencia

normal tiene un comportamiento similar. Este modelo se aproxima utilizando la función logística y paralelamente un polinomio cúbico. Para el caso del polinomio cúbico, los puntos mínimo, de inflexión y máximo, proporcionan: el nacimiento, máximo interés y muerte de computación PCDS respectivamente.

4.1. Distribución temporal de resúmenes para computación PCDS.

Cuando se presenta un problema científico y su nombre cuenta con algunas «palabras claves»; éstas se pueden encontrar como títulos de artículos, de congresos o de revistas. Según la opinión en los grandes círculos científicos, todo tema suficientemente grande es distribuido en una proporción 1:3 en revistas cuyo título corresponde al nombre del tema; revistas; cuyas secciones contienen el nombre del tema; y otras revistas que no contienen el nombre del tema ni en su título ni en sus secciones.

La hipótesis propuesta es que la misma distribución de artículos sobre computación PCDS se encuentra en la sección de congresos. Una restricción principal es que los congresos sobre este tema no dependen del idioma de trabajo; se consideran solamente congresos en inglés y no se investiga si esto es verdadero para otros idiomas (ruso, chino, etc.), por lo que se concluye en límites de esta hipótesis que es posible construir un muestreo representativo coleccionando todos los artículos relacionados con el tema de computación PCDS de la base de datos de la librería digital de IEEE como se comentó en la sección 3.1. En este caso el muestreo es representativo, pero esto es verdadero si la hipótesis es verdadera.

4.1.1. Distribución de resúmenes agrupados por año de publicación para computación PCDS.

La figura 4.1 muestra el histograma de los datos de la tabla 3.1 de la sección 3.1.1. La gráfica color azul muestra la distribución de los 2300 resúmenes coleccionados sobre computación PCDS para el período que comprende de 1990 al 2004. Se observa que la tendencia es creciente hasta el año 1997. En este año se tiene una colección de 212 resúmenes que tratan sobre temas relativos a computación PCDS. En el año 1998 el interés por estos temas disminuye a 118 resúmenes; una posible causa es que el interés de los investigadores se haya enfocado a nuevos temas de investigación, o el apoyo económico para estos temas se haya reducido. Para el año 2000 se tiene un total de 215 resúmenes, estos temas vuelven a tener interés dentro de la comunidad científica aumentando gradualmente hasta el año 2004, donde se tiene un total de 229 resúmenes. La gráfica en color rojo muestra el tamaño en kBytes de cada uno de los archivos agrupados por año de publicación, según la distribución de la sección 3.1.1. Las dos gráficas

tienen un comportamiento similar a excepción del año 2004, donde la gráfica azul (correspondiente a la cantidad de artículos) tiene un incremento, mientras la gráfica roja (correspondiente al tamaño) sigue un comportamiento decreciente; esto se debe a que la cantidad de palabras en los resúmenes para este año, disminuyó por debajo del promedio (el cual corresponde a 1.2 kB por resumen).

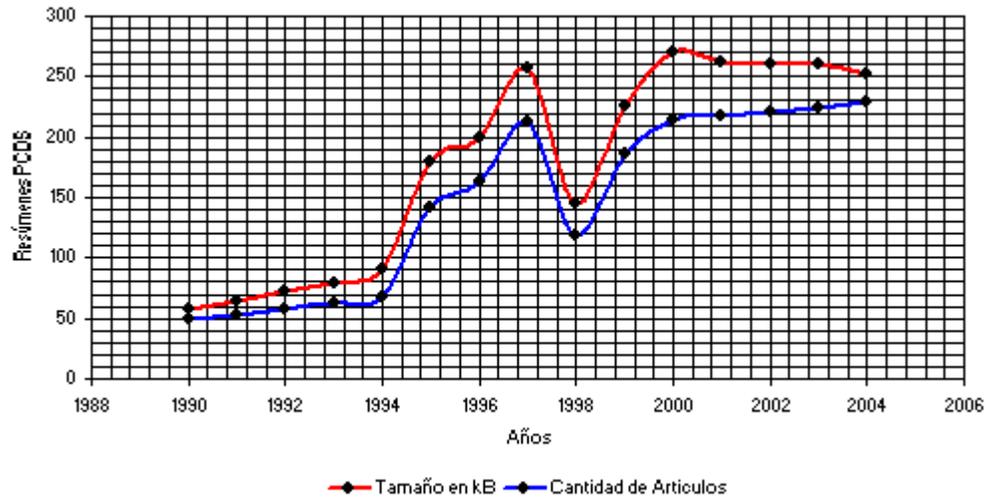


Figura 4.1. Histograma de distribución de los 2300 resúmenes de la Librería Digital de IEEE para Computación PCDS durante el período de 1990 al 2004. La gráfica en color rojo muestra el tamaño en kB y la gráfica en color azul muestra la cantidad de artículos.

4.1.2. Comportamiento individual de computación PCDS.

El desarrollo individual de las ramas que forman computación PCDS se muestra en la figura 4.2, se observa que cada una sigue un comportamiento irregular; en algunos años crecen mientras que en otros decrecen. Esta gráfica es una aproximación del comportamiento individual de cada rama, debido a que existen resúmenes que pertenecen a dos o más áreas de computación PCDS. El criterio de selección, fue darle prioridad a los temas más relevante dentro del resumen. **Computación paralela** tiene un total de 773 resúmenes publicados durante el período 1990-2004, alcanza su máximo desarrollo en 1997, y en 1998 presenta el mayor decremento. **Computación concurrente** tiene su máximo desarrollo en 1996, decrece significativamente en 1998, continúa de forma irregular con una tendencia creciente hasta el año 2001, y a mediados del año 2002 decrece nuevamente, y después continúa con una tendencia creciente hasta el año 2004. **Computación distribuida** es el área con mayor productividad durante el período de investigación, ya que cuenta con un total de 799 publicaciones; alcanza una tendencia creciente hasta 1992; en 1993 tiene un pequeño decremento que es compensado al siguiente año; continúa con una tendencia creciente hasta 1997, al siguiente año muestra el mayor decremento de todo

el período de investigación, al igual que computación paralela. Después presenta una tendencia creciente hasta mediados del año 2002, donde alcanza su mayor productividad con 97 publicaciones, y por último decrece en el año 2004. De las cuatro áreas que forman computación PCDS, el área de **computación simultánea** es la más reciente y con mayores perspectivas de desarrollo. Cuenta con un total de 247 publicaciones para el período de investigación, con ligeros decrementos en los años de 1994, 1998 y 2003. Para los demás años presenta una tendencia creciente hasta el año 2004. En la figura se observa que el par **paralelo-distribuido** presenta una tendencia decreciente a partir del año 2003, en tanto que el **par concurrente-simultáneo** muestra una tendencia creciente a partir de ese mismo año. En las próximas secciones se realiza un estudio minucioso basado en modelos matemáticos, para determinar el desarrollo temporal de computación PCDS.

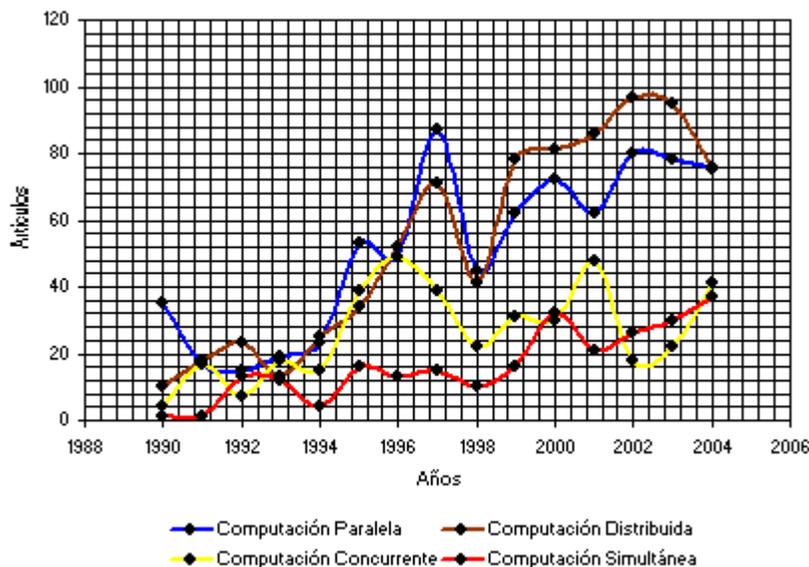


Figura 4.2. Histograma de distribución de los 2300 resúmenes, separados por tema. La gráfica en color azul muestra la distribución de computación Paralela, en color amarillo computación Concurrente, en color café computación Distribuida y en color rojo computación Simultánea.

4.1.3. Curva de distribución acumulada y curva S para computación PCDS.

No es posible conocer el tiempo exacto o el lugar donde emergerá un nuevo descubrimiento dentro de una rama científica, pero es posible predecir el desarrollo de su ciencia normal, la cual se basa en resultados anteriores, y el estudio de estos antecedentes da la posibilidad de predecir la tendencia de su desarrollo. El estudio de la ciencia normal para computación PCDS, permite conocer las etapas de inicio, máximo interés, así como la etapa cuando ésta será obsoleta. La gráfica color rojo de la figura 4.3 muestra la distribución acumulada para los 2300 resúmenes, mientras que la gráfica color azul muestra la línea de tendencia de la serie de datos. Esta curva

ascendente tiene un carácter exponencial en forma de **curva J** [37]. Según la **ley de Verhulst** [URL-13] ningún sistema con crecimiento exponencial puede seguir indefinidamente. Eventualmente, se alcanza un límite después de cierto período, y la curva J va tomando la forma de una curva S. Aprovechando la información de la curva J es posible restaurar completamente la curva S. Esta última curva se puede aproximar mediante un polinomio cúbico o la función logística; para conocer la discrepancia se implementan ambos métodos.

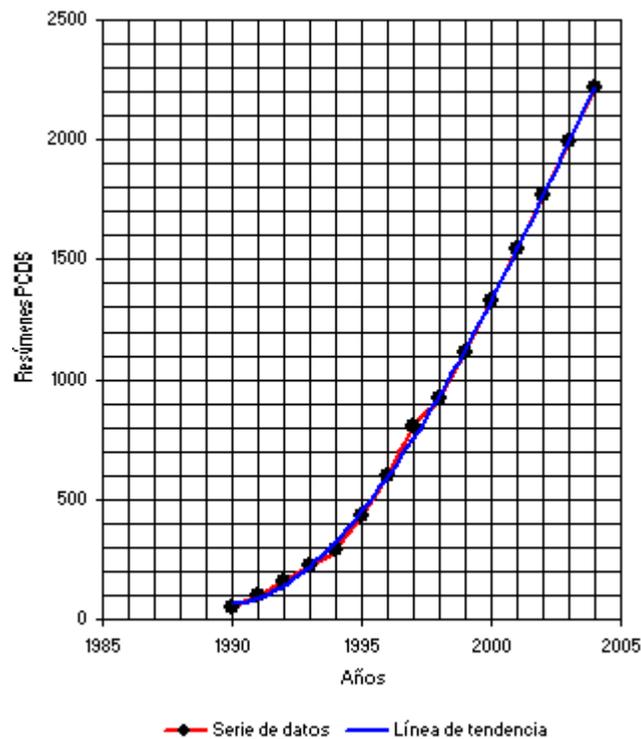


Figura 4.3. Distribución acumulada (color rojo) y aproximación de la curva mediante un polinomio de tercer orden (curva azul) para computación PCDS.

4.1.3.1. Ajuste del polinomio cúbico por mínimos cuadrados.

Supongamos que existe una relación funcional $y = f(x)$ entre dos cantidades x e y , con f desconocida y se conocen valores y_k que aproximan a $f(x_k)$, es decir:

$$f(x_k) = y_k + \epsilon_k \quad \text{donde } k = 0, 1, 2, \dots, n \quad (4.1)$$

con ϵ_k desconocido. Se trata de recuperar la función f a partir de los datos aproximados y_k , $k = 0, 1, \dots, n$. Este problema se conoce como «ajuste de datos» o «ajuste de curvas» (caso discreto)

Si f es una función polinómica, digamos $f(x) = P_m(x)$, entonces el problema se convierte en:

Dados $n + 1$ puntos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ con x_0, x_1, \dots, x_n números reales distintos, se trata de encontrar un polinomio:

$$P_m(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m \quad (4.2)$$

con $m < n$ que «mejor se ajuste» a los datos. Lo de «mejor ajuste» se entenderá en el sentido de que:

$$\left[\sum_{k=0}^n (P_m(x_k) - y_k)^2 \right]^{1/2} \quad (4.3)$$

Sea mínimo, es decir, que

$$\sum_{k=0}^n (P_m(x_k) - y_k)^2 \quad (4.4)$$

Este criterio de mejor ajuste, se conoce como «mínimos cuadrados», y el método para obtener los polinomios que mejor se ajustan según mínimos cuadrados se llama «regresión polinomial».

4.1.3.2. Cálculo de los coeficientes del polinomio cúbico.

Supongamos que se conocen los datos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ con x_0, x_1, \dots, x_n números reales distintos, y se desea encontrar un polinomio

$$P_3(x) = a_0 + a_1x + a_2x^2 + a_3x^3, \quad \text{con } 3 \leq n \quad (4.5)$$

tal que:

$$\varphi(a_0, a_1, a_2, a_3) = \sum_{k=0}^n (P_3(x_k) - y_k)^2 = \sum_{k=0}^n (a_0 + a_1x_k + a_2x_k^2 + a_3x_k^3 - y_k)^2 \quad (4.6)$$

sea mínima.

El grado del polinomio $P_m(x)$ se escogió en base a los datos, ya que este polinomio es el que más se aproxima a la forma en que están distribuidos.

Volviendo a la función $\varphi(a_0, a_1, a_2, a_3)$, una **condición necesaria** para la existencia de un mínimo relativo de esta función es que las **derivadas parciales** de $\varphi(a_0, a_1, a_2, a_3)$ con respecto a a_j , $j = 0, 1, 2, 3$ sean cero.

Resultan entonces las siguientes cuatro ecuaciones lineales con incógnitas a_0, a_1, a_2, a_3 :

$$\begin{aligned} \frac{\partial \varphi}{\partial a_0} &= \sum_{k=0}^n 2 a_0 + a_1 x_k + a_2 x_k^2 + a_3 x_k^3 - y_k = 0 \\ \frac{\partial \varphi}{\partial a_1} &= \sum_{k=0}^n 2 a_0 + a_1 x_k + a_2 x_k^2 + a_3 x_k^3 - y_k \quad x_k = 0 \\ \frac{\partial \varphi}{\partial a_2} &= \sum_{k=0}^n 2 a_0 + a_1 x_k + a_2 x_k^2 + a_3 x_k^3 - y_k \quad x_k^2 = 0 \\ \frac{\partial \varphi}{\partial a_3} &= \sum_{k=0}^n 2 a_0 + a_1 x_k + a_2 x_k^2 + a_3 x_k^3 - y_k \quad x_k^3 = 0 \end{aligned} \quad (4.7)$$

De las ecuaciones (4.7) se cancela el número dos, se desarrollan los paréntesis y se usa que:

$$\sum_{k=0}^n a_0 = (n+1) \bar{a}_0 \quad (4.8)$$

se obtiene:

$$\begin{aligned} n+1 a_0 + \left(\sum_{k=0}^n x_k \right) a_1 + \left(\sum_{k=0}^n x_k^2 \right) a_2 + \left(\sum_{k=0}^n x_k^3 \right) a_3 &= \left(\sum_{k=0}^n y_k \right) \\ \left(\sum_{k=0}^n x_k \right) a_0 + \left(\sum_{k=0}^n x_k^2 \right) a_1 + \left(\sum_{k=0}^n x_k^3 \right) a_2 + \left(\sum_{k=0}^n x_k^4 \right) a_3 &= \left(\sum_{k=0}^n x_k y_k \right) \\ \left(\sum_{k=0}^n x_k^2 \right) a_0 + \left(\sum_{k=0}^n x_k^3 \right) a_1 + \left(\sum_{k=0}^n x_k^4 \right) a_2 + \left(\sum_{k=0}^n x_k^5 \right) a_3 &= \left(\sum_{k=0}^n x_k^2 y_k \right) \\ \left(\sum_{k=0}^n x_k^3 \right) a_0 + \left(\sum_{k=0}^n x_k^4 \right) a_1 + \left(\sum_{k=0}^n x_k^5 \right) a_2 + \left(\sum_{k=0}^n x_k^6 \right) a_3 &= \left(\sum_{k=0}^n x_k^3 y_k \right) \end{aligned} \quad (4.9)$$

Este es un sistema de cuatro ecuaciones lineales con 4 incógnitas a_0, a_1, a_2, a_3 , llamado **sistema de ecuaciones normales**. Este sistema se puede escribir en forma simplificada como sigue:

$$\sum_{i=0}^3 a_i \sum_{k=0}^n x_k^{i+j} = \sum_{k=0}^n x_k^j y_k \quad \text{con } j = 0, 1, 2, 3 \quad (4.10)$$

Estas ecuaciones se pueden reproducir a partir de:

$$p_3 \ x_k = a_0 + a_1 x_k + a_2 x_k^2 + a_3 x_k^3 = y_k \quad (4.11)$$

Multiplicando a ambos lados por x_k^j , $j = 0, 1, \dots, 3$

$$a_0 x_k^j + a_1 x_k^{1+j} + a_2 x_k^{2+j} + a_3 x_k^{3+j} = x_k^j y_k \quad (4.11a)$$

Sumando sobre k

$$a_0 \sum_{k=0}^n x_k^j + a_1 \sum_{k=0}^n x_k^{1+j} + a_2 \sum_{k=0}^n x_k^{2+j} + a_3 \sum_{k=0}^n x_k^{3+j} = \sum_{k=0}^n x_k^j y_k \quad \text{con } j = 0, 1, 2, 3 \quad (4.12)$$

La ecuación (4.12) permite calcular los coeficientes del polinomio de la ecuación (4.5); los cuales se muestran a continuación:

$$\begin{aligned} a_0 &= 63.2000 \\ a_1 &= 10.5024 \\ a_2 &= 15.0394 \\ a_3 &= -0.3422 \end{aligned} \quad (4.12a)$$

Para simplificar el cálculo de éstos coeficientes, fue necesario realizar un cambio de escala en el eje de las abscisas; sustituyendo 1990 por cero, 1991 por uno y así sucesivamente hasta el año 2004 al cual le corresponde 14. Este cambio de escala se muestra en la figura 4.4.

Ahora, sustituyendo los coeficientes obtenidos, en la ecuación (4.5), se obtiene el siguiente polinomio:

$$P_3(x)_{PCDS} = -0.3422x^3 + 15.0394x^2 + 10.5024x + 63.2000 \quad (4.13)$$

4.1.4. Curva S de computación PCDS aproximada con un polinomio cúbico.

Para conocer los puntos mínimo y máximo del polinomio cúbico de la ecuación (4.13), se calcula la primera derivada:

$$P_3'(x) = -1.0266x^2 + 30.0788x + 10.5024 \quad (4.14)$$

Con la fórmula general de segundo grado se obtienen los puntos mínimos y máximo del polinomio.

$$\begin{aligned}x_{\min(PCDS)} &= -0.3451 \\x_{\max(PCDS)} &= 29.6446\end{aligned}\tag{4.15}$$

Las rectas horizontales que pasan por los puntos $x_{\min(PCDS)}$, $x_{\max(PCDS)}$, se encuentran al sustituir los puntos mínimo y máximo de la ecuación (4.15) en la ecuación (4.13).

$$\begin{aligned}y_{\min(PCDS)} &= 61.3808 \\y_{\max(PCDS)} &= 4676.292\end{aligned}\tag{4.16}$$

Conocer el punto de inflexión es esencial debido a que nos revela el momento en el que la ciencia normal pasa de un estado de desarrollo progresivo a un estado de desarrollo decadente. El punto de inflexión se presenta donde cambia la concavidad de la gráfica, el signo de la segunda derivada $P_3''(x)$ cambia en estos puntos.

$$P_3''(x) = -2.0532x + 30.0788\tag{4.17}$$

De manera que, para localizar el punto de inflexión, se calcula los valores de x para los que $P_3''(x) = 0$

$$x_{\text{inf}(PCDS)} = 14.6497\tag{4.18}$$

Sustituyendo la ecuación (4.18) en la ecuación (4.13), se tiene la ordenada del punto de inflexión:

$$y_{\text{inf}(PCDS)} = 2368.8519\tag{4.19}$$

Estos resultados se muestran en la figura 4.4. Se observa que el polinomio se encuentra entre las rectas horizontales $y_{\min(PCDS)} = 61.3808$ y $y_{\max(PCDS)} = 4676.2924$ que pasan por los puntos $x_{\min(PCDS)} = -0.3451$, $x_{\max(PCDS)} = 29.6446$ respectivamente. Estas rectas horizontales sirven de asíntotas al polinomio.

La ecuación (4.18) proporciona el punto de inflexión del polinomio cúbico, el cual tiene un valor de $x_{\text{inf}(PCDS)} = 14.6497$ y representa al mes de julio del 2004. En esta fecha se manifestó el mayor progreso de computación PCDS. Para fechas posteriores, el interés decrece hasta el punto $x_{\max(PCDS)} = 29.6446$, el cual corresponde al mes de agosto del 2019. Momento en que el tema de computación PCDS, llegará a la etapa de decadencia.

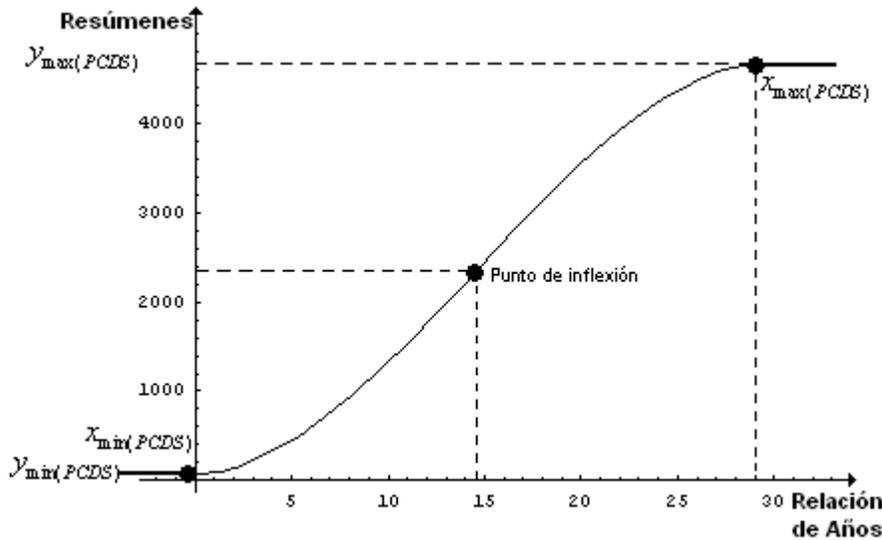


Figura 4.4. Aproximación de la curva S de computación PCDS, con el polinomio cúbico de la ecuación 4.13.

El punto $x_{\min(PCDS)} = -0.3451$ indica el inicio de computación PCDS, el cual representa el mes de abril de 1989. Esta fecha no es exacta, ya que existen artículos publicados antes de este tiempo que hablan sobre computación PCDS. Una razón es que las investigaciones sobre computación PCDS en la década de los 80's se realizaban de manera secreta en supercomputadoras Cray y todos los resultados eran reservados para fines militares o gubernamentales [URL-14]. Además en esta fecha no había congresos relacionados con el tema de computación PCDS, lo que hace difícil determinar el valor mínimo o nacimiento exacto de esta rama científica. Para poder obtener resultados reales es necesario hablar de ciencia normal abierta, lo cual implica tener congresos y difusión abierta de la ciencia.

4.1.5. Curva S de computación PCDS aproximada con la función logística.

En 1939 Graham propuso un modelo matemático bastante simple, el cual describe el crecimiento de una población, relacionando la tasa instantánea de incremento neto, con el tamaño o la densidad de la población [38]. Este modelo se basa en el concepto que, bajo condiciones de equilibrio, la tasa instantánea de incremento de la población, es directamente proporcional al tamaño de la misma población, y a la cantidad de población que falta para llegar a ocupar plenamente el espacio, o ambiente disponible. Lo anterior se expresa por la ecuación diferencial:

$$\frac{dl}{dt} = \frac{\beta l(\alpha - l)}{\alpha} \quad (4.20)$$

donde: l es el tamaño o densidad de la población; α la población máxima que el ambiente puede soportar, esto es, representa la capacidad de carga del medio, que se supone constante; β es una constante que representa la tasa neta de incremento específico de la población, cuando la densidad de la población y el efecto limitante del ambiente se aproxima a cero; t es el tiempo.

En esta ecuación no es posible identificar ninguno de los parámetros poblacionales, ya que lo que se toma en cuenta es el resultado final, es decir, la tasa de cambio de la población en función al tamaño de la población misma.

Integrando la ecuación (4.20) se obtiene la ecuación siguiente:

$$l_{PCDS} = \frac{\alpha}{1+e^{-\beta(x-\gamma)}} \quad (4.21)$$

que describe el crecimiento en peso de una población y corresponde a una curva logística de Verlhust, donde l_{PCDS} es la población en cualquier instante x . Como punto de referencia para medir el tiempo se toma el momento γ , en que la población alcanza la mitad de su crecimiento (es decir, cuando $x = \gamma$ y $l = \frac{\alpha}{2}$) y corresponde al punto de inflexión de la curva logística. El modelo de regresión logística se fundamenta en una curva en forma de S alargada, acotada en el intervalo abierto $[0, \alpha]$ tal que el valor esperado de la variable independiente tiende a cero cuando los valores de las variables explicativas tienden a ser muy grandes y convergen a α . La función logística se aplica a una serie de fenómenos cuyo denominador común es el proceso de desarrollo. Así, la curva logística describe adecuadamente la tendencia de producción de la ciencia normal para computación PCDS, tomando en cuenta que durante el período de análisis no existen innovaciones radicales (ciencia revolucionaria). El punto de inflexión se obtiene cuando $x = \gamma$.

$$l_{\text{inf}(PCDS)} = \frac{\alpha}{2} \quad (4.22)$$

Esto quiere decir que el punto de inflexión de la función logística se encuentra a la mitad del crecimiento, y representa el momento en el que el crecimiento desaparece y se inicia el decremento del desarrollo hasta llegar a α . Para hallar el valor de α , la condición necesaria y suficiente es que los puntos de inflexión del polinomio cúbico y de la función logísticas sean iguales, matemáticamente se puede escribir:

$$P(x)_{\text{inf}(PCDS)} = l(x)_{\text{inf}(PCDS)} \quad (4.23)$$

Calculando la primera derivada del polinomio cúbico de la ecuación (4.5) se tiene:

$$P'_3(x) = 3a_3x^2 + 2a_2x + a_1 \quad (4.24)$$

Utilizando la ecuación general de segundo grado se encuentran los puntos mínimo y máximo de la función.

$$x_{(PCDS)\min} = -\frac{a_2}{3a_3} + \frac{\sqrt{4a_2^2 - 12a_3a_1}}{6a_3} \quad x_{(PCDS)\max} = -\frac{a_2}{3a_3} - \frac{\sqrt{4a_2^2 - 12a_3a_1}}{6a_3} \quad (4.25)$$

El punto de inflexión para el polinomio cúbico esta dado por:

$$P_{\inf(PCDS)} = \frac{x_{\min(PCDS)} + x_{\max(PCDS)}}{2} = -\frac{a_2}{3a_3} \quad (4.26)$$

Sustituyendo la ecuación (4.26) en la ecuación (4.5) se obtiene la ecuación general para encontrar el punto de inflexión del polinomio cúbico

$$P_{\inf(PCDS)} = \frac{2}{3} \left(\frac{a_2^3}{(3a_3)^2} \right) - \frac{a_2a_1}{3a_3} + a_0 \quad (4.27)$$

Igualando las ecuaciones (4.22) y (4.27) (puntos de inflexión) se obtiene la ecuación para calcular $\alpha_{(PCDS)}$.

$$\alpha_{(PCDS)} = 2 \left[\frac{2}{3} \left(\frac{a_2^3}{(3a_3)^2} \right) - \frac{a_2a_1}{3a_3} + a_0 \right] \quad (4.28)$$

Sustituyendo los coeficientes de la ecuación (4.12a) en la ecuación (4.28) se obtiene el valor de $\alpha_{(PCDS)}$.

$$\alpha_{(PCDS)} = 4737.7038 \quad (4.28a)$$

La condición necesaria y suficiente para calcular el valor de β , es que al derivar la ecuación (4.23) se siga cumpliendo la igualdad, esto es:

$$P'(x)_{\inf(PCDS)} = l'(x)_{\inf(PCDS)} \quad (4.29)$$

Sustituyendo la ecuación (4.26) en la ecuación (4.24) se tiene la primera derivada del polinomio cúbico en el punto de inflexión, esto es la parte izquierda de la ecuación (4.29).

$$P'(x)_{\inf(PCDS)} = a_1 - \frac{a_2^2}{3a_3} \quad (4.30)$$

Derivando la ecuación (4.21) se tiene:

$$l'_{(PCDS)} = \frac{\alpha\beta e^{-\beta(x-\gamma)}}{1+e^{-\beta(x-\gamma)}^2} \quad (4.31)$$

El punto de inflexión se obtiene cuando $x = \gamma$, lo que implica que la ecuación (4.31) se convierte en

$$l'(x)_{\text{inf}(PCDS)} = \frac{\alpha\beta}{4} \quad (4.32)$$

igualando las ecuaciones (4.30) y (4.32) bajo la condición dada por la ecuación (4.29), se tiene que:

$$\beta_{(PCDS)} = \frac{4}{\alpha} a_1 - \frac{a_2^2}{3a_3} \quad (4.33)$$

sustituyendo los coeficientes de la ecuación (4.12a) en la ecuación (4.33) se obtiene el valor de: $\beta_{(PCDS)} = 0.1948$

Como se menciono anteriormente que γ se encuentra a la mitad del crecimiento de la ecuación logística, el cual corresponde al punto de inflexión y esta dado por:

$$\gamma_{(PCDS)} = -\frac{a_2}{3a_3} \quad (4.34)$$

sustituyendo los coeficientes a_2 y a_3 se tiene el valor de γ

$$\gamma_{(PCDS)} = 14.6497 \quad (4.34a)$$

Este valor coincide con el punto de inflexión del polinomio cúbico ecuación (4.18). Con los valores hallados para α , β y γ , la ecuación logística toma la siguiente forma:

$$l_{(PCDS)} = \frac{4737.7038}{1+e^{-0.1948(x-14.6497)}} \quad (4.35)$$

En la figura 4.5 se muestra la gráfica para la ecuación (4.35). Se aprecia que la curva sigue un comportamiento en forma de S. En la parte central se aprecia el punto de inflexión, con coordenadas (14.6497, 2368.8519). Este punto representa la fecha de máximo desarrollo para computación PCDS, así como la cantidad de resúmenes recopilados hasta esa fecha. Estos datos corresponden al mes de Julio del 2004 y a un total de 2368.8519 resúmenes respectivamente. El desarrollo de computación PCDS se encuentra limitada por $\alpha_{(PCDS)} = 4737.7038$; ésta indica el número de publicaciones (resúmenes) recopiladas, desde los primeros años de la década de los

80's (según se aprecia en la parte inferior de la curva), hasta el año 2027; momento en que llegará a su final.

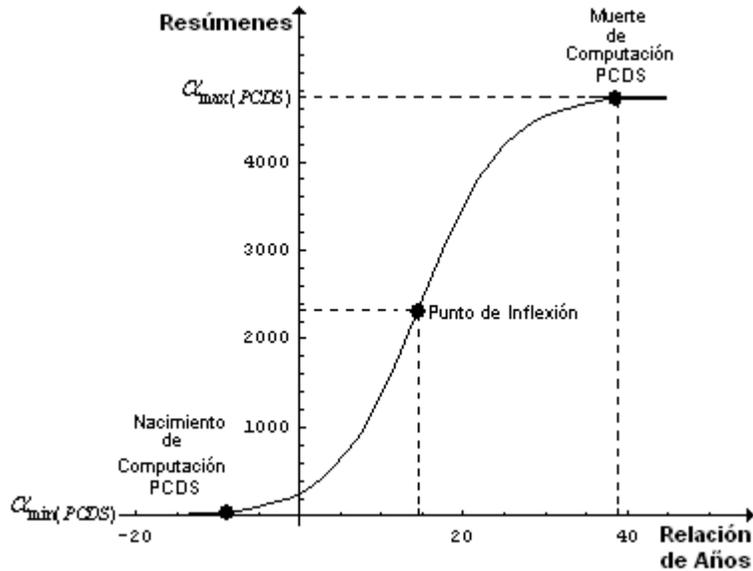


Figura 4.5. Nacimiento, máximo desarrollo y muerte de computación PCDS, según la curva logística.

4.1.6. Superposición de las curvas cúbica y logística para computación PCDS.

Después de realizar el análisis para ambas curvas es necesario conocer la discrepancia que existe entre ellas. Se debe recordar que tanto el punto de inflexión como la inclinación de la tangente en este punto, es la misma para ambas curvas. Esto se muestra en la figura 4.6. La curva punteada corresponde al polinomio cúbico, mientras que la curva continua corresponde a la función logística. La asíntota superior de la ecuación logística se sitúa en $\alpha_{\max(PCDS)} = 4737.7038$, por encima de la recta horizontal superior del polinomio cúbico y se encuentra en $y_{\max(PCDS)} = 4676.2924$. La asíntota inferior de la función logística se localiza en $\alpha_{\min(PCDS)} = 0.0306$, mientras que de la recta horizontal mínima del polinomio cúbico en $y_{\min(PCDS)} = 61.3808$. Los puntos de inflexión coinciden en la parte central de la figura y tienen por coordenadas (14.6497, 2368.8519). En la figura 4.6 se observa que el polinomio cúbico alcanza más rápidamente sus puntos mínimo y máximo, lo que demuestra que tiene un período mas corto de desarrollo, comparado con el modelo de la función logística.

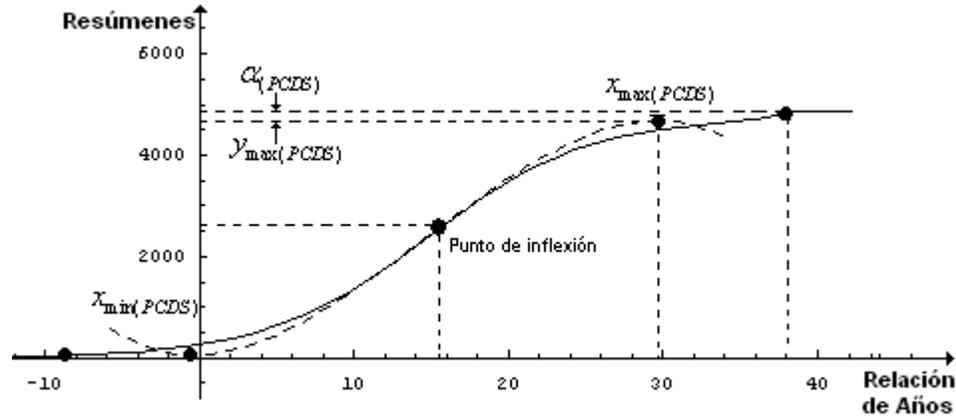


Figura 4.6. Comportamiento de la función logística y el polinomio cúbico para computación PCDS.

Para calcular la diferencia entre la asíntota superior del polinomio cúbico y la asíntota superior de la función logística, se utiliza la siguiente ecuación.

$$\Delta_{\max(PCDS)} = \alpha_{\max(PCDS)} - y_{\max(PCDS)} = 61.4114 \quad (4.36)$$

De igual forma, para calcular la diferencia entre la asíntota inferior del polinomio cúbico y la asíntota inferior de la función logística, se tiene:

$$\Delta_{\min(PCDS)} = \alpha_{\min(PCDS)} - y_{\min(PCDS)} = 61.4114 \quad (4.37)$$

Se observa que el resultado de las ecuaciones (4.36) y (4.37) son iguales, por lo tanto

$$\Delta_{\max(PCDS)} = \Delta_{\min(PCDS)} \quad (4.37a)$$

Lo anterior indica que las curvas son simétricas con relación al punto de inflexión, por lo que, la discrepancia es la misma, tanto en la parte superior como en la parte inferior de la curva, con un error equivalente al 1.3%. Esto se muestra en la figura 4.7.

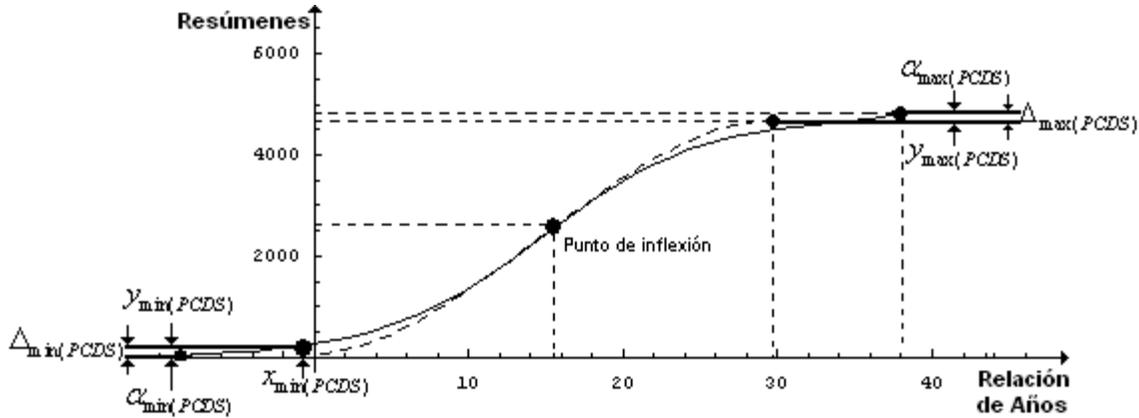


Figura 4.7. Discrepancia entre la función logística y la curva cúbica para Computación PCDS.

4.1.7. Curva S de VLSI y tolerancia a fallos aproximada con un polinomio cúbico y la función logística.

Los temas VLSI y tolerancia a fallos han tenido un importante desarrollo en el área de computación PCDS. Este desarrollo se ha dado de manera recíproca, ya que como ley causa-efecto, los sistemas VLSI utilizan ideas de paralelismo para solucionar los conflictos generados por el aumento en la escala de integración; y a su vez, computación PCDS utiliza los avances tecnológicos obtenidos en VLSI para lograr nuevos retos científicos. Se puede observar por los resultados obtenidos en la parte 1, que la relación entre ambas áreas es muy estrecha, ya que el porcentaje de información de VLSI y tolerancia a fallos contenida en los textos sobre computación PCDS es muy alto; lo que obliga a investigar estos temas mediante modelos más finos como son los utilizados en la sección anterior.

4.1.7.1. Distribución de resúmenes agrupados por año de publicación para VLSI y tolerancia a fallos.

De los 2300 resúmenes recopilados de la librería digital de IEEE, 604 corresponden a temas relacionados con VLSI y tolerancia a fallos como muestra la tabla 4.0 en las columnas 2, 5 y 8 lo que equivale a un 26.26%. Mientras que en las columnas 3, 6 y 9 de la misma tabla se muestra los datos acumulados para el periodo 1990 a 2004.

| Año | Cantidad de Resúmenes | Datos acumulados | Año | Cantidad de Resúmenes | Datos acumulados | Año | Cantidad de Resúmenes | Datos acumulados |
|------|-----------------------|------------------|------|-----------------------|------------------|------|-----------------------|------------------|
| 1990 | 37 | 37 | 1995 | 37 | 187 | 2000 | 54 | 411 |
| 1991 | 26 | 63 | 1996 | 36 | 223 | 2001 | 55 | 466 |
| 1992 | 28 | 91 | 1997 | 57 | 280 | 2002 | 51 | 517 |
| 1993 | 27 | 118 | 1998 | 36 | 316 | 2003 | 49 | 566 |
| 1994 | 32 | 150 | 1999 | 41 | 357 | 2004 | 38 | 604 |

Tabla 4.0. Cantidad de resúmenes y datos acumulados para el periodo 1990-2004 para VLSI-FT.

En la figura 4.8 se muestra el histograma de esta información, y se observa que durante todo el período de investigación presenta un comportamiento similar, al manifestado por computación PCDS de la sección 4.1.1.

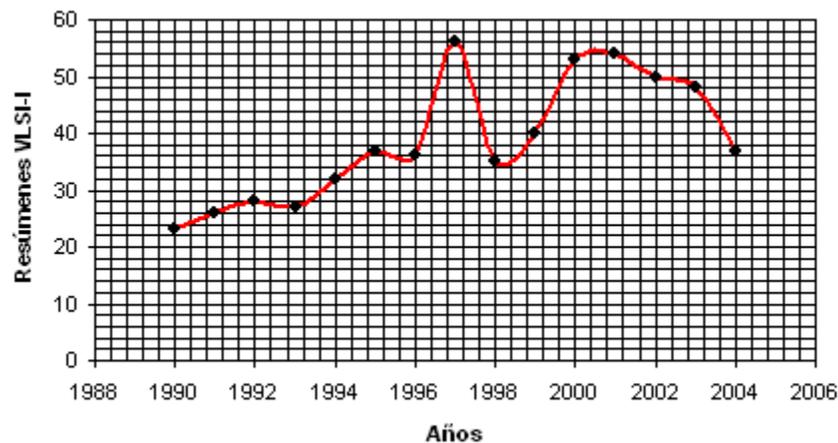


Figura 4.8. Histograma de los 604 resúmenes relacionados con VLSI y Tolerancia a fallos.

En la figura anterior se observa que a partir de 1990 y hasta 1997, la gráfica presenta una tendencia «casi» creciente, a excepción de 1993 y 1996 que manifiestan pequeños decrementos. A mediados de 1997 se tiene el máximo desarrollo, mientras que al finalizar éste año, se inicia un desinterés por estos temas, prolongándose hasta los primeros meses de 1998. Las posibles causas (como se mencionó en la sección anterior) son: que el apoyo económico se haya reducido o que los investigadores hayan elegido un nuevo tema de investigación; ya que también la ciencia tiene temas que se ponen de moda durante ciertos períodos. Continuando con la gráfica, se observa que al final de 1998, se inicia un crecimiento gradual que se extiende hasta el 2001. Para fechas posteriores a éste año, el interés decrece paulatinamente hasta el 2004. A continuación se realiza un estudio para conocer el desarrollo temporal de VLSI y tolerancia a fallos.

4.1.7.2. Curva S y curva de distribución acumulada para VLSI y tolerancia a fallos.

La gráfica en color rojo de la figura 4.9 muestra la distribución acumulada para VLSI y tolerancia a fallos, mientras que la gráfica en color azul muestra la línea de tendencia de esta serie de datos. De la misma forma que se hizo para computación PCDS; la curva J se aproxima mediante un polinomio cúbico.

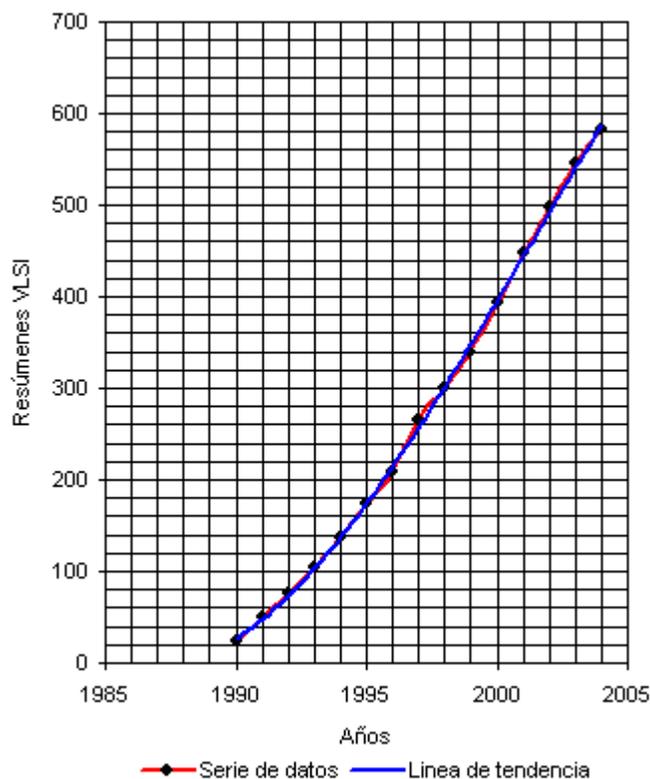


Figura 4.9. Distribución acumulada (curva color rojo) y aproximación de la curva mediante un polinomio de tercer orden (curva color azul) para VLSI y Tolerancia a fallos.

La gráfica anterior se considera como un histograma de valores, que tiene una distribución acumulada en forma de curva J, y que en algún momento alcanza su límite superior y se convierte en una curva S; esto sucede cuando el desarrollo de VLSI y tolerancia a fallos, haya iniciado la etapa de decadencia.

Para la serie de datos (línea roja) de la figura 4.9, la curva que mas se ajusta es el polinomio de la ecuación (4.38), el cual se encuentra limitado entre dos rectas horizontales $y_{\min(VLSI-Fault)} = 10.8943$ y $y_{\max(VLSI-Fault)} = 968.4412$ que pasan por sus puntos mínimo $x_{\min(VLSI-Fault)} = -3.1003$ y máximo $x_{\max(VLSI-Fault)} = 26.1944$; como muestra la figura 4.10.

$$P_3(x)_{VLSI-Fault} = -0.075289x^3 + 2.6081x^2 + 18.3430x + 40.4510 \quad (4.38)$$

El punto de inflexión $x_{\text{inf}(VLSI-Fault)} = 11.5470$, representa el mes de junio del 2001 y se localiza en la parte central de la curva S. En esta fecha se tiene el máximo desarrollo de VLSI y tolerancia a fallos. Para fechas posteriores al punto de inflexión; el interés disminuye gradualmente hasta el punto $x_{\text{max}(VLSI-Fault)}$ cuando el tema se considera obsoleto y corresponde al mes de febrero del 2016. El punto $x_{\text{min}(VLSI-Fault)} = -3.1003$ indica el nacimiento de VLSI y tolerancia a fallos de manera conjunta, y corresponde al mes de enero de 1987. Revisando [41] y [42], se puede ver que los primeros sistemas VLSI que incorporaron tolerancia a fallos, surgen a finales de la década de los 80's, lo cual concuerda con nuestro análisis.

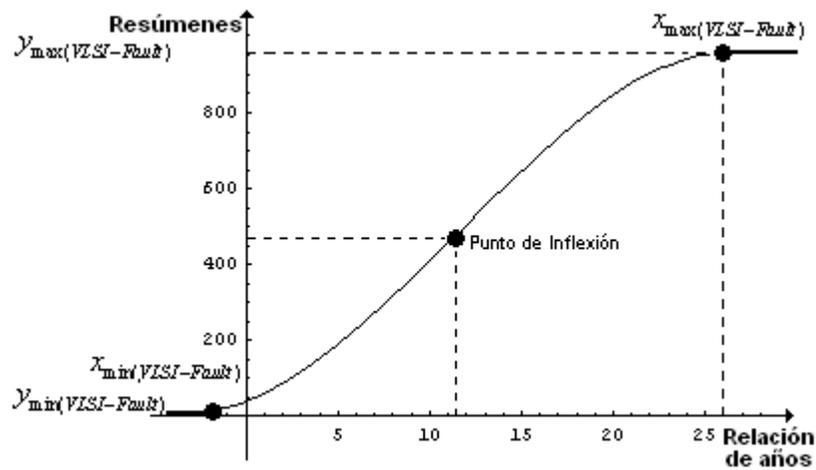


Figura 4.10. Aproximación de la curva S de VLSI y tolerancia a fallos con el polinomio cúbico de la ecuación 4.38.

Siguiendo el desarrollo de la metodología de la sección 4.1.5 se obtiene la ecuación para la función logística:

$$l_{(VLSI-Fault)} = \frac{968.4528}{1+e^{-0.2001(x-11.5470)}} \quad (4.39)$$

donde $\beta_{(VLSI-Fault)} = 0.2001$, $\gamma_{(VLSI-Fault)} = 11.5470$ y $\alpha_{\text{max}(VLSI-Fault)} = 968.4528$. La figura 4.11 muestra los detalles de la ecuación (4.39). La asíntota superior de la curva logística se encuentra en $\alpha_{\text{max}(VLSI-Fault)} = 968.4528$, en tanto que la asíntota inferior en $\alpha_{\text{min}(VLSI-Fault)} = 10.8827$.

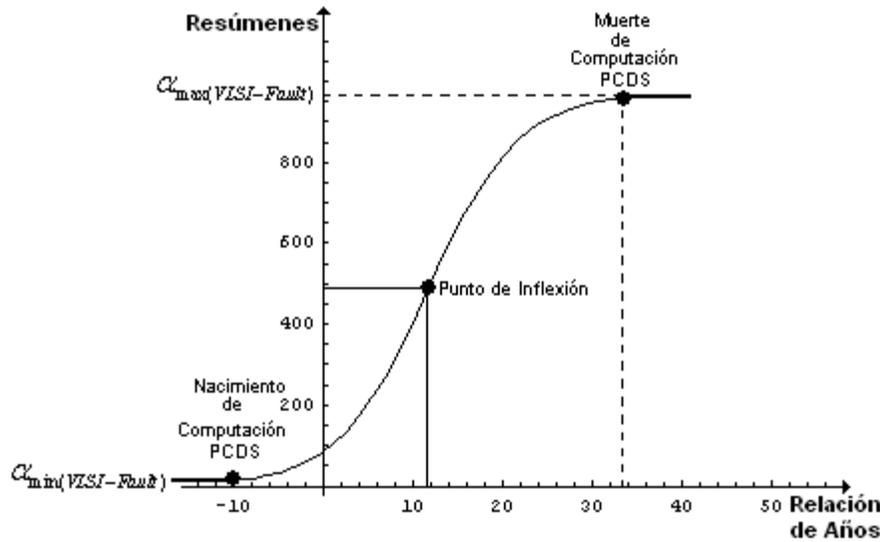


Figura 4.11. Aproximación de la curva S mediante la función logística para VLSI y Tolerancia a fallos.

La superposición de ambas curvas se muestra en la figura 4.12. Ambas curvas son simétricas con relación al punto de inflexión. La curva punteada corresponde al polinomio cúbico mientras que la curva continua corresponde a la función logística. El desarrollo de ésta última es más lento que la curva cúbica que tiene un desarrollo más vertiginoso.

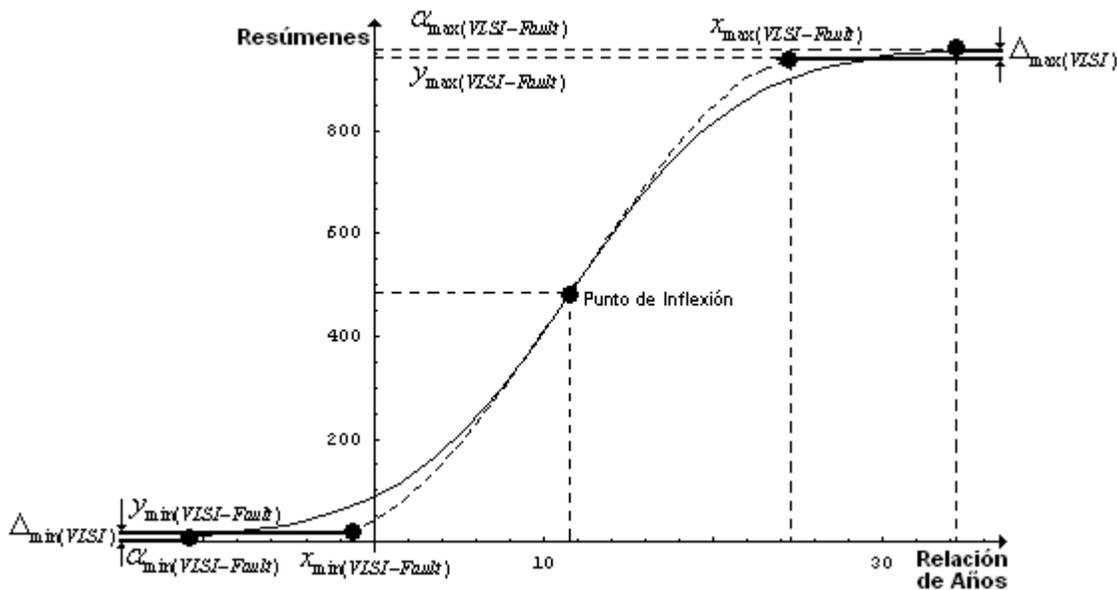


Figura 4.12. Superposición de las curvas logística y cúbica para VLSI y Tolerancia a fallos.

La discrepancia entre ambas curvas es de 0.0116, lo que implica un error del 1.16%. Por lo tanto, se puede decir que este porcentaje de error es bastante aceptable, y que cualquiera de los modelos utilizados, proporcionan resultados confiables para investigar la tendencia de desarrollo de VLSI y tolerancia a fallos.

4.2. Nuevo enfoque de la ley de Moore.

En 1975 Moore cambió los míticos 12 meses por 24 (ninguna de las dos cifras explican el número de transistores que hoy tiene un microprocesador), y a partir de este año ha sido corregida y distorsionada de forma constante. En la década de los 80's se actualiza para cambiar los 24 meses de 1975 por 18. A principios de los 90's cambió de nuevo para centrarse únicamente en el número de transistores que podían integrarse en un chip, una cifra que actualmente se dobla cada 18 meses. Algunos estudiosos del fenómeno, como Ilkka Tuomi, van más allá y consideran que la ley de Moore “jamás ha predicho con exactitud ningún desarrollo en el campo de la microelectrónica”, y que se trata en realidad de un mito y una atribución sin sentido, dado que varios científicos de la época conocían también este crecimiento exponencial, y predijeron con mayor exactitud los futuros niveles de integración. Más que una ley, se debería hablar de una tendencia [50] [URL-15].

Existen distintos enfoques [51] [52][URL-16] para explicar y representar la ley de Moore, la más popular es la cantidad de transistores que pueden empotrarse en un chip. Algunos autores prefieren la **frecuencia de reloj**, otros la **disipación de potencia**, el **costo de fabricación**, unos mas optan por la **tendencia de cambio**, etc. Esta situación conectada con los puntos débiles de cada enfoque, hace que no exista una posición estándar para formular la denominada ley. Por ejemplo, el enfoque más popular: la **duplicación de transistores**, es criticada por que no es una determinación, ya que se ha cambiado varias veces el período de tiempo (desde 12 hasta 24 meses). Al mismo tiempo, el punto débil más importante es que la cantidad de transistores no es una característica universal, ya que el aumento de éstos en un chip, implica disminución en su tamaño y a la vez presenta una tasa de fallo muy elevada, que los hace poco apropiados para aplicaciones que requieren alta fiabilidad; una de las razones principales que obligó utilizar sistemas tolerantes a fallos dentro de los sistemas VLSI.

Como medida para luchar con el problema de duplicación y reservación de esquema, se han investigado formas de integración que permitan mantener el ritmo esperado, y empotrar cada vez más transistores en el mismo o en menor espacio de silicio. Una de ellas es la **tecnología doble núcleo** (double core) [53] [URL-12], que se ha convertido rápidamente en parte central para los principales productores de chips, en un esfuerzo por mejorar el desempeño de los procesadores. Por ejemplo si la cantidad de transistores aumenta dos órdenes de integración, estos procesos de duplicación disminuyen la cifra de elevación efectiva de dos órdenes a una,

implicando una disminución en la cantidad de transistores y reduciendo la posibilidad de utilizar sistema tolerante a fallos. Lo anterior se observa en la figura 4.13.

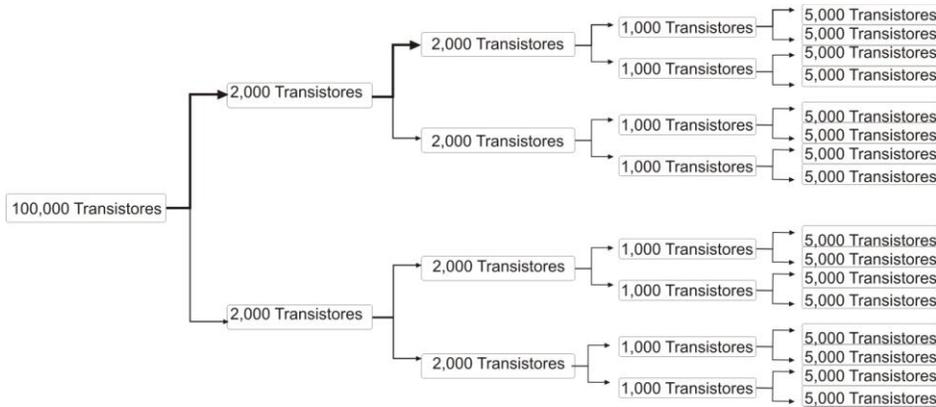


Figura 4.13. Esquema de la tecnología de doble núcleo.

Otra característica que cambia con la tendencia de duplicación de la cantidad de transistores en un chip, es la **disminución del tamaño del transistor**. Esta propiedad ha evolucionado en base al tamaño de la terminal de control del transistor, como se observa en la tabla 4.1.

| Dispositivo | Fecha | $-\text{Log}(\text{tamaño})$ | Dispositivo | Fecha | $-\text{Log}(\text{tamaño})$ |
|-----------------|-------|------------------------------|--------------------------|-------|------------------------------|
| Efecto Edison | 1879 | -8.23044892 | Procesador 80286 | 1982 | -3.17609126 |
| Bulbo | 1904 | -7.954242 | Procesador 80386 DX | 1985 | -3.17609126 |
| Transistor | 1947 | -7.69897 | Procesador 80486 DX | 1989 | -3 |
| C.I | 1958 | -6.74036269 | Procesador Pentium | 1993 | -2.90308999 |
| C.I | 1960 | -4.47712125 | Procesador Pentium Pro | 1995 | -2.77815125 |
| C.I | 1970 | -4 | Procesador Pentium II | 1997 | -2.54406804 |
| Procesador 4004 | 1971 | -4 | Procesador Pentium III | 1999 | -2.39794001 |
| Procesador 8008 | 1972 | -4 | Procesador Pentium IV | 2000 | -2.25527251 |
| Procesador 8080 | 1974 | -3.77815125 | Procesador Pentium IV HT | 2004 | -2.11394335 |
| Procesador 8085 | 1976 | -3.47712125 | Procesador Pentium IV HT | 2005 | -1.95424251 |
| Procesador 8086 | 1978 | -3.47712125 | Procesador Pentium D950 | 2006 | -1.81291336 |
| Procesador 8088 | 1979 | -3.47712125 | | | |

Tabla 4.1. Evolución de los dispositivos electrónicos basado en el tamaño de su terminal de control.

La tabla comienza con el efecto Edison, considerado momento cero o punto de partida de la era electrónica. Este dispositivo nace en 1879 y mide alrededor de 17 cm (del filamento a la placa metálica interna) [34]. Posteriormente se tiene el primer bulbo cuyo tamaño inicial es de 9 cm entre ánodo y cátodo. Siguiendo con la evolución de los dispositivos electrónicos, se alcanza el desarrollo del Procesador Pentium D950 con tecnología de 65 nm en el año 2006 [URL-12] [URL-17]; para el 2007 y 2009 las tecnologías de 45 nm y 32 nm respectivamente [54]; en tanto que para el 2011 los microprocesadores Ivy Bridge e Intel Atom SoC de 22 nm [URL-20].

La distribución de la tabla 4.1 se preparó para **menos logaritmo del tamaño de la terminal de control de los dispositivos electrónicos**, como muestra la ecuación (4.40).

$$-\log(\text{tamaño de la terminal de control}) \quad (4.40)$$

El signo menos se aplica para comodidad de observar un progreso con tendencia creciente (curva S como modelo). El **logaritmo del tamaño** es necesario para disminuir el rango de escala que se tiene desde el efecto Edison (17 cm), hasta los microprocesadores Ivy Bridge e Intel Atom SoC de 22 nm. Con esta ecuación se obtiene la curva mostrada en la figura 4.14. Esta curva conforma un simple eslabón fractalizado de una conformación similar en cadena. Cada una de las curvas S individuales, va cediendo paso a otra curva S inmediatamente superior. Entre los eslabones (curvas S individuales) en lugar de los tradicionales elementos de continuidad, aparecerán rupturas (discontinuidades y situaciones caóticas). Así, los cambios de paradigma representan momentos revolucionarios menores, dentro de una cadena mayor de características evolutivas.

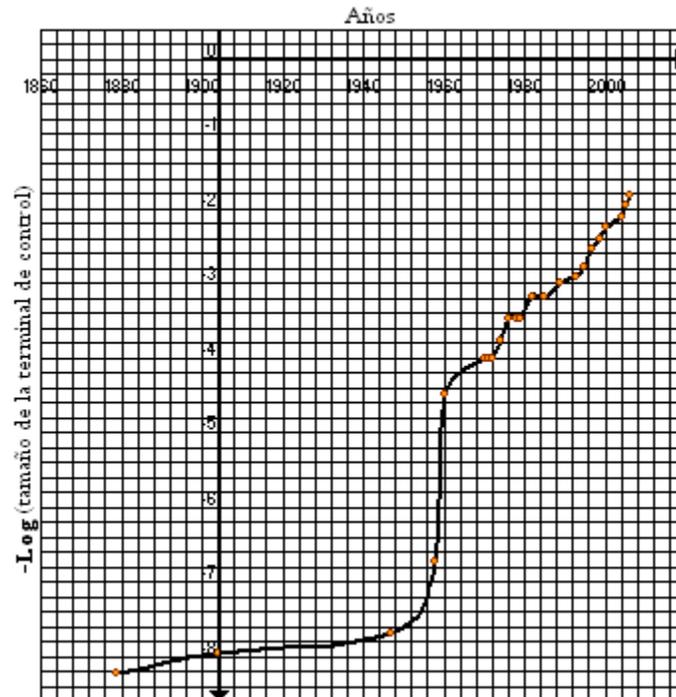


Figura 4.14. Conjunto de curvas S de la evolución de los dispositivos electrónicos.

No es difícil explicar esta estructura de curvas S en escalera; cada nuevo escalón (subcurvas S) enlaza el cambio de tecnología de fabricación del transistor. Para conocer la tendencia de evolución del tamaño del transistor, se traza la tangente a las múltiples curvas S. Por su

generalización, no existe un método para construir la curva tangente de manera similar al método de mínimos cuadrados. Para este análisis se propone dos métodos los cuales se enuncian a continuación:

1. selección de los puntos que tienen mayor grado de concavidad,
2. selección de puntos mediante un método recurrente.

Para el primer método no es necesario considerar el conjunto de curvas S mostradas en la figura 4.14; basta con tomar en cuenta los puntos de las curvas que tienen mayor grado de concavidad. La tabla 4.2 resume estos puntos, y en la figura 4.15 se muestra la gráfica.

| Año | -Log(tamaño) | Año | -Log(tamaño) |
|------|--------------|------|--------------|
| 1960 | -4.47712125 | 1982 | -3.17609126 |
| 1976 | -3.47712125 | 2006 | -1.81291336 |

Tabla 4.2. Puntos seleccionados que presentan el mayor grado de concavidad.

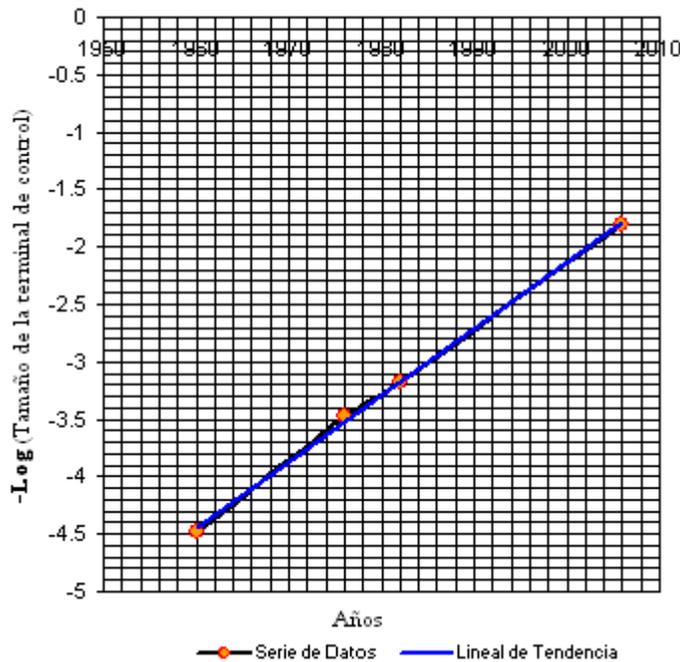


Figura 4.15. Puntos con mayor grado de concavidad de la figura 4.14.

Siguiendo la metodología utilizada en el capítulo anterior, se tiene que el polinomio de la ecuación (4.41) es la curva que más se ajusta a los puntos de la tabla 4.2.

$$y = -0.0000137x^3 + 0.000592x^2 + 0.05722x - 2.2563 \quad (4.41)$$

En esta ecuación se observa que los dos primeros miembros (parte no lineal) son despreciables comparados con los dos últimos (parte lineal); esto permite utilizar la parte lineal del polinomio cúbico para aproximar los puntos.

La recta tangente se traza desde el punto que tiene coordenadas: (**año de invención, tamaño del efecto Edison**); punto A de la figura 4.16. Ahí la tangente es horizontal y continúa por los puntos que se muestran en la figura 4.15, hasta el punto E que tiene por coordenadas (**año de fabricación del procesador D950, tamaño de la puerta de los transistores en el procesador Pentium D950**). Prolongando esta recta después del punto E, se observa que intercepta la recta horizontal $x = 0.3010$ en el punto G. Este punto indica el final de la electrónica de silicio y corresponde al año 2043. La recta horizontal muestra el límite mínimo que puede tener un dispositivo, que conserva la información en base al cambio de energía del electrón. Esta distancia es la diferencia entre la banda de conducción y la banda de valencia, y tiene un valor aproximado de 5 \AA ; después de este tamaño no es posible proponer un dispositivo mas pequeño para **electrónica**.

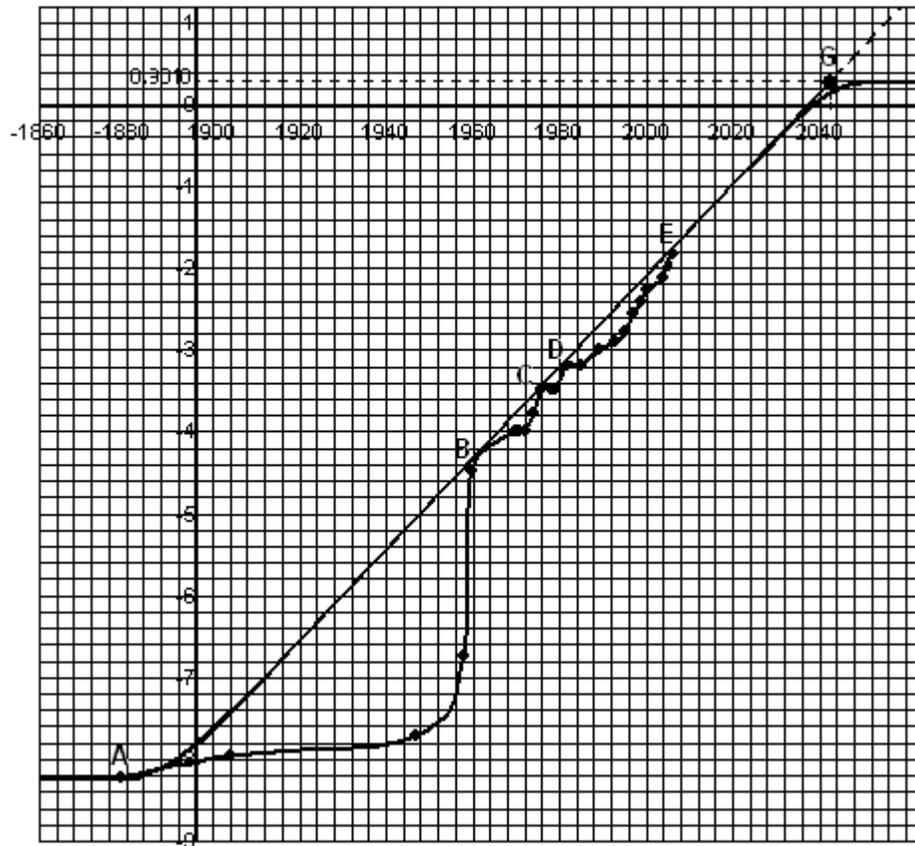


Figura 4.16. Evolución del tamaño de los dispositivos electrónicos. Conjunto de curvas S y recta tangente despreciando los dos primeros términos del polinomio cúbico.

La discrepancia que se obtiene al utilizar la parte lineal del polinomio es menor del 3% para los primeros 10 años mientras que para 20 años se tiene un error de aplicación del 11.44%.

El método anterior proporciona una aproximación debido a que los puntos tangenciales se eligieron basados en la inspección ocular de la figura 4.14. El segundo método basado en la aproximación de la curva tangente mediante parábolas, permite calcular los puntos exactos mediante una ecuación de recurrencia y evitar arbitrariedades.

La figura 4.17 muestra el método para hallar la ecuación de recurrencia. El punto x_0 , se fija como punto inicial o punto de previa iteración que corresponde al año de inicio de la era electrónica. En este punto $F(x_0) = S_0 = -8.23044892$ y $F'(x_0) = S'_0 = 0$. Para las x_i restantes se fijan las condiciones $F_i(x_i) = S_i(x_i)$ y $F'_i(x_i) = S'_i(x_i)$, donde la tangente F , la curva S_i y sus derivadas tienen el mismo valor en el punto x_{i+1} pero la abscisa es desconocida.

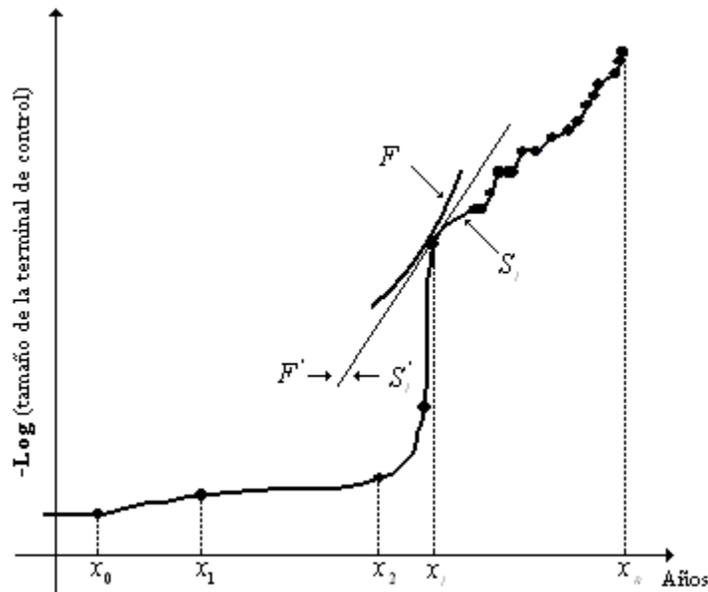


Figura 4.17. Metodología para hallar los puntos tangentes con la ecuación de recurrencia.

Considerando que se conoce: x_1 , $F(x_1)$ y $F'(x_1)$. Con $F_1(x_1) = S_1(x_1)$ y $F'_1(x_1) = S'_1(x_1)$.

Se obtienen cuatro ecuaciones algebraicas con cuatro incógnitas (coeficientes de parábola A, B, C y punto x_{i+1}), al mismo tiempo.

$$A_1 x_0^2 + B_1 x_0 + C_1 = S_0 \quad (4.42)$$

$$A_1 x_1^2 + B_1 x_1 + C_1 = a_1 x_1^2 + b_1 x_1 + c_1 \quad (4.43)$$

Derivando las ecuaciones (4.42) y (4.43) se tiene

$$2A_1x_0 + B_1 = S_0' \quad (4.44)$$

$$2A_1x_1 + B_1 = 2a_1x_1 + b_1 \quad (4.45)$$

Restando a la ecuación (4.45), la ecuación (4.44) se tiene que:

$$A_1 = \frac{a_1x_1 + \frac{1}{2}(b_1 - S_0')}{(x_1 - x_0)} \quad (4.46)$$

De la ecuación (4.44) se despeja B_1

$$B_1 = S_0' - 2A_1x_0 \quad (4.47)$$

A la ecuación (4.43) se le resta la ecuación (4.42)

$$A_1(x_1^2 - x_0^2) + B_1(x_1 - x_0) = a_1x_1^2 + b_1x_1 + c_1 - S_0 \quad (4.48)$$

sustituyendo la ecuación (4.46) y la ecuación (4.47) en la ecuación (4.48) se tiene que

$$x_1 = -\frac{2(c_1 - s_0) + x_0(b_1 + s_0')}{2a_1x_0 + (b_1 - s_0')} \quad (4.49)$$

Para el caso general:

$$x_{i+1} = -\frac{2(c_{i+1} - s_i) + x_i(b_{i+1} + s_i')}{2a_{i+1}x_i + (b_{i+1} - s_i')} \quad (4.50)$$

En la figura 4.18 se muestra los resultados de este análisis y al igual que la curva de la figura 4.16 no tiene restricciones en la parte superior derecha.

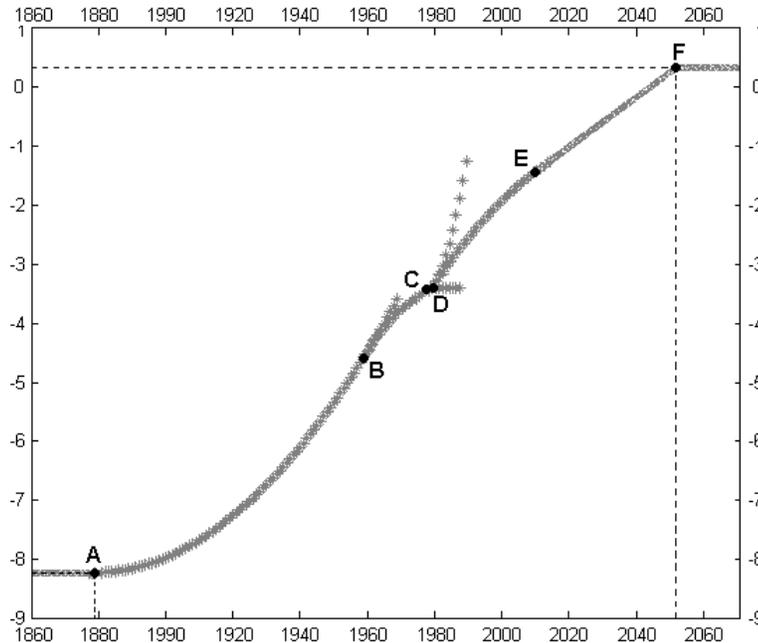


Figura 4.18. Conjunto de curvas S aproximadas mediante parábolas.

Este límite se alcanza en el punto F, donde la tangente encuentra la recta horizontal $x = 0.3010$ y corresponde al año 2051. El punto F indica el fin del desarrollo de la electrónica como ciencia límite que emplea el flujo de electrones para transmisión, almacenamiento y tratamiento de información.

4.2.1. Conclusiones parciales.

Los modelos matemáticos aplicados, revelan que la etapa de máximo desarrollo para computación PCDS ha finalizado, por lo que ha dado comienzo la etapa de madurez. Según la trayectoria del modelo de curva S, esta etapa requiere de mayor esfuerzo económico e intelectual para obtener avances significativos. Lo anterior implica que el ciclo de vida para computación PCDS se acerca a su límite, el cual tendrá lugar a final de la tercera década entre el período 2019 y 2027.

Los temas VLSI y tolerancia a fallos han tenido un importante desarrollo dentro de computación PCDS, sobre todo a mediados del año 2001 cuando alcanzan su máximo desarrollo. Este desarrollo se ha dado de manera recíproca, ya que como ley causa-efecto, los sistemas VLSI utilizan ideas de paralelismo para solucionar los conflictos generados por el aumento en la escala de integración; y a su vez, computación PCDS utiliza los avances tecnológicos obtenidos en VLSI para lograr nuevos retos científicos. Los modelos matemáticos empleados en la sección

4.1.7.2 indican que la etapa de mayor interés para estos temas ha culminado y ha iniciado la etapa de decadencia la cual finalizará aproximadamente en el año 2016.

La metodología original para la construcción de la tangente al conjunto de curvas-S mediante parábolas, ha proporcionado información para conocer la tendencia de los dispositivos electrónicos. Lo que indica el final de la era de los dispositivos semiconductores tal como se observó anteriormente en la sección 4.2. Por lo que se concluye que el ciclo de vida para la electrónica de silicio, ha entrado en la etapa de madurez, y que para el período que comprende del 2043 al 2051 habrá llegando a su fin para dar paso a nuevas tecnologías alternativas, tales como: electrónica cuántica, molecular o fotónica; y de esta forma iniciar un nuevo escalón (ciencia revolucionaria) en este campo.

Capítulo 5

Conclusiones y líneas futuras.

El trabajo se orienta a la creación de una metodología que permite el análisis del desarrollo de una rama específica del conocimiento. Para demostrar su utilidad se aplicó a un cuerpo de documentos sobre computación paralela, concurrente, distribuida y simultánea. Gracias al análisis resultante es posible detectar tendencias, observar cambios y predecir futuras líneas de oportunidad. La trascendencia del tema abordado por éste trabajo es de gran importancia en nuestro tiempo, donde el conocimiento es indispensable para el desarrollo de una sociedad. Por otro lado, conocer las líneas futuras de oportunidad es de gran relevancia, sobre todo en un país con fuertes limitantes económicas en el campo de la investigación científica como es el caso de México.

5.1. Conclusiones.

En este último capítulo se exponen las conclusiones obtenidas de la investigación, para ello se retoman conceptos ya explicados en los capítulos anteriores.

1. Para evaluar la etapa de desarrollo de este dominio, se coleccionaron 2300 resúmenes de la librería digital de IEEE (cantidad total de resúmenes en la base de datos de IEEE), distribuidos durante un período de 15 años, estos resúmenes contienen un promedio de 200 palabras por lo que son considerados ‘objetos de conocimiento pobre’. Se agruparon por año de publicación para contrarrestar este problema y convertirlos en ‘objetos de conocimiento enriquecido’; el análisis de los datos agrupados da una visión más clara de la dinámica de desarrollo de los temas, es decir que la suma anual de textos se utilizó como un nuevo objeto de investigación.
2. El análisis de los datos numerosos facilita la posibilidad de descubrir el cambio de interés en las direcciones más importantes de computación PCDS; como pudo observarse durante el período 1997-1998, donde se encontró que dos clústeres cambian su valor y tienen una productividad casi constante antes y después de este período. El cluster 1 es el más importante, representa el cambio de época en el desarrollo de computación PCDS, como muestra el escalón en el período 1997-1998. Esta última observación permite bajo investigación subdividir el período en dos etapas: de 1990 a 1997 y de 1998 al 2004. La comparación de estos dos estados permite determinar las tendencias de desarrollo de la ciencia. Es más, un análisis similar para un período más grande, permite detectar un mayor número de etapas, y así, mas detalles.
3. Los resultados obtenidos proporcionan información importante, para que un «experto» en computación PCDS construya una ontología. Cuando no se cuenta con un «experto» o conjunto de expertos en esta área, entonces se muestra que es posible sustituir a éstos, por información contenida en libros, revistas y congresos. Por ejemplo, si se tienen palabras claves del DOD en algunos textos anuales y en algunos libros, el nombre de estos libros o sus capítulos, tienen el nombre del subtema que enlaza a este conjunto de textos anuales; y con esto se puede conocer el lapso de tiempo de desarrollo de este subtema (nombre del libro).
4. La idea de ontología basada en cadena (tema, concepto y palabra) es productiva, porque da la posibilidad de estructurar las palabras en el DOD de manera jerárquica, y al mismo tiempo proporciona claridad sobre el nivel de abstracción de cada palabra en el DOD. Realmente la ontología es una estructura jerárquica del DOD, que permite analizar el cambio temporal de ésta. De tal manera es posible ver el cambio de valor en la palabra, cuando ésta baja o sube de nivel jerárquico, o cuando las palabras de menor abstracción

toman posición en los conceptos o perdieron su valor; pasando de un nivel de conceptos, a un nivel de palabras. Lo cual muestra pérdida de interés en los subtemas. Con esto, no solamente se tiene un método para construcción de ontologías, sino también la posibilidad de investigar y medir la dinámica de desarrollo del tema, a través de la dinámica de la ontología, la cual es la diferencia entre las estructuras del DOD para los dos lapsos de tiempo.

5. Para comparar dos ontologías o sus estructuras se puede aprovechar el método de medición de distancias entre dos estructuras dendroides. Este algoritmo es similar al algoritmo conocido como edit tree algorithm (distancia de edición en árboles). La estructura jerárquica de los documentos permite determinar la relación padre-hijo entre palabras: por ejemplo, una palabra que aparece en el título de un artículo es candidato a ser padre de las palabras que aparecen en el cuerpo de éste; si tal relación se repite varias veces, se coloca el par **padre-hijo** en la ontología. Usando los artículos correspondientes se construye la ontología para cada período independientemente. Comparando las ontologías (usando distancia de edición en árboles) se encuentran las tendencias de evolución temporal para computación PCDS.

6. Los resultados experimentales muestran que la ontología construida es significativa. Específicamente, puede ser utilizada para el análisis comparativo (del estado y desarrollo) de una rama científica. Se puede aplicar a cualquier jerarquía de textos estructurados, por ejemplo, páginas web HTML. Sin embargo, el uso más probable del método es el PRO, para resultados de más alta calidad. Los conceptos más interesantes son: **analysis**, **network** y **vlsi**, donde la palabra clave es VLSI y los conceptos **fault** y **tolerance**. En los niveles más bajos de la ontología, estas palabras se encuentran casi siempre juntas. En el segundo nivel **fault** y **tolerance** forman el concepto **fault-tolerance**. Se puede decir que los conceptos **VLSI** y **fault-tolerance** se encuentran enlazados, debido a que VLSI está conectado directamente con el tema de hardware, mientras que **fault** y **tolerance** con los problemas más actuales de VLSI. Para **fault** y **tolerance** se tienen nueve conceptos para el primer período; veinticinco para el segundo y cuatro para ambos períodos. Estas cifras muestran la tendencia de **fault** y **tolerance** en la segunda etapa. Sumando los conceptos del primer, segundo y ambos períodos da un total de treinta y ocho conceptos. El núcleo de palabras que se encuentran en todo el período son solamente cuatro, lo que implica que el núcleo conserva solamente el 10.5% de información. La parte que no cambia demuestra la evidencia de que la dinámica en esta temática es muy alta. Considerando el mismo nivel para VLSI, en donde se tienen siete **conceptos** contruidos de ochenta y un **palabras** para el primer período; trece **conceptos** contruidos de ciento cuarenta y tres **palabras** para el segundo período, y veinticuatro **conceptos** en ambos períodos. Se tienen que el núcleo conserva 54.5% de información. Esta temática es aún más estable, ya que cuenta con un diccionario más desarrollado,

tanto en el primer como en el segundo período; en éste último se tiene el doble de **conceptos nuevos** que muestran la tendencia de VLSI. El análisis de la estructura y la dinámica de las ontologías para estos casos parciales, proporciona la dirección de mayor interés para éste período.

7. Como meta para la evaluación de la ontología obtenida, se utilizó una estructura elaborada de capítulos, secciones, subsecciones y partes pequeñas de libros de texto correspondientes a los temas (es decir, libros de texto con el título correspondiente a un concepto de la ontología). Se observa en el árbol que los conceptos de los nodos inferiores se parecen a los capítulos y a las secciones de los libros de texto. Esta es una prueba de la alta calidad de la ontología obtenida automáticamente. Por supuesto, al estar la ontología basada en libros de texto, no hace el trabajo inútil, por el hecho de proporcionar resultados ya conocidos. La ontología extraída de un libro de texto, se puede construir solamente en un pasado muy remoto, algo para el cual el libro de texto ya existe. Sin embargo, para las áreas de investigación activa, los libros de texto aún no existen y aparecerán solamente en el futuro. Realmente, la ontología refleja la estructura de un libro de textos a futuro, y puede ser utilizado para planear un nuevo libro de texto. La buena correspondencia entre la ontología y los libros de texto, indica que la ontología construida con el mismo método para un período reciente, es correcta. Sin embargo, una evaluación más rigurosa es un trabajo a futuro.
8. Los resultados experimentales demuestran que dentro de la estructura jerárquica de computación PCDS, los temas VLSI y tolerancia a fallos tienen un desarrollo significativo durante todo el período de investigación. Los clústeres relacionados con estos temas son numerosos en ambos períodos (1990-1997 y 1998-2004), y en la ontología se manifiesta una relación intrínseca entre VLSI, tolerancia a fallos y computación PCDS. Los modelos matemáticos aplicados, revelan que la etapa de mayor producción para computación PCDS ha finalizado, por lo que ha dado comienzo la etapa de madurez. Según los resultados del modelo de curva S, esta etapa requiere de mayor esfuerzo económico e intelectual para obtener avances significativos. Lo anterior implica que el ciclo de vida para computación PCDS se acerca a su límite, el cual tendrá lugar a fines del año 2020.
9. Los temas VLSI y tolerancia a fallos han tenido un importante progreso dentro de computación PCDS, sobre todo a mediados del año 2001 cuando alcanzan su máximo desarrollo. Este desarrollo se ha dado de manera recíproca, ya que como ley causa-efecto, los sistemas VLSI utilizan ideas de paralelismo para solucionar los conflictos generados por el aumento en la escala de integración; y a su vez, computación PCDS utiliza los avances tecnológicos obtenidos en VLSI para lograr nuevos retos científicos.

Los modelos matemáticos empleados indican que el ciclo de vida para VLSI y tolerancia a fallos concluye aproximadamente en el año 2016.

10. El pronóstico científico del desarrollo de la electrónica (basada en los electrones como los conservadores de información) se encontró como resultado de fundamentar en una nueva presentación física de la ley de Moore y de un análisis propio, más profundo. La metodología de construcción de ontologías, el descubrimiento dinámico del dominio científico y, la metodología de construcción de *curvas S* y *curva tangente*, fueron utilizadas para analizar el desarrollo de la electrónica desde su concepción (efecto Edison) hasta los actuales nanoprocesadores.
11. Existen al menos dos tipos de límites más, relacionados con el ulterior progreso de la electrónica de semiconductores: económicos y termodinámicos. Los límites económicos, se refiere al coste de una nueva fábrica. Desgraciadamente, la ley de Moore sobre cambio exponencial se aplica también a los costes de fabricación. En los últimos veinticinco años, el coste de una fábrica ha aumentado en más de una orden de magnitud. Lo importante de esto, es el enorme riesgo que una determinada empresa tiene que afrontar cuando decide construir una nueva fábrica. Este riesgo se está haciendo casi insoportable y lo único que va a hacer es empeorar. Para el siguiente límite se debe tomar en cuenta el principio de Landauer. Este principio puede llegar a ser un problema en el desarrollo de la microelectrónica y/o nanoelectrónica si no se le da la importancia que amerita. Según este principio, la temperatura es insignificante cuando la información cambia en un transistor, pero cuando se tienen millones de estos dispositivos (como es el caso de los últimos microprocesadores y nanoprocesadores) la temperatura comienza a ser un obstáculo en el desarrollo de los dispositivos electrónicos.
12. La metodología y herramienta es aplicable al sector gubernamental en el sentido de que los órganos encargados de regular y guiar el desarrollo de la ciencia y tecnología (SEP, CONACyT, Consejo Consultivo de Ciencias, Secretaria de Educación Pública, entre otros) puedan contar con una alternativa para evaluar, y fomentar/regular el desarrollo de ciertas áreas, o para encontrar rezagos y/o deficiencias.
13. Los resultados del apéndice D demuestran que el pronóstico conseguido con los 5039 resúmenes recopilados de la librería digital ACM Guide to Computing “siguen” los resultados presentados en el capítulo 4 utilizando los 2300 resúmenes de la librería digital de IEEE para el período 1998-2004. Se observa un final para los temas VLSI y tolerancia a fallos en el mes de febrero del 2016; mientras que el resultado utilizando la información de la librería digital ACM para el período 1998-2010 culmina en el mes de junio del 2014. Por lo que se puede concluir que independientemente de la base de datos que se considere, los resultados obtenidos son muy similares ya que en los últimos 6 años

los temas de computación PCDS, VLSI y tolerancia a fallos han tenido un comportamiento muy similar al pronosticado en el capítulo 4.

Continuando con este breve apartado de conclusiones, es necesario en este momento remontarnos a las motivaciones que originaron esta investigación como son las hipótesis y objetivos, y que están básicamente reflejadas en el capítulo 1.

Para la primera hipótesis se inició con el enriquecimiento de los textos con conocimiento pobre para luego obtener la lista de frecuencia de los textos científicos sobre computación PCDS; aprovechando la estructura jerárquica de los documentos científicos, se construyó el prototipo rápido de ontología y se estableció un nuevo punto de vista sobre ontología: no como una colección estática de hechos acerca del lenguaje y el mundo; sino como una «foto instantánea» del estado del lenguaje y del mundo en un momento específico; esto permitió darle una nueva aplicación a la ontología como una herramienta para describir y estudiar la evolución del dominio. Lo anterior valida la primera hipótesis.

Después de realizar un estudio sobre *la teoría para la resolución de problemas inventivos* se encontró que las únicas leyes de desarrollo técnico aplicable al progreso del hardware y software son: *ley dinámica de transformación del macronivel al micronivel* (tendencia de los dispositivos electrónicos, sección 2.3.2), y *ley de la transformación de cantidad en calidad* (ciclo de vida de un sistema, sección 2.3.4); y esto se debe a que la teoría para la resolución de problemas inventivos fue creada basada en principios de la mecánica clásica y se desarrolló considerando aspectos de procesos macrofísicos. Por lo que no es posible aplicarlos a procesos de micro y nanoelectrónica, debido a que éstos se encuentran regidos por leyes de la mecánica cuántica, por lo que la segunda hipótesis es válida.

La última hipótesis se validó en el capítulo 4, específicamente con las figuras 4.3, 4.4 y 4.5. La figura 4.3 muestra la curva acumulada o curva-J para los 2300 resúmenes sobre computación PCDS. Aplicando la ley de Verhulst y aproximando los datos mediante una curva de tendencia se obtiene las curvas S de la figuras 4.3 y 4.5.

En cuanto al cumplimiento de los objetivos se observa que los primeros dos objetivos específicos fueron cubiertos con los resultados del capítulo 3, específicamente con las secciones 3.2 y 3.3, donde se presenta la construcción del prototipo rápido de ontología y el análisis dinámico de la ontología. El resto de los objetivos específicos fueron cubiertos en el capítulo 4.

Para terminar esta sección se puede señalar a manera de conclusión general, que la metodología basada en el tamaño de la terminal de control de los dispositivos electrónicos (bulbo, transistor, mosfet etc.), ha proporcionado un enfoque para conocer la tendencia de los dispositivos electrónicos. Esta tendencia ha provocado que en la actualidad se hayan obtenidos circuitos

integrados con escalas de integración muy grandes, como es el caso de los micro y nanoprocesadores. El aumento en la escala de integración depende crucialmente de la capacidad para fabricar transistores cada vez más pequeños. Pero este proceso tienen un límite: la estructura atómica del semiconductor (silicio u otro material). De seguir la tendencia actual, tarde o temprano se tropezará con los 5 \AA que detendrá el proceso y obligará a una nueva y revolucionaria tecnología. Este tamaño equivale a la diferencia entre las bandas de conducción y valencia en el átomo del semiconductor (silicio), y es la distancia mínima que puede tener un dispositivo que conserve la información en base al cambio de energía del electrón. Después de este tamaño no es posible proponer un dispositivo más pequeño para **electrónica**. Esto implica el final de la era de los dispositivos semiconductores, tal como se observó en los modelos desarrollados y la aplicación de métodos para la construcción de las curvas tangentes. Por lo que se puede concluir que el ciclo de vida para la electrónica de silicio, se encuentra en la etapa de madurez, y que para el período que comprende del 2043 al 2051 habrá llegando a su fin, para dar paso a nuevas tecnologías alternativas, tales como: electrónica molecular o fotónica; y de esta forma iniciar un nuevo escalón (ciencia revolucionaria) en este campo. Lo eminente de este resultado es que el transistor es la parte esencial del hardware, y éste se encuentra vinculado de manera cooperativa con el software (es decir que los cambios en el hardware se presentan como retos para el desarrollo del software; al mismo tiempo las posibilidades teóricas y demandas prácticas del software, imponen nuevas demandas para el desarrollo del hardware), lo que implica que al haber cambios sustanciales en la electrónica; el hardware y software tendrán que evolucionar a nuevas arquitecturas.

5.2. Trabajo futuro.

Enseguida se proponen algunos temas relacionados con el presente trabajo y que pueden dar continuidad a esta investigación con el objetivo de ampliar el alcance de esta línea de investigación.

1. Existen muchas direcciones futuras de este trabajo, la mayoría han sido mencionadas a través del desarrollo del texto. Como trabajo futuro se planea desarrollar esta metodología, utilizando métodos de filtración y descubrimiento de palabras del diccionario orientado a dominio de manera más fina. Para esto se necesita aislar las palabras del supersistema HSW, y considerar únicamente palabras sobre computación PCDS; en este caso, muchas palabras del supersistema que contienen otras ramas de desarrollo del HSW son eliminadas del diccionario orientado a dominio.
2. Por otro lado, para revelar conceptos más precisos, es mejor preparar una investigación comparativa de las mismas ramas científicas, utilizando distintos idiomas; lo que

permitirá descubrir conceptos análogos y comparar la diferencia de los resultados parciales entre idiomas.

3. Realizar el análisis por lo menos cada seis años.
4. Se puede aprovechar la metodología desarrollada para construcción de ontologías en temas pequeños.
5. Estudiar e investigar más leyes físicas que restringen el desarrollo de la nanoelectrónica.
6. Planeación de libros de texto con ontologías.

Referencias

- [1] Salton G., "Automatic content analysis in information retrieval" (University of Pennsylvania, PA, 1968).
- [2] Noyons E.C.M. (1999) *Bibliometric Mapping as a Science Policy and Research Management Tool*. DSWO PRESS. Science Studies. Leiden University, The Netherlands.
- [3] Kuhn Thomas S. (1970). *The Structure of Scientific Revolutions*. The University of Chicago Press.
- [4] Makagonov P., Ruiz Figueroa A. (2004). *Study of Knowledge Evolution in Parallel Computing by Short Texts Analysis*. In: Progress in Pattern Recognition, Image Analysis and Applications, CIARP2004, Lecture Notes in Computer Science N3287, Springer.

- [5] Perelló Cabrera, J. L. (1986) *Modelación matemática en el estudio bibliométrico de la información especializada*. Actualidades de la Información Científica y Técnica La Habana) 1(126): 79-90,.
- [6] Gorbea-Portal, S. (1994). *Principios teóricos y metodológicos de los estudios métricos de la información*. Investigación Bibliotecológica 8(17): 23-32.
- [7] Hulme, E.W. *Statistical Bibliography in relation to the growth of modern civilization*.
- [8] Ranganathan, S.R. (1969) *Librarymetry and its scope in subject analysis for document of translation service*. Indian Statistical Institute (Bangalore, India). Documentation. Research and Training Centre (Hrsg.) Bangalore, Indian Statistical Institute, (DRIC Annual Seminars 7, vol.1: papers).
- [9] Morales-Morejón, M. y Cruz-Paz A. (1995) *La Bibliotecología, la Cienciología y la Ciencia de la Información y sus disciplinas instrumentales: Su alcance conceptual*. Ciencias de la Información (La Habana) 26(2): 70-88.
- [10] Bernal J. D. (1939) *The social function of science*. Londres, Routledge.
- [11] Golsmith, M. y MACKAY, A. L. (1964): *Tite Science of Science*. Londres. Sonvenir.
- [12] Price D. J. S.(1963). *Little Science, Big Science*. New York. Columbia University Press.
- [13] Dobrov G. M. (1966). *Nauka o Nauke*. Kiev, Naukova.
- [14] Otlet P. (1934). *Traité the Documentation*. Le livre sur livre Théorie et pratique. Bruselas, Mundaneum, 411 Pp.
- [15] Tushman, M. y CO'Reilly. (1999). *"Winning through innovation"*.
- [16] Makagonov, P., Ruiz Figueroa A. (2005). *A Method of Rapid Prototyping of Evolving Ontologies*. In: Lecture Notes in Computer Science 3406, Springer-Verlag.
- [17] Makagonov P., Alexandrov, M., Sboychakov, K. *A toolkit for development of the domain-oriented dictionaries for structuring document flows*. In: H.A. Kiers et al (Eds), "Data Analysis, Classification, and Related Methods", Springer, 2000 (Studies in classification, data analysis, and knowledge organization), pp. 83-88.
- [18] Hearst, M. (1992). *Automatic acquisition of hyponyms from large text corpora*. In: Proc. 14th International Conference on Computational Linguistics (COLING-92), p. 539-545.
- [19] Paşka, M. (2004). *Acquisition of Categorized Named Entities for Web Search*. In: Proc. 2004 ACM CIKM Intern. Conf. on Information and Knowledge Management. ACM.
- [20] Martins, A., H. S. Pinto, Oliveira A. L., (2004). *Towards Automatic Learning of a Structure Ontology for Technical Articles*. In: Proc. Workshop on the Semantic Web at SIGIR-04, U.K.

- [21] Gelbukh, A., Alexandrov M., S. Y. Han. (2004). *Detecting Inflection Patterns in Natural Language by Minimization of Morphological Model*. In: Lecture Notes in Computer Science 3287, Springer-Verlag, p. 432–438.
- [22] Makagonov, P., Ruiz Figueroa A. (2005). *A Method of Rapid Prototyping of Evolving Ontologies*. In: Lecture Notes in Computer Science 3406, Springer-Verlag.
- [23] Sboychakov K. IRBIS *Library Automated System*; Programmer of IRBIS System; Russian National Public Library for Science and Technology, Moscow.
- [24] Makagonov P, Ruiz Figueroa A., Sboychakov K., Gelbukh A. (2005). *Learning a Domain Ontology from Hierarchically Structured Texts*. Proc. of Workshop “Learning and Extending Lexical Ontologies by using Machine Learning Methods” at 22nd International Conference on Machine Learning, ICML, Bonn, Germany.
- [25] Wagner J. A., III and Gooding R. Z., (1987). *Effects of Societal Trends on Participation Research*, Administrative Science Quarterly, 32, pp. 241–262.
- [26] Taylor, R. : *The Informaíon Sciencess, Lihrary Joarnal* (88), 1963 19, p. 4161.
- [27] Borko H. (1968) *Information Science. What is it?*», American Documentation (19), 1.
- [28] Terrada M. L. y López Piñeiro. *Historia del concepto de Documentación*», Teorema, 4, p. 238.
- [29] Terrada M. L.: *La documentación médica como disciplina*. Valencia, Centro de Documentación e Informática Biomédica. Universidad de Valencia, 1983, p. 127.
- [30] Petrov V. *Laws of Dialectics in Technology Evolution*. Israel.
- [31] Foster R. *Innovation: The attackers Advantage* (Hardcover). ISBN: 0671622501.
- [32] Robert L. Boylestad and Louis Nashels. *Electronic Devices and Circuit Theory*. By Prentice Hall, Inc. Sixth Edition, pp 114-115.
- [33] John Bardeen et al. *Three-Electrode Circuit Element Utilizing Semiconductive Materials*. Patent issued to Bell Labs in 1950. U. S. Patent 2,524,035.
- [34] Parra L. *Televisión a Color Moderna*. Centro Japonés de Información Electrónica S. A de C. V. ISBN 968-7356-86-3.
- [35] Kilby J. *Miniaturized Electronic Circuits*. Patent issued to Texas Instruments in 1964. U. S. Patent 3,138,743.

- [36] William F. Brinkman, Douglas E. Haggan, and William W. Troutman. *A History of the Invention of the Transistor and Where It Will Lead Us*. IEEE Journal of Solid-State Circuits. Vol. 32, No. 12, December 1997.
- [37] Ramirez Hassan A. (2003). *Condición Marshall-Lerner y efecto curva J: una aproximación al caso colombiano*. Medellín, Colombia.
- [38] Graham M., (1939). *The sigmoid curve and the overfishing problem*. Rapp.P.-V.Ré. CIEM, 110:15–20.
- [39] McLellan P. *Effective data management for VLSI design*. Proceedings of the 22nd ACM/IEEE conference on Design automation (DAC '85). Publication: June 1985. pp. 652-657.
- [40] Hiroshi A. (1985). *Automatic routing algorithm for VLSI*. Proceedings of the 22nd ACM/IEEE conference on Design automation (DAC '85). pp. 785-788.
- [41] Koren I., Adit D. (1990). *Singh Fault Tolerance in VLSI Circuits*. Computer. pp. 73-83.
- [42] Singh. A. D. (1988). *An Area Efficient Fault Tolerance Scheme for Large Area VLSI Processor Arrays*. IEEE Transactions on Computers. pp. 1398-1410.
- [43] Duran R. (2003) *Internet Electronic Journal Nanociencia et Moletrónica NANOMÁQUINAS Y NANOROBOTS*. Vol. 1; N°2, págs. 112-119.
- [44] RUA N. A. La domesticación de la tecnología. Monografias.com.
- [45] Rengel Estévez R. (2002). *Análisis de los fenómenos de transporte y ruido electrónico en transistores MOSFET y SOI submicrométricos*. Salamanca.
- [46] Wilson A. H. (1931). Proc. Royal Soc. 133A 458.
- [47] Bardeen J. Y Brattain W. H. (1948). *"The transistor, a semiconductor triode"*. Phys. Rev 74230.
- [48] Shockley W., (1951) *"Circuit element utilizing semiconductive material"*, US Patent 2 569 347. Application filed June 26, granted Sep 25.
- [49] Gordon E. Moore. *Cramming more components onto integrated circuits*. Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.
- [50] Ilkka Tuomi. (2002). *The Lives and Death Moore's Law. First Monday*. Vol. 7, number 11.
- [51] Alamo J. *Microelectrónica de Si*. De la era de las PC a la era de Internet. Departamento de Ingeniería Electrónica y Ciencias de la Computación. Massachusetts Institute of Technology.

- [52] García L. F., Aguirre M. *Curso de Doctorado. Microelectrónica Digital*. Departamento de Ingeniería Electrónica.
- [53] *Intel Developer Forum (IDF) 2005 Spring* celebrated of 3 to 5 March in San Francisco, Cal.
- [54] Koji Shiro. (2002). *Nanotechnology Fueling Moore's Law*. Technology & Manufacturing Group-Japan. Copyright Intel Corporation.
- [55] Callon, M, Courtial, Pierre J. y Penan, Hervé (1993): "*La Scientométrie*", *Que sais-je?*, Numero 2727, Presses Universitaires de France, Paris.
- [56] Proyecto CONACYT N-39011-A. *Construcción de Diccionarios Orientados a Dominio*.
- [57] Petrov V.M. & Zlotina E.S. (1990). *Laws of System Evolution*, -Leningrad.
- [58] C H Bennet and Landauer R., (1985) '*The Fundamental Physical Limits of Computation*', Scientific American, pp. 38-46.
- [59] Nikil Dutt. (2005). *Introduction to Embedded and Ubiquitous Systems*. ICS 212 Fall.
- [60] Ruiz Figueroa A, Makagonov (2006). *Modelos de desarrollo del Hardware y Software basado en el estudio de Computación Paralela*. Revista Interciencia, Caracas Venezuela.
- [61] Camps D. (2008) *Limitaciones de los indicadores bibliométricos en la evaluación de la actividad científica biomédica*. Corporación Editora Medica del Valle. Colombia Médica. Vol. 39 N° 1.
- [62] Camps D, Recuero Y, Avila R, Samar M. (2006). *Estudio bibliométrico de un volumen de la revista Archivos de Medicina*. (sede Web). Oviedo, España. Arch Med 2: 3. Disponible en: <http://www.archivosdemedicina.com/ojs/index.php/archmed/article/view/0203001/0>
- [63] Pérez-Matos N. (2006). *La bibliografía, bibliometría y las cienciasafines*. (sede Web). Ciudad de La Habana, Cuba. ACIMED. 2002; 10: 1-2. Disponible en: http://scielo.sld.cu/scielo.php?pid=S1024-94352002000300001&script=sci_arttext.
- [64] Bordons M, Zulueta M. (1999) *Evaluación de la actividad científica a través de indicadores bibliométricos*. Rev Esp Cardiol.; 52: 790-800.
- [65] Rousseau R. (2006). *Indicadores bibliométricos y econométricos en la evaluación de instituciones científicas*. (sede Web). Ciudad de La Habana, Cuba. ACIMED. 2001; 9: 23-29. Disponible en: http://scielo.sld.cu/scielo.php?pid=S1024-94352001000400008&script=sci_arttext
- [66] Spinak E. (1998). *Indicadores cientiométricos*. Ci Inf.; 27: 141-8.
- [67] Institute for Scientific Information. (2006). *The ISI impact factor*. (sede Web) Philadelphia, USA. Disponible en: www.isinet.com

- [68] Leon-Sarmiento FE, León ME, Contreras VA. *El factor de impacto: ¿Mito o realidad?* Colomb Med. 2007; 38: 290-6.
- [69] Leon-Sarmiento F, Bayona-Prieto J, León ME. *Concepciones, confusiones y contraindicaciones del factor de impacto en Colombia*. Rev Esp Salud Publica. 2007; 81: 147-54.
- [70] Gruber, T. R. A (1993). *Translation Approach to Portable Ontology Specifications*. Knowledge Acquisition, 5 199-200.
- [71] Bateman J., Kasper T. Moore J. y Whitney R. (1990). "A General Organization of Knowledge for Natural Language Processing: The Penman Upper Model". Technical Report, USC/ISI. Marina del REY, CA (USA).
- [72] Bernaras A., Laresgoti I. y Corera J. "Building and Reusing Ontologies for electrical Network Applications". European Conference of Artificial Intelligence (ECAI'96) Editado por Wiley and Sons Lid. Págs. 298
- [73] Beners-Lee T., Hendler J. y Lassila O. (2001). "The Semantic Web". *Scientific American*. Vol. 284, no. 5. Págs. 34-43.
- [74] Borst W. (1997). "Construction of Engineering Ontologies". Tesis Doctoral, university of Twente. Enschede.
- [75] Farquhar A., Fikes R. y Rice J. (1996). "The Ontolingua Sever: A Tool for Collaborative Ontology Construction". Proceedings of the 10 Knowledge Acquisition for Knowledge-based Systems Workshop, vol. 19 Págs. 1-44, Banff. Alberta. Canada.
- [76] Fernández López M. (1999) "Overview of Methodologies for Building Ontologies". Proceedings of the IJCAI '99, Workshop on Ontologies and PSMs. Stockholm. Sweden.
- [77] Geneseth M. y Fikes R. (1992). "Knowledge Interchange Format, version 3.0, Reference Manual". Technical Report. Logic-92. Computer Science Department, Stanford University. California. USA.
- [78] Gómez-Pérez A. (1998). "Knowledge Sharing and Reuse". The Handbook of Applied Expert Systems. Editado por J. Liebowitz, CRC Press. Londres. Reino Unido.
- [79] Gómez-Pérez A. y Benjamins R. (1999) "Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods". Proceedings of the IJCAI '99. Workshop on Ontologies and PSMs. Stockholm. Sweden.
- [80] Gómez-Pérez A. y Benjamins R. (1999) "Applications of Ontologies and Problem-solving Methods. AI Magazine vol. 20, no. 1, págs. 119-122.
- [81] Gruber T. (1993). "A Traslation Approach to Portable Ontology Specifications". Knowledge Acquisition, vol. 5, no. 2, págs. 199-220.

- [82] Gruber T. (1995) “*Toward Principles for the Design of Ontologies Used for Knowledge Sharing*”. International Journal of Human-Computer Studies, 43. Págs. 907-928.
- [83] Guarino N. y Giarretta P. (1995). “*Ontologies and Knowledge Bases: Towards a Terminological Clarification*” Toward Very Large Knowledge Bases. Knowledge Building & Knowledge Sharing. Editado por IOS Press. Págs. 25-32.
- [84] Guarino N. (1998). “*Some Ontological Principles for Design Upper Lexical Resources*”. Proceedings of the 1 International Conference on Language Resources and Evaluation. Págs. 527-534. Granada. España.
- [85] Guarino N. y Welty C. (2000). “*A Formal Ontology of Properties*”, Proceedings of the ECATOO Workshop on Applications of Ontologies and PSMs. Berlin.
- [86] Knight K. Luk S. (1994). “*Building a Large Knowledge Base for Machine Translation*”. American Association for Artificial Intelligence (AAAI). Págs. 773-778. Seattle. USA.
- [87] Lenat D. y Guha R.V. (1990). “*Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*”. Editado por Addison-Wesley Publishing Company, Inc. California. USA.
- [88] Mahesh K. (1996). “*Ontology Development for Machine Translation: Ideology and Methodology*”. Technical Report. Computer Research Laboratory. New Mexico State University.
- [89] Mizoguchi R. (1993). “*Knowledge Acquisition and Ontology*”. Proceedings of the KB&KS '93. Págs. 121-128. Tokyo.
- [90] Mizoguchi R. Vanwelkenhuysen J. e Ikeda M. (1995). “*Toward Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing*”. IOS Press. Págs. 46-59.
- [91] Neches R., Fikes R., Finin T., Gruber T., Patil R., Senator T. y Swartout W. (1991). “*Enabling Technology for Knowledge Sharing*”. AI Magazine. Págs. 36-56. Otoño.
- [92] Niles I. y Pease A. (2001). “*Origin of the IEEE Standard Upper Ontology*”, Proceedings of the IJCAI '01, Workshop on the IEEE Standard Upper Ontology Seattle. Washintong. USA.
- [93] Studer S., Benjamins R. y Fensel D. (1998). “*Knowledge Engineering: Principales and Methods*”. Data and Knowledge Engineering, 25, 161-197.
- [94] Swartout B., Patil R., Knight K. y Russ T. (1997). “*towards Distributed Use of Large-Scale Ontologies*” AAAI'97 Spring Symposium Series on Ontological Engineering. Págs. 138-148.
- [95] Uschold M. y Grüninger M. (1996). “*Ontologies : Principles, Methods and Applications*”. Knowledge Engineering Review, Vol. 2 Pags. 93-155.

- [96] Van Heist G., Schreiber A. y Wielinga B. (1997). *“Using Explicit Ontologies in KBS Development”*. International Journal of human-computer Studies, vol.45 Págs. 183-292.
- [97] Bollinger T. y Pfleeger S. (1990). *“The economics of Reuse: Issues and Alternatives”*. Proceedings of the 8th Annual National Conference on Ada Technology. Págs. 436-447. Atlanta.
- [98] Nacional Institute of Standards and Technology (1994). *“Glosary of Software Reuse Terms”*. Diciembre.
- [99] Poulin J.S. (1997). *“Measuring Software Reuse: Principles, Practices, and Economic Models”*. Addison Wesley Longman Inc.
- [100] Real academia Española. (1999). *“Diccionario de la Lengua Española”*. 21^a edición. Espasa Calpe, Madrid.
- [101] del Río Pérez J. (2007). *Revisión de las Investigaciones Psicológicas sobre Creatividad Observatorio (OBS*)*. Journal, 2 253-302 Universidad de Navarra, España.
- [102] La historia de los procesadores desde la ENIAC hasta la Nehalem. Intel Corporation.
- [103] Castro de D., et al., (2004). *Automatic Web News Extraction Using Tree Edit Distance*, Department Federal University of Minas Gerais, Belo Horizonte, Brasil.
- [104] Godinez-Fernández, E.; Martínez-Luna, G.L.; Guzmán-Arenas, A.; (2010).  2-s2.0-78650602 *SATC tool for analysis of the evolution of scientific knowledge*. Pp 42-46. In: ICCTD 2010 2nd International Conference on Computer Technology and Development Proceedings. ISBN: 978-1-4244-8844-5. Cairo, Egipto.
- [105] Kermanidis K; (2009). *The effect of borderline examples on language learning*. Pp. 19-42. In: Journal of Experimental & Theoretical Artificial Intelligence. ISSN: 0952813X Volume 21 Issue 1, March 2009. Department of Informatics, Ionian University, Corfu, Greece.
- [106] Kermanidis K; Fakotakis, N; (2007). *One-sided Sampling for Learning Taxonomic Relations in the Modern Greek Economic Domain* Pp 354 - 361 in: Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference. ISBN: 978-0-7695-3015-4. Volume: 2. Artificial Intelligence Group, Department of Electrical and Computer Engineering, University of Patras, 26500 Rio, Greece.
- [107] Vaquero A.; Saenz F.; Alvarez F.; De Buenaga M.; (2006). *Conceptual design for domain and task specific ontology-based linguistic resources*. Pp. 855-862. In: On the Move to Meaningful Internet Systems 2006 CoopIS DOA GADA and ODBASE Proceedings Part I Montpellier France (2006). ISBN: 3540482873; 978-354048287-1. Volume 4275 LNCS - I, 2006. Berlin, Germany.
- [108] Vaquero A.; Saenz F.; Alvarez F.; De Buenaga M.; (2006) *Methodologically Designing a Hierarchically Organized Concept--Based Terminology Database to Improve Access to Biomedical*

Documentation. Pp. 658-668 in: On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. ISSN: 03029743. Volume 4277 LNCS - I, 2006, Berlin, Germany.

[109] Vaquero A.; Saenz F.; Alvarez F.; De Buenaga M.; (2006). *Thinking Precedes Action: Using Software Engineering for the Development of a Terminology Database to Improve Access to Biomedical Documentation*. Pp. 207-218. In: 7th International Symposium on Biological and Medical Data Analysis, ISBMDA 2006. ISBN: 3540680632; 978-354068063-5 Volume 4345 LNCS, 2006, Berlin, Germany.

[110] Sheremetov, L.; Batyrshin, I.; Filatov, D.; Martínez-Muñoz, J.; (2006). An uncertainty model for diagnostic expert system based on fuzzy algebras of strict monotonic operations. Pp. 165-175. In: 5th Mexican International Conference on Artificial Intelligence, MICAI 2006. ISBN: 3540490264; 978-354049026-5. Volume 4293 LNAI, 2006.

Páginas de Internet.

[URL-1] Bibliometría y cienciometría. Septiembre del 2006.
<http://egcti.upr.edu/Observa/html/procedimientos.htm>.

[URL-2] Cienciometria. 4 de Octubre del 2006.
http://148.216.10.83/VIGILANCIA/capitulo_4.htm.

[URL-3] Araujó Ruiz J.C., Arencibia R. Bibliometría y cienciometría. Diciembre del 2005.
http://bvs.sld.cu/revistas/aci/vol10_4_02/aci040402.htm.

[URL-4] Acevedo Pineda E.B. Lo que la cienciometría no alcanza a medir. Abril del 2005.
<http://www.campus-oei.org/salactsi/elsa6.htm>.

[URL-5] Marzo del 2004. <http://www.acm.org>. Abstracts

[URL-6] Abstracts. Abril del 2004. <http://www.computer.org/publications/dlib>.

[URL-7] Abstracts. Abril del 2004. <http://www.springer.com>.

[URL-8] Abstracts. Junio del 2004. <http://citeseer.nj.nec.com>.

[URL-9] Abstracts. Junio 2004. <http://researchindex.org>.

[URL-10] 100 años de radio. Enero 2006. <http://perso.wanadoo.es/electronics/radio/100radio2.html>.

[URL-11] Basil M. Al Hadithi. Tendencias en la diseño Electrónico. Septiembre 2005.
http://www.uax.es/publicaciones/archivos/TECELS04_001.pdf.

[URL-12] Microprocessor. Septiembre 2006.
<http://www.intel.com/pressroom/kits/quickreffam.htm>.

- [URL-13] Haemig. Leyes de la ecología poblacional. Febrero 2006.
<http://www.ecologia.info/leyes-1.htm>.
- [URL-14] La Cray en investigaciones gubernamentales y militares. Abril 2004.
<http://www.cita.es/textos/Cray-militares>.
- [URL-15] The lives and death of Moore's law. Octubre del 2006.
http://firstmonday.org/issues/issue7_11/tuomi/index.html.
- [URL-16] Los grandes desafíos en tecnologías de la información. Julio 2006.
<http://www.idg.es/computerworld/conocimiento/pdfs/mooreenfoques>.
- [URL-17] Dual core technology. Agosto 2006. <http://blogdrake.net/node/3712>.
- [URL-18] La investigación en el campo de la documentación e investigación científica.
<http://www.geocities.com/zaguan2000/512.html>.
- [URL-19] Jiménez H., Espínola D., Abascal R., Gómez R. Creación de una ontología para el dominio cultural. <http://www.cua.uam.mx/dccd/cc/memorias/innova/HJS.pdf>.
- [URL-20] Rod Cabrera. La primera demostración de transistores 3-D de 22 nm
<http://www.spk.la/2011/05/nuevos-transistores-3d-de-intel-de-22-nanometros/>

Apéndice A

Algoritmo: conjunto de directrices e instrucciones, que debidamente completadas nos llevan al resultado final correcto de un problema. Su nombre deriva del nombre del matemático Mohammed Ibn-Musa Al Khwarizmi, quien formó parte de la corte real en Bagdad –780-850–.

Advanced Micro Devices: empresa fundada en 1969, se introdujo en el mercado en 1975 al fabricar memorias RAM, actualmente es uno de los mayores fabricantes de microprocesadores, pasando por chipsets y placas base.

Análisis de información: la aplicación de técnicas de procesamiento automático del lenguaje natural. Esta definición corresponde al análisis asistido por computadora. En general, por análisis de la información se entiende la fase de interpretación que el usuario realiza de una

manera directa y manual. Los límites de este tipo de análisis son evidentes desde el momento que se trabaja sobre una cantidad importante de datos, y se trata además de incorporar el análisis en un sistema de producción de información elaborada o especializada.

Árbol: En general, un árbol es un grafo no dirigido, conexo, y sin lazos ni ciclos.

Artículos científico: es un informe escrito que comunica por primera vez los resultados de una investigación. Los artículos científicos publicados en miles de revistas científicas componen la literatura primaria de la ciencia. Los libros y los artículos de síntesis –review articles– resumen el conocimiento de un tema, componen la literatura secundaria. El artículo científico tiene seis secciones principales:

- Resumen o abstract: resume el contenido del artículo.
- Introducción: provee un trasfondo del tema e informa el propósito del trabajo.
- Materiales y Métodos: explica cómo se hizo la investigación.
- Resultados: presenta los datos experimentales.
- Discusión: explica los resultados y los compara con el conocimiento previo del tema.
- Literatura Citada: enumera las referencias citadas en el texto.

Athlon: Microprocesador fabricado por AMD, fue el primero en dar mayor velocidad de reloj, aumentándolo hasta 200 MHz, fue el primero en llegar a 1 GHz.

Base de datos: Conjunto de información guardada de forma correlativa y organizada de la cual posteriormente se puede extraer uno o los datos necesarios.

Bibliometría: es la ciencia que estudia la naturaleza y el curso de una disciplina mediante la estadística y el análisis de las diversas facetas de la comunicación escrita

Block de notas: es un editor de texto simple incluido en los sistemas operativos de Microsoft desde 1985. Su funcionalidad es muy sencilla, con capacidades de formato –negrita, cursiva– propias de procesadores de textos, o facilidades de programación –edición hexadecimal, reemplazado de texto–.

Bulbo o trífodo: es un componente electrónico basado en la propiedad que tienen los metales calientes de liberar electrones desde su superficie –efecto Edison–.

Búsqueda en anchura: se usa para aquellos algoritmos en donde resulta crítico elegir el mejor camino posible en cada momento como sucede en Dijkstra. A diferencia con la búsqueda por profundidad, ahora se visitan todos los vecinos de un vértice antes de pasar al siguiente. Por tanto no hay necesidad de retroceder. Una vez etiquetados todos los vecinos de un vértice X, se continúa con el primer vértice alcanzado después de X en la búsqueda.

Búsqueda en profundidad: se usa cuando queremos probar si una solución entre varias, cumple con ciertos requisitos como sucede en el problema del camino que debe recorrer un caballo para pasar por las 64 casillas del tablero. Se comienza en cualquier vértice y en cada paso se avanza a un nuevo vértice adyacente siempre que se pueda. Cuando todos los adyacentes a X hayan sido visitados, se retrocede al vértice desde el que se alcanzó X y se prosigue. Así se consigue etiquetar –visitar– todos los vértices de la componente conexa en que se encuentre el vértice inicial.

Byte: Unidad de almacenamiento en una computadora; se compone de 8 bits, el tamaño que ocupa es equivalente a un carácter o símbolo.

| | |
|-----------------|------------|
| 1 KiloByte (KB) | 1024 bytes |
| 1 Megabyte (MB) | 1024 KB |
| 1 GigaByte (GB) | 1024 MB |
| 1 TeraByte (TB) | 1024 GB |

Celeron: Procesador del fabricante Intel, de bajo costo, se diferencia con los otros por tener menos caché de nivel 2.

Centrino: Tecnología de Intel basada en los procesadores Pentium M, integrado en ordenadores portátiles, destacan por su bajo consumo de energía, haciendo más aprovechable la propia batería del portátil, y su integración de soporte para tecnología inalámbrica WiFi.

Ciclo de vida: Todo proyecto de ingeniería tiene unos fines ligados a la obtención de un producto, proceso o servicio, el cual es necesario generar a través de diversas actividades. Algunas de estas actividades pueden agruparse en fases porque globalmente contribuyen a obtener un producto intermedio, necesario para continuar hacia el producto final y facilitar la gestión del proyecto. Al conjunto de las fases empleadas se denomina «ciclo de vida». Sin embargo, la forma de agrupar las actividades, los objetivos de cada fase, los tipos de productos intermedios que se generan, etc. pueden ser muy diferentes dependiendo del tipo de producto o proceso a generar y de las tecnologías empleadas.

Ciencia de la ciencia: ver **cienciometría**.

Ciencia normal: significa investigación basada firmemente en una o más realizaciones científicas pasadas, realizaciones que alguna comunidad científica particular reconoce durante cierto tiempo, como fundamento para su práctica posterior.

Ciencia revolucionaria: Si el número o la importancia de problemas no pueden ser resueltos dentro de un paradigma, puede sobrevenir una crisis y cuestionarse la validez del paradigma. Entonces la ciencia pasa al estado de ciencia extraordinaria o ciencia revolucionaria en el que los científicos ensayan teorías nuevas.

Cienciometría: estudio cuantitativo de la ciencia que tiene sus inicios en lo que Price denomina «ciencia de la ciencia» cuyo núcleo central se encuentra en la idea de aplicar los recursos de la propia ciencia a si misma. En otros términos, se intenta aplicar los mismos métodos que las diferentes disciplinas ocupan en sus diversos objetos de estudio, pero tomando ahora como objeto de estudio a la misma ciencia. De alguna manera, intenta identificar las leyes que rigen la actividad científica.

Circuito integrado: pequeño circuito electrónico utilizado para realizar una función electrónica específica. Se combina por lo general con otros componentes para formar un sistema más complejo y se fabrica mediante la difusión de impurezas en silicio monocristalino, que sirve como material semiconductor, o mediante la soldadura del silicio con un haz de flujo de electrones.

Círculos científicos: Grupo de científicos que comparten una misma idea.

Clústeres de palabras: representa un pequeño grupo de palabras "asociadas" o sea un pequeño grupo de palabras coocurrentes en los contextos elementales del cuerpo de textos.

Computación paralela: La computación paralela se orienta a resolver rápidamente una tarea empleando múltiples procesadores simultáneamente. Esta práctica se vuelve popular a partir de fines de los 80's. La posibilidad de construir hardware paralelo de alto rendimiento no fue suficiente, el real desafío fue el software. Los ambientes de programación paralela fueron difíciles de usar y más, todavía, estaban atados a arquitecturas particulares. Solo se desarrolló esta práctica en problemas científicos y de ingeniería; el mercado era muy restringido. Sin embargo la perspectiva de la programación paralela se ha tornado mas atractiva con las redes de PCs o estaciones de trabajo llamados clusters. Es la solución de un problema usando múltiples computadoras o computadoras con multiples procesadores. Es un término genérico para una variedad de arquitecturas, incluyendo multiprocesamiento simétrico (SMP Symmetric MultiProcessing), agrupamiento de sistemas SMP y procesadores paralelos masivos (MPP).

Computación concurrente: implica "pseudoparalelismo". Computación altamente paralela: idem computación paralela, excepto que tiene un gran número de procesadores (100-1000). Computación de alto rendimiento o supercomputación: velocidad de computación. Redes computacionales: conexiones flojamente acopladas. Computación descentralizada: no hay un control central. Computación colaborativa.

Computación distribuida: Un sistema distribuido es una colección de computadoras autónomas que están conectadas unas con otras y cooperan compartiendo recursos (por ejemplo: impresoras y bases de datos). Son usados en aplicaciones comerciales y procesamiento de datos. Van desde pequeñas configuraciones cliente-servidor sobre redes en varias escalas hasta Internet. Los sistemas distribuidos no son vendidos como tal, pero crecen natural e incrementalmente. Las computadoras individuales son interconectadas por redes locales y éstas pueden estar interconectadas en una red amplia en la medida que la necesidad lo amerita. El crecimiento de la computación distribuida ha creado una serie de nuevos problemas que están en investigación. Los problemas elementales incluyen la ausencia de reloj común y la posibilidad de fallas de transmisión.

Computación PCDS: es el grupo formado por computación paralela, concurrente, distribuida y simultánea.

Coocurrencia: se dice cuando dos palabras van juntas en el título, en el abstract, por ejemplo, o en el número de descriptores o de identificadores. Si aparecen juntas en muchas ocasiones, significa que existe una relación entre estas palabras.

CPU o Unidad de Proceso Central: es un chip que sirve como cerebro a la computadora. En el interior de este componente electrónico existen millones de transistores.

Enfoque cualitativo: se destacan las cualidades del objeto de interés o de investigación; se señalan las cualidades de los componentes del proceso de investigación o las virtudes individuales de los sujetos; e incluso, se pone más interés en los procesos que en los propios resultados

Enfoque cuantitativo: se privilegia la experimentación con control de variables; se sustenta en la estadística y en la filosofía positiva; y el investigador permanece lejano al contexto del problema.

Curva J: gráfica con forma parecida a la letra J, y que representa un crecimiento de tipo exponencial.

Curva logística o función logística: modela la función sigmoidea de crecimiento de un conjunto P. El crecimiento inicial es aproximadamente exponencial; al cabo de un tiempo, aparece la competición entre algunos miembros de P por algún recurso crítico K –"cuello de botella"– y la tasa de crecimiento disminuye; finalmente, en la madurez, el crecimiento se detiene.

Curva S: nivelación de una gráfica de crecimiento exponencial, curva en J, cuando la población en rápido crecimiento rebasa la capacidad sustentadora de su medio ambiente y cesa de aumentar. Describe un aspecto universal del ciclo de vida de muchas dimensiones del mundo natural e intelectual. Este ciclo de fases puede ser aplicado a: temporadas del año, ideas, productos, la vida de una persona, generaciones, datos demográficos, tendencias sociales, y muchas otras áreas.

Derivadas parciales: En matemática, una derivada parcial de una función de diversas variables es su derivada respecto a una de esas variables en tanto las otras se mantienen constantes.

Dispositivos semiconductores: son dispositivos electrónicos contruidos de material semiconductor.

Diccionario Orientado a Dominio: representa un tema específico y es construido en base al análisis de una colección de documentos que representa este tema, el cual es seleccionado por un grupo de expertos. El criterio principal consiste en tomar las frecuencias del cuerpo de textos (resúmenes), se forma un diccionario de todas las palabras del cuerpo de textos y se compara con la frecuencia de la lista de frecuencias del lenguaje común. Si es mayor que tres veces la frecuencia de la lista de palabras comunes, pasa al diccionario.

Efecto Edison: el flujo de electrones entre dos elementos en un tubo al vacío.

Electrónica de estado sólido: la electrónica basada en el transistor.

Electrónica de vacío: es la electrónica basada en bulbos.

Electrónica molecular: a veces llamada moletrónica, es la rama de la ciencia que estudia el uso de moléculas orgánicas en la electrónica. Los primeros trabajos acerca de la transferencia de cargas eléctricas entre moléculas fueron realizados por Robert Mulliken y Albert Szent-Gyorgi en 1940. Sin embargo, el primer dispositivo moletrónico fue desarrollado hasta 1974, año en el que se construye un biestable de melanina. En 1988 fue descrita una molécula capaz de actuar como un transistor de efecto de campo. Los recientes avances en nanotecnología y el descubrimiento de polímeros conductores y semiconductores, han permitido avances espectaculares en la materia. Dispositivos como los diodos orgánicos emisores de luz –OLED por sus siglas en inglés–, transistores orgánicos de efecto campo –OFET por sus siglas en inglés– o paneles solares orgánicos son ya conocidos en la industria.

Escala de integración: La historia de la microelectrónica va marcada por la sucesión de diversas tecnologías según el número de componentes integrados en el mismo circuito. A comienzos de los 60, la tecnología SSI –Small Scale of Integration– permitía la integración de hasta 100

transistores por circuito integrado. A mediados de los 60 aparece la tecnología MSI –Medium Scale of Integration– capaz de integrar varios cientos de transistores en un chip. En la década de los 70 aparece la tecnología LSI –Large Scale of Integration–, aplicada fundamentalmente a circuitos digitales, con la que se integran varios miles de transistores. Durante esta década los microprocesadores fueron evolucionando desde los primitivos microprocesadores de 4 bits como el 4004 de Intel en 1971, hasta los microprocesadores de 16 bits, al final de la década. Este hecho extendió de forma importante el campo de aplicación de la electrónica hacia las aplicaciones de computación, que a su vez permitieron fabricar nuevos circuitos más complejos. Durante la década de los 80's se puede hablar ya de la tecnología VLSI –Very Large Scale of Integration– para designar a los circuitos arriba de 10 mil dispositivos en un chip. Estas tecnologías, con resoluciones de línea del orden de las 5 μ m, se dedicaron a la realización casi exclusiva de microprocesadores y memorias.

Exclusión Mutua: es la comunicación requerida entre dos o más procesos que se están ejecutando en paralelo y que necesitan a la vez el uso de un recurso no compartible. Consiste en asignar el recurso no compartible a uno de los procesos, mientras que los otros deben permanecer a la espera hasta que finalice la utilización de dicho recurso por el proceso al que se le asignó. Cuando este proceso termine, el recurso será asignado a uno de los procesos en espera. Se asegura el correcto uso del recurso.

FET: es un dispositivo semiconductor que controla un flujo de corriente por un canal semiconductor, aplicando un campo eléctrico perpendicular a la trayectoria de la corriente. Está compuesto de una parte de silicio tipo N, a la cual se le adicionan dos regiones con impurezas tipo P llamadas compuerta (gate) y que están unidas entre si. Las otras dos terminales se llaman drenador (drain), fuente (surce). La región que existe entre el drenador y la fuente, y que es el camino obligado de los electrones se llama "canal". La corriente circula de drenaje a Fuente.

Flujo de resúmenes: análisis basado en resúmenes.

Fotolitografía: el proceso que utiliza técnicas fotográficas para transferir las máscaras de los circuitos integrados a las obleas de silicio, y técnicas de grabado por ataque químico.

Fotónica: refleja el importante vínculo entre la óptica aplicada y la electrónica; forjado por el creciente papel que los materiales y dispositivos semiconductores juegan en los sistemas fotónicos. El campo de la óptica es amplio y continúa manteniendo alto potencial de explotación. Así como la electrónica involucra el control de flujo de carga eléctrica en el vacío o en la materia, la fotónica involucra el control de fotones en el espacio libre o en la materia.

GaAs(Arseniuro de Galio): material semiconductor que presenta una movilidad electrónica superior a la del silicio y por tanto permite la fabricación de circuitos más rápidos. Sin embargo su materia primas es más cara que el silicio y de uso menor extendido, lo que implica un proceso tecnológico mucho más caro.

Germanio: es un elemento químico con número atómico 32, y símbolo Ge, perteneciente al grupo 14 de la tabla periódica de los elementos. Es un metaloide sólido duro, cristalino, de color blanco grisáceo lustro, que conserva el brillo a temperaturas ordinarias. Presenta la misma estructura cristalina que el diamante y resiste a los ácidos y álcalis. Forma gran número de

compuestos órganos metálicos y es un importante material semiconductor utilizado en transistores y fotodetectores. A diferencia de la mayoría de semiconductores, el germanio tiene una pequeña banda prohibida –band gap– por lo que responde de forma eficaz a la radiación infrarroja y puede usarse en amplificadores de baja intensidad

Google: es un motor de búsqueda a gran escala. Su principal objetivo es proveer información de alta calidad a pesar del rápido crecimiento de la World Wide Web, para ello incorpora una serie de técnicas para mejorar la calidad de la búsqueda, incluyendo el Page Rank y la proximidad de la información. Google es una arquitectura completa para capturar páginas web, indexarlas e interpretar las consultas. El análisis de la estructura de enlaces vía Page Rank, permite a Google evaluar la calidad de las páginas web y proporcionar mayor relevancia a sus respuestas.

Grafo dirigido acíclico: es un conjunto finito no vacío de nodos y un conjunto de pares de nodos ordenados, llamados arcos directos -sin ciclos-.

Granularidad: se basa en el tamaño de las unidades léxicas en las que se puede segmentar un texto. Los textos de granularidad más gruesa son los que contienen combinaciones mayores de palabras: fórmulas, frases hechas, nombres propios, términos compuestos. La densidad de un texto es inversamente proporcional a la granularidad. A mayor densidad, menor granularidad.

Hertz o Hercio: Unidad de frecuencia que indica la cantidad de ciclos por segundo.

Indexación semántica latente: el análisis de la semántica latente –LSA– nace como evolución de una técnica de recuperación de información bibliográfica denominada «indexación semántica latente» -LSI-. Esta técnica permite recuperar títulos de documentos relacionados con un patrón de búsqueda aunque estos títulos no incluyan explícitamente ese patrón. El sistema realiza inferencias basadas en los términos que los títulos comparten entre sí. LSA se basa en este enfoque, y permite representar el significado de términos, y compararlos entre sí, basándose en la coocurrencia de los términos en los distintos párrafos de un documento.

Internet: es un sistema mundial de computadoras conectados en red creado por ARPA(Advanced Research Projects Agency) del gobierno de EE.UU. Primeramente se llamó ARPANET y su principal objetivo fue la intercomunicación entre universidades y centros de desarrollo tecnológico, pudiendo transmitirse la información en más de una dirección, así los datos de una de las partes no se perdían por ataque militar u otros desastres. En la actualidad Internet es una cooperativa pública y autosuficiente a la cual cualquier persona se puede conectar.

Itanium: primer microprocesador basado en la IA-64, arquitectura de 64 bits. Fue la principal motivación para crear nuevo software basado en 64 bits y para su incorporación en servidores y estaciones de trabajo.

Jerarquía: es la disposición de documentos en orden ascendente o descendente, según criterios de clase o cualquier otro asunto que conduzca a un sistema de clasificación.

Journals: características de journals académicos:

1. se enfocan en un tema específico y una audiencia especializada.

Ejemplos:

- a. Journal of Property Finance.

- b. Journal of the American Medical Association.
 - c. Comparative Education.
 - d. Political Science Quarterly.
2. Los artículos están escritos en un lenguaje académico por uno o más autores con mucho conocimiento y experiencia en la disciplina.
 3. Los artículos tienden a ser largos y tienen cierta estructura. Muchas veces incluyen un abstract –resumen– al inicio y deben contener bibliografía de obras consultadas.
 4. Normalmente los journals no son ilustrados, aunque pueden contener tablas o gráficas.
 5. Los journals en su formato impreso tienen un aspecto serio y sencillo.
 6. Puede ser que los journals sean arbitrados. Esto significa que un comité de expertos en la disciplina han leído el artículo y determinado que tiene valor y es digno de ser publicado. Los Journals arbitrados son considerados de muy buena reputación. En inglés, se dice "refereed" o "peer-reviewed journals" para indicar a este tipo de publicación.

Large Scale Integration: ver escala de integración.

Ley causa-efecto: proclama que todo sucede por alguna razón, se conozca ésta o no. No existen hechos accidentales. Vivimos en un universo ordenado, regido estrictamente por leyes; esta idea es crucial para la comprensión del funcionamiento conjunto de todas las leyes o principios.

Ley de Bradford: plantea que si un número determinado de publicaciones periódicas se ordenan en forma decreciente, según su productividad de artículos y se dividen en tres zonas de análisis, cada una con la misma cantidad de contribuciones, entonces existirá una zona núcleo con un reducido número de fuentes donde se concentrará la mayor cantidad de artículos y dos zonas restantes con una mayor dispersión.

Ley de Lotka: describe una relación cuantitativa entre los autores y los artículos producidos en un tema, durante un período de tiempo. Por ejemplo en una determinada especialidad pocos autores publican un gran número de artículos, mientras que el gran número de autores restante publica muy poco. Es habitual que el 50% de los trabajos sobre una especialidad sean efectuados, aproximadamente, por sólo el 10% de los autores totales, un 25% por un 15% de los autores, y el restante 25% de los artículos corresponde al 75% de los autores.

Ley de Moore: La evolución tecnológica de los circuitos integrados está resumida en la denominada ley de Moore. Esta ley es un razonamiento formulado el 19 de abril de 1965 por el Dr. Gordon E. Moore, en su artículo titulado "Cramming more components onto Integrated Circuits" publicado en el 35 aniversario de la revista Electronics la cual enuncia que: el número de elementos que la industria sería capaz de colocar en un chip se duplicaría cada 12 meses.

Ley Zipf: enuncia que las palabras que aparecen en un texto lo suficientemente largo se cuentan y clasifican por la frecuencia de aparición.

Leyes cuánticas: son las que rigen el movimiento de sistemas en los cuales los efectos cuánticos son relevantes. Se ha documentado que tales efectos son importantes en materiales mesoscópicos (unos 1000 átomos).

Ley de Verhulst: establece que en algún momento la tasa de crecimiento per capita de una población esta limitada directa e inmediatamente por su propia densidad, a través del proceso de

competencia intra específica. En otras palabras ningún sistema con crecimiento exponencial puede seguir indefinidamente.

Maquina de John von Neumann: Utiliza un procesador y memoria donde se guardan datos y programa, es decir, una máquina secuencial que procesa datos escalares. La idea central del modelo consiste en almacenar las instrucciones del programa en su propia memoria, logrando así que la máquina siga los pasos definidos por su programa almacenado.

Matriz «palabra / palabra»: es una matriz construida con la cantidad de textos que contiene los pares de palabras de columna y fila.

Matriz «texto / palabra»: es una matriz construida con la cantidad de palabras de DOD (columna) en cada texto (fila).

Matriz «texto / texto»: es una matriz construida con la cantidad de las mismas palabras de DOD en cada par de textos.

Medium Scale Integration: ver escala de integración.

Message Passing Interface: La interface de pases de mensajes MPI, por sus siglas en Inglés, (Message Passing Interface), es una biblioteca de funciones y subrutinas que pueden ser usadas en programas C, FORTRAN y C++. Con el uso de MPI en programas que modelan algún fenómeno o proceso de Ciencias e Ingeniería; se intenta explotar la existencia de múltiples procesadores a través del pase de mensajes. MPI fue desarrollado en los años 1993-1994 por un grupo de investigadores de la industria y la comunidad académica. Hoy en día MPI es una biblioteca estándar en la programación paralela basada en el pase de mensajes.

Microprocesador: Es un circuito integrado que contiene un procesador. Es casi el corazón de cada computadora y consta de miles de transistores y otros componentes, integrados en una pastilla o chip. Ver CPU.

Método de mínimos cuadrados: es un conjunto de pasos utilizado para aproximar un conjunto de puntos a un modelo, el cual puede ser lineal, cuadrado, exponencial, etc. Se basa en el principio de reducir la varianza al mínimo, adecuándolo a uno de los modelos anteriormente citados. Además nos brinda información importante sobre la tendencia que tendrá alguna variable en cuestión.

MOSFET: son las siglas de Metal Oxide Semiconductor Field Effect Transistor. Consiste en un transistor de efecto de campo basado en la estructura MOS. Fue ideado teóricamente por el alemán Lilinfield en 1930, aunque debido a problemas de carácter tecnológico y el desconocimiento acerca de cómo se comportan los electrones sobre la superficie del semiconductor; se pudieron fabricar décadas más tarde. Básicamente un MOSFET está formado por una placa de metal y un semiconductor, que hace las veces de soporte físico, separados por una zona de óxido del semiconductor de unos 100 nm de espesor. En este sistema se disponen cuatro electrodos: compuerta (gate), simbolizado con G; que se conecta a la placa metálica. Fuente (Source) y Drenador (Drain), ambos simétricos, que se internan en el cuarto electrodo llamado sustrato (Body), generalmente conectado eléctricamente con la fuente.

Muestreo: es la técnica para la selección de una muestra a partir de una población. Al elegir una muestra, se espera que sus propiedades sean extrapolables a la población. Este proceso permite

ahorrar recursos, obteniendo resultados parecidos a los obtenidos, si se realizara un estudio de toda la población.

Multithreads: se refiere a que dos o más tareas se ejecutan "aparentemente" a la vez, dentro de un mismo programa. Ver thread.

Objetos de conocimiento enriquecido: ver sección 3.1.3.

Objetos de conocimiento pobre: ver sección 3.1.3.

Ontología: Es un conjunto de términos estructurados jerárquicamente que describen un dominio. Es el esqueleto sobre el cual se construye la base del conocimiento.

PageRank: Es uno de los mejores secretos guardados de google. Se puede entender como una valoración que google tiene de las páginas webs. Una vez que las páginas son indexadas en la base de datos de google, cada cierto tiempo se calcula este valor, normalmente con una periodicidad aproximada de un mes. El Page Rank de una página se puede conocer con la google toolbar, la conocida barra de herramientas de google

Páginas Web: es un documento de la World Wide Web (que intenta ser estandarizada por el World Wide Web Consortium, también llamada W3C), normalmente en formato HTML.

Palabras claves: son palabras que representan un concepto concreto de la materia que estamos buscando. Las palabras clave junto con la descripción, el autor y el título, son las etiquetas de mayor interés para la identificación del contenido de un documento.

Paradigma: realizaciones científicas universalmente reconocidas que, durante cierto tiempo, proporcionan modelos de problemas y soluciones a una comunidad científica.

Pentium: Familia de microprocesadores fabricados por Intel, existen de gama baja como Celeron, de gama media como Pentium, Pentium II, Pentium III y Pentium IV o bien especialmente diseñados para funcionar en servidores y estaciones de trabajo como Pro y Xeon. Su nombre se debe a que iba a ser el sucesor del 486, y en vez de llamarlo 586 se cogió la palabra Pentium ideada a partir del griego Penta de significado 5.

Personal Computer: Computadora diseñado para uso individual, éste término se empezó a usar para indicar que una computadora era IBM (compatible que usaban Disk Operating System-DOS) con microprocesadores Intel, y así diferenciarlo de la arquitectura de un ordenador Apple basada en microprocesadores Motorola y otros sistemas operativos.

Período característico: es el tiempo en que cada rama científica produce cambios sustanciales.

Placa de Circuito Impreso: Placa lisa de material aislante, por ejemplo plástico o fibra de vidrio, sobre la que se montan chips y otros componentes electrónicos, generalmente en orificios previamente taladrados. Los componentes de una placa de circuito impreso, y más concretamente los orificios, están conectados eléctricamente mediante pistas de metal conductor definidas con anterioridad e impresas sobre la superficie de la placa. Las puntas metálicas que sobresalen de los componentes electrónicos se sueldan a las pistas metálicas conductoras formando las conexiones.

Principio de Landauer: al borrar un bit de información, se transfiere al ambiente por lo menos una energía $kBT \ln 2$. Al borrar un bit de información, la entropía del ambiente se incrementa al menos $kB \ln 2$. La computación reversible se puede realizar sin consumo de energía. Si el

sistema es reversible la entropía es constante. En una computadora cuántica, la interacción con el ambiente puede introducir ruido o decoherencia, lo que hace que el proceso no sea unitario, o sea reversible. Las medidas no son reversibles, y aumentan la entropía.

Proceso distribuido: es una forma de proceso en la que los datos y las funciones están distribuidos en los distintos elementos de una configuración o sistema que implica la presencia de una red de área local o una red de área amplia.

Proceso Paralelo: Es un tipo de proceso asimilable a los grandes sistemas. Consiste básicamente en procesar varias operaciones de modo simultáneo por distintas unidades centrales. En realidad, es un proceso cooperativo que implica una arquitectura de microprocesadores ligados entre sí y compartiendo tareas.

Procesos tecnológicos: consiste en una serie de pasos ordenados rigurosamente para la transferencia del diseño gráfico de un chip a una oblea de silicio monolítico o monobloque. Se trata de reproducir los rasgos del diseño: por fotolitografía o por barrido con haz de iones.

Punto de Inflexión: es el punto donde la concavidad de una curva cambia de arriba abajo –o viceversa–.

Revolución científica: se consideran como aquellos episodios de desarrollo no acumulativo en que un antiguo paradigma es reemplazado, completamente o en parte, por otro nuevo e incompatible.

Silicio: es un elemento químico no metálico situado en el grupo 14 de la tabla periódica de los elementos, formando parte de la familia de los carbonoideos. Es el segundo elemento más abundante en la corteza terrestre (27,7% en peso) después del oxígeno. Se presenta en forma amorfa y cristalizada; el primero es un polvo pardusco, más activo que la variante cristalina, que se presenta en octaedros de color azul grisáceo y brillo metálico. Se utiliza en aleaciones, en la preparación de las siliconas, en la industria de la cerámica técnica y, debido a que es un material semiconductor muy abundante, tiene un interés especial en la industria electrónica y microelectrónica como material básico para la creación de obleas o chips que se pueden implantar en transistores, pilas solares y una gran variedad de circuitos electrónicos.

Sistema distribuido: consiste en un conjunto de computadoras autónomas, comunicadas por una red y equipadas con un software, el cual permite a dichas computadoras coordinar sus actividades y compartir los recursos disponibles HSW.

Sistema paralelo consiste en un conjunto de procesadores, comunicados por una memoria común o por envío de mensajes, los cuales en forma conjunta ejecutan uno o varios programas, con el fin de aumentar la eficiencia del procesamiento e incrementar la velocidad de respuesta.

Small Scale Integration: ver escala de integración.

Software: Término general para nombrar a los programas y/o aplicaciones. Se puede describir como la parte lógica de un computador.

Supercomputadora: Una supercomputadora es el tipo de computadora más potente y más rápido que existe en un momento dado. Estas máquinas están diseñadas para procesar enormes cantidades de información en poco tiempo y son dedicadas a una tarea específica. Cuentan con un control de temperatura especial, esto para disipar el calor que algunos componentes alcanzan

a tener. Unos ejemplos de tareas a las que son expuestas las supercomputadoras son los siguientes: Búsqueda y estudio de la energía y armas nucleares.

Synset: es un conjunto de palabras de la misma categoría gramatical (pueden ser nombres, verbos, adjetivos o adverbios) con el mismo significado en un contexto dado. A partir de estos elementos, y mediante un conjunto limitado de relaciones, se establece una red semántica. Algunas de estas relaciones son sinonimia, antonimia, hiponimia, meronimia.

Sparseness: se refiere a la extrema dispersión de los ejemplos en el espacio de atributos

Tecnociencia: es un concepto ampliamente usado en la comunidad interdisciplinaria de estudios de ciencia y tecnología para designar el contexto social y tecnológico de la ciencia. La idea muestra un reconocimiento común de que el conocimiento científico, no es solo un código situado en la sociedad y la historia, sino que se sustenta y se hace durable por redes materiales no humanas

Tecnología doble núcleo: se refiere a procesadores que tienen en realidad dos unidades centrales de proceso. Es decir, dos cerebros que pueden trabajar simultáneamente en una única tarea o realizar trabajos diferentes sin que el rendimiento del otro se vea afectado. Esta nueva tecnología de microprocesadores permite aumentar el rendimiento sin consumir más energía ni generar un exceso de calor, ambos núcleos conviven dentro de un mismo ambiente, es decir, dentro del mismo chip.

Tecnología planar: Esta descripción tecnológica consiste en un conjunto de máscaras, que representan los diferentes materiales que forman tanto los dispositivos como las conexiones, que dan lugar al sistema electrónico diseñado. Este conjunto de máscaras recibe el nombre de layout.

Teoría unificada de la física: es aquella que, con una (o varias) ecuaciones, explicaría el complejo Universo en el que vivimos.

Texto plano: son archivos compuestos únicamente por texto sin formato, sólo caracteres. Estos caracteres se pueden codificar de distintos modos dependiendo de la lengua usada. Algunos de los sistemas de codificación más usados son: ASCII, ISO-8859-1 o Latín-1, etc. Se les conoce también como archivos de texto llano por carecer de información destinada a generar formatos (negritas, subrayado, cursivas, tamaño, etc.) y tipos de letra (por ejemplo, Arial, Times, Courier, etc.). Las aplicaciones destinadas a la escritura y modificación de archivos de texto se llaman editores de texto.

Thread o hilo, flujo de control del programa: representa un proceso individual ejecutándose en un sistema. A veces se les llama procesos ligeros o contextos de ejecución. Cada hilo controla un único aspecto dentro de un programa, como puede ser supervisar la entrada en un determinado periférico o controlar toda la entrada / salida del disco. Todos los hilos comparten los mismos recursos, al contrario que los procesos, en donde cada uno tiene su propia copia de código y datos (separados unos de otros).

Tiempo de acceso: se define como el tiempo mínimo entre dos accesos seguidos, que permite un determinado dispositivo. Este tiempo de acceso relaciona dos términos que son : latencia (básicamente es una búsqueda del "sitio" correcto del dispositivo para garantizar un correcto almacenamiento de los datos) y tiempo de transferencia.

Tolerancia a fallos: es la capacidad de un sistema a responder a un suceso inesperado, como puede ser un fallo de suministro eléctrico o un fallo de hardware o software; de forma que no se pierdan datos.

Transistor: es la contracción de transfer resistor, es decir, de resistencia de transferencia. El transistor es un dispositivo electrónico semiconductor que se utiliza como amplificador o conmutador electrónico. Es un componente clave en nuestra civilización ya que toda la electrónica moderna los utiliza en: circuitos integrados, microprocesadores, controladores de motores eléctricos de corriente continua y de pasos, y actualmente están integrados a todos los dispositivos electrónicos de utilización diaria: radios, televisores, grabadores, reproductores de audio y vídeo, hornos de microondas, lavarropas automáticos, automóviles, equipos de refrigeración, alarmas, relojes de cuarzo, calculadoras, impresoras, lámparas de iluminación fluorescentes, equipos de rayos X, tomógrafos, ecógrafos, etc.

Transistor BJT: se fabrica básicamente sobre un monocristal de Germanio, Silicio o Arseniuro de Galio. Sobre el sustrato de cristal se contaminan en forma muy controlada tres zonas, dos de las cuales son del mismo tipo, NPN o PNP; quedando formadas dos uniones NP.

Transistor de Efecto de Campo: ver FET.

Válvula de vacío: ver Bulbo.

Visual Heuristic Cluster Analysis for Texts: ver apéndice A

VLSI: ver escala de integración.

WordNet: Es una base de datos que contiene una red semántica y ha sido desarrollada por George Miller y su grupo de investigación en la Universidad de Princeton (Miller et al. 1993). Está diseñada utilizando un modelo semántico relacional y una base de datos también relacional y ha sido elaborada desde el punto de vista de la Psicolingüística. Está considerada como la red semántica del inglés (o de cualquier otra lengua) más completa que existe. Los elementos primitivos de WordNet son tres: Words (palabras, cadenas de caracteres), Meanings (significados, synsets, esto es, conjuntos de sinónimos, representados por un número identificador) and Relationships (relaciones, punteros relacionales).

Xeon: Microprocesador creado especialmente para funcionar en servidores o estaciones de trabajo. Diseñado para un perfecto trabajo en paralelo, basado en la arquitectura Pentium P6 de Intel.

Apéndice B

B.1. Listado de Programas.

B.1.1. Programa en MATLAB para calcular los coeficientes del polinomio cúbico para computación PCDS.

```
Y=[50,103,161,223,290,432,595,807,925,1112,1327,1544,1765,1990,2250]';  
% Y- vector columna de artículos  
X=[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14]'  
%X -vector -columna de (años -1990)  
plot(X,Y,'g')
```

hold on

```
%Mínimos cuadrados
```

```
x1=X;
```

```
x2=X.*X;
```

```
x3=x2.*X;
```

```
A3=[ones(size(x1)),x1,x2,x3];
```

```
C3=A3\Y;
```

```
Yteor3=A3*C3;
```

```
plot(X,Yteor3,'b')
```

```
C3
```

hold on

```
ytest=x3+2*x2+3*x1+4;
```

```
Ctest=A3\ytest
```

```
plot(X,ytest,'r*')
```

```
Ctest
```

```
salida →
```

$$a_0 = 63.2000$$

$$a_1 = 10.5024$$

$$a_2 = 15.0394$$

$$a_3 = -0.3422$$

B.1.2. Programa en MATHEMATICA para graficar el polinomio cúbico y la curva logística para computación PCDS.

```
Plot[{ -0.3422x3 + 15.0394x2 + 10.5024x + 63.2000 }, { x, -0.3451, 29.6446 }];
```

```
f[x_]:={-0.3422x3 + 15.0394x2 + 10.5024x + 63.2000};
```

```
Solns = Solve[ f'[x] == 0, x]
```

```
{{ x → -0.345098, { x → 29.6445}}
```

```
Plot[ {  $\frac{4737.7038.}{1 + e^{-0.1948(x-14.6497)}}$ , x, -4, 33 }];
```

```
Plot[ { -0.3422x3 + 15.0394x2 + 10.5024x + 63.2000,  $\frac{4737.7038.}{1 + e^{-0.1948(x-14.6497)}}$ , { x, -4, 33 },
```

```
PlotStyle → {Dashing[{0.009, 0.009}],
```

```
GrayLevel[0.0],
```

```
{Thickness[{0.01}]}}];
```

B.1.3. Programa en MATLAB para calcular los coeficientes del polinomio cúbico para VLSI y tolerancia a fallos.

```
x=[0:1:14]'
y=[37, 63, 91, 118, 150, 187, 223, 279, 314, 354, 407, 461, 511, 559, 596]'
x2=x.*x;
x3=x2.*x;
A=[ones(size(x)), x, x2, x3];
c=A\y
p(1)=c(4)
p(2)=c(3)
p(3)=c(2)
p(4)=c(1)
zz=polyval(p, x);
plot(x, y, 'r')
hold on
plot(x, zz)
```

salida →

$$\begin{aligned}
 a_0 &= 40.451 \\
 a_1 &= 18.343 \\
 a_2 &= 2.6081 \\
 a_3 &= -0.075289
 \end{aligned}$$

B.1.4. Programa en MATHEMATICA para graficar el polinomio cúbico y la curva logística para VLSI y tolerancia a fallos.

```
Plot[{ -0.075289x^3 + 2.6081x^2 + 18.343x + 40.451 }, { x, -4, 28 }];
f[x_]:= {-0.075289x^3 + 2.6081x^2 + 18.343x + 40.451 };
Solns = Solve[ f'[x] == 0, x ]
{{ x → -3.10033, { x → 26.1945 }}
Plot[ { {  $\frac{968.4528}{1 + e^{-0.2001(x-11.5470)}}$  }, x, -14, 43 } ];
Plot[ { -0.075289x^3 + 2.6081x^2 + 18.343x + 40.451,  $\frac{968.4528}{1 + e^{-0.2001(x-11.5470)}}$  }, { x, -8, 33 },
PlotStyle → {Dashing[{0.009, 0.009}],
GrayLevel[0.0],
{Thickness[{0.01}]}}];
```

B.1.5. Programa en MATHEMATICA para graficar el histograma de computación PCDS.

```
ListPlot[{{1990,50}, {1991,53}, {1992,58}, {1993,62}, {1994,67},{1995,142}, {1996,163}, {1997,212},  
{1998,118}, {1999,187}, {2000,215}, {2001,217}, {2002,221},{2003,225},{2004,229}}, PlotJoined →  
True];
```

B.1.6. Programa en MATHEMATICA para graficar la curva acumulada para computación PCDS.

```
ListPlot[{{1990,50}, {1991,103}, {1992,161}, {1993,223}, {1994,290},{1995,432}, {1996,595},  
{1997,807}, {1998,925}, {1999,1112}, {2000,1327}, {2001,1544},  
{2002,1765},{2003,1990},{2004,2219}}, PlotJoined → True];
```

B.1.7. Programa en MATLAB para aproximar mediante parábolas el desarrollo de la ley de Moore.

```
x=[60,70,72]'  
y=[-4.47712125 -4 -4]'  
c1=polyfit(x,y,2)  
t=60:0.5:72;  
y1=polyval(c1,t);  
plot(x,y,'r*'),  
hold on;  
plot(t,y1);  
hold on;
```

```
x=[72,74,76,78]';  
y=[-4 -3.77815125 -3.47712125 -3.47712125]';  
c2=polyfit(x,y,2)  
t=72:0.5:78;  
y2=polyval(c2,t);  
plot(x,y,'r*'),  
hold on;  
plot(t,y2);  
hold on;
```

```
x=[78,79,82,85]';  
y=[-3.47712125 -3.47712125 -3.17609126 -3.17609126]';  
c3=polyfit(x,y,2)  
t=78:0.5:85;  
y3=polyval(c3,t);  
plot(x,y,'r*'),  
hold on;  
plot(t,y3);  
hold on;
```

```
x=[85,89,93]';  
y=[-3.17609126,-3,-2.90308999]';  
c4=polyfit(x,y,2)  
t=85:0.5:93;  
y4=polyval(c4,t);  
plot(x,y,'r*'),  
hold on;  
plot(t,y4);  
hold on;  
  
x=[95,97,99,100,104]';  
y=[-2.77815125 -2.54406804,-2.39794001,-2.25527251,-2.11394335]';  
c5=polyfit(x,y,2)  
t=95:0.5:104;  
y5=polyval(c5,t);  
plot(x,y,'r*'),  
hold on;  
plot(t,y5);  
hold on;  
  
x=[105,106,107,109,111]';  
y=[-1.95424251,-1.81291336,-1.65321251,-1.50514998,-1.34242268]';  
c6=polyfit(x,y,2)  
t=105:0.5:111;  
y6=polyval(c6,t);  
plot(x,y,'r*'),  
hold on;  
plot(t,y6);  
hold on;  
  
salida →
```

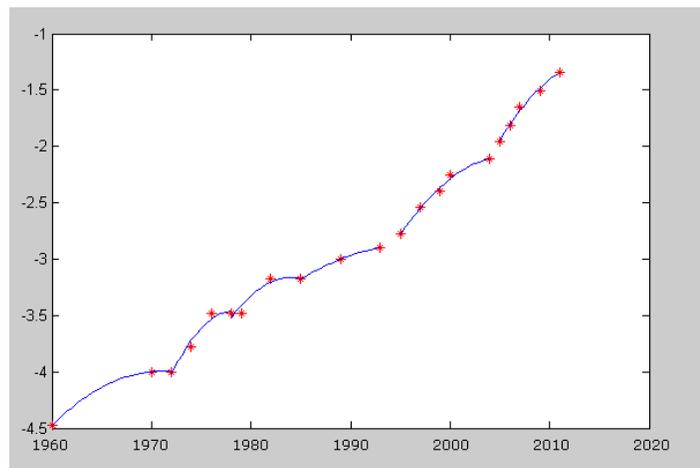


Figura B.1. Salida del programa B.1.7.

B.1.8. Programa en MATLAB para obtener la curva tangente al conjunto de curvas S del desarrollo de la ley de Moore.

```
s0=-8.2304
s0prim=0
x0=-21
for i=1:4,
a1=input('a1: ');
b1=input('b1: ');
c1=input('c1: ');
x1=-((c1-s0)*2+x0*(b1+s0prim))/(2*a1*x0+b1-s0prim);
A1=(a1*x1+0.5*(b1-s0prim))/(x1-x0);
B1=s0prim-2*A1*x0;
C1=s0-A1*x0*x0-B1*x0;
s1=A1*x1*x1+B1*x1+C1;
s1prim=2*A1*x1+B1;
xx=x0:1:x1+10;
F1=A1*xx.*xx+B1*xx+C1;
plot(xx,F1,'r*'),
hold on;
ss1=a1*xx.*xx+b1*xx+c1;
plot(xx,ss1,'g--');
hold on;
s0=s1
s0prim=s1prim
x0=x1
end

s0 = -8.2304
s0prim = 0
x0 = -21
a1: -0.003976
b1: 0.56459
c1: -24.039
s0 = -4.5289
s0prim = 0.09205

salida →
```

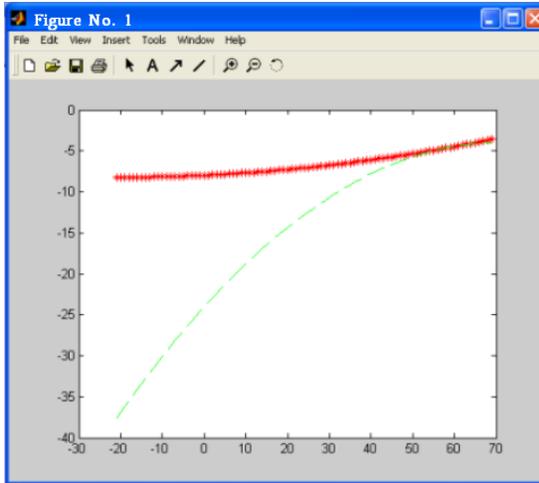


Figura B.2. Salida del programa B.1.8, primera iteración.

$x_0 = 59.424$
 $a_1 = -0.013866$
 $b_1 = 2.1733$
 $c_1 = -88.619$
 $s_0 = -3.4719$
 $s_{0prim} = 0.025158$
 $x_0 = 77.461$

salida →

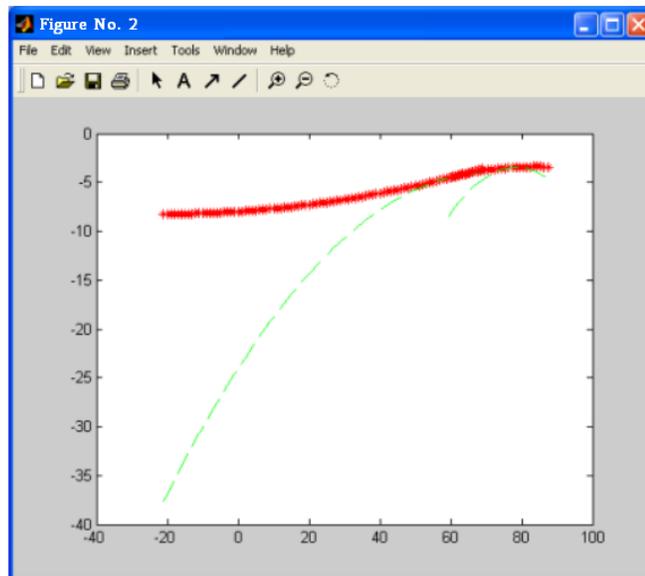


Figura B.3. Salida del programa B.1.8, segunda iteración.

a1: -0.0091221
b1: 1.5371
c1: -67.911
s0 = -3.3523
s0prim = 0.083855
x0 = 79.655

salida →

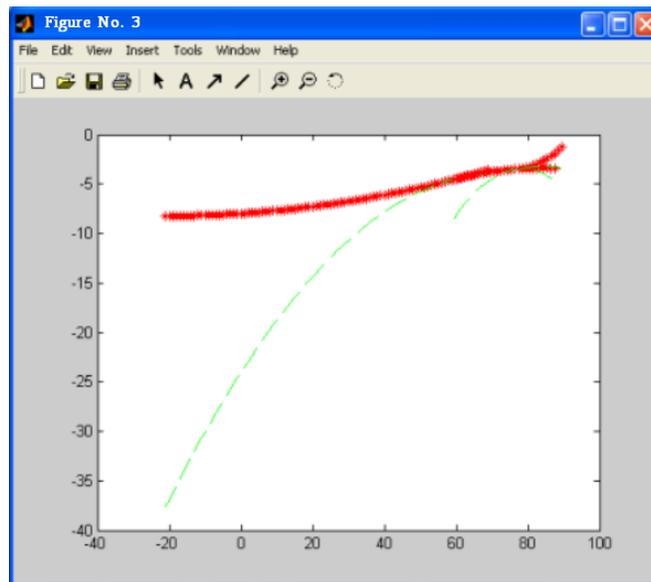


Figura B.4. Salida del programa B.1.8, tercera iteración.

a1: -0.0085432
b1: 1.9452
c1: -112
s0 = -1.3311
s0prim = 0.044011
x0 = 111.27

salida →

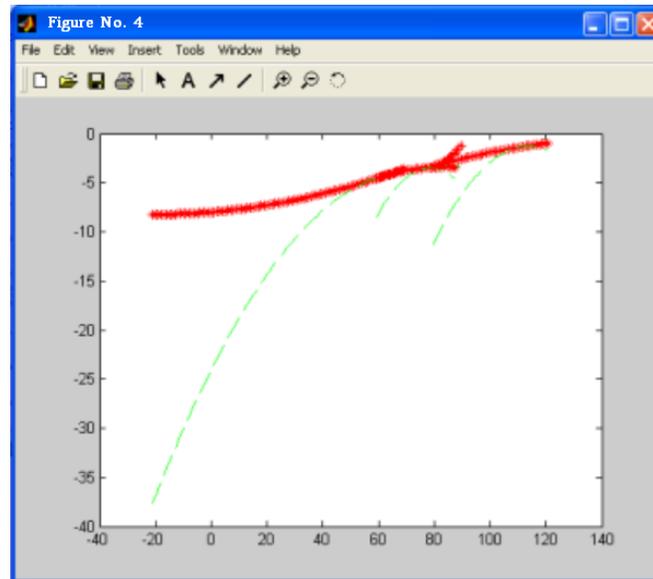


Figura B.5. Salida del programa B.1.8, cuarta iteración.

```
s0(1)=-8.2304;  
s1(1)=0;  
s1prim(1)=0;  
s0prim(1)=0;  
s0(2)=0;  
s1(2)=0;  
s1prim(2)=0;  
s0prim(2)=0;  
s0(3)=0;  
s1(3)=0;  
s1prim(3)=0;  
s0prim(3)=0;  
s0(4)=0;  
s1(4)=0;  
s1prim(4)=0;  
s0prim(4)=0;  
x0(1)=-21;  
x0(2)=0;  
x0(3)=0;  
x0(4)=0;  
x1(1)=0;  
x1(2)=0;  
x1(3)=0;  
x1(4)=0;  
a1(1)=-0.003976;  
b1(1)=0.56459;  
c1(1)=-24.039;
```

```

a1(2)=-0.013866;
b1(2)=2.1733;
c1(2)=-88.619;
a1(3)=-0.0091221;
b1(3)=1.5371;
c1(3)=-67.911;
a1(4)=-0.0091858;
b1(4)=2.0795;
c1(4)=-119.03;
for i=1:4,
x1(i)=-((c1(i)-s0(i))*2+x0(i)*(b1(i)+s0prim(i)))/(2*a1(i)*x0(i)+b1(i)-s0prim(i));
A1(i)=(a1(i)*x1(i)+0.5*(b1(i)-s0prim(i)))/(x1(i)-x0(i))
B1(i)=s0prim(i)-2*A1*x0(i)
C1(i)=s0(i)-A1(i)*x0(i)*x0(i)-B1(i)*x0(i)
s1(i)=A1(i)*x1(i)*x1(i)+B1(i)*x1(i)+C1(i)
s1prim(i)=2*A1(i)*x1(i)+B1(i)
xx=x0(i):1:x1(i)+5;
F1=A1(i)*xx.*xx+B1(i)*xx+C1(i);
plot(xx,F1,'r*'),
hold on;
xxx=(x1(i)-5):1:(x1(i)+5);
%ss1(i,:)=a1(i)*xxx.*xxx+b1(i)*xx+c1(i);
%plot(xxx,ss1,'g--');
hold on;
if (i < 4) s0(i)=s1(i)
s0prim(i)=s1prim(i)
x0(i)=x1(i)
end
end

salida →
    
```

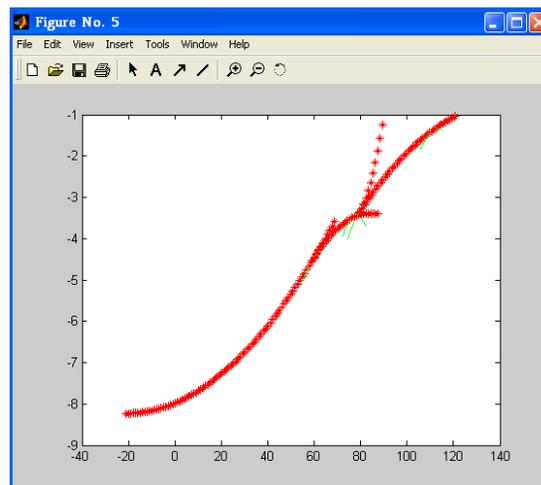


Figura B.6. Salida del programa B.1.8, quinta iteración.

Apéndice C

C.1. Tutorial del sistema automatizado Visual Heuristic Cluster Analysis.

El sistema incluye dos programas:

- 1.- Administrador de bases de datos de textos completos.
- 2.- Análisis heurístico de clústeres para matrices numéricas.

Después de la instalación del sistema, se observa en la figura C.1 el menú completo de los íconos correspondientes al programa.

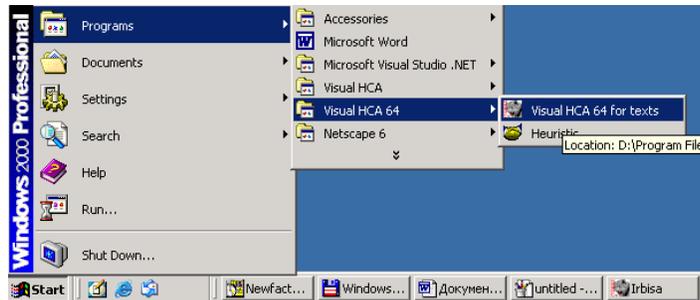


Figura C.1. Menú de VHCA.

Para el manejo del sistema VHCA se consideran las siguientes etapas:

1.- Construcción de la base de datos del arreglo de textos. Los textos deben ser preparados en codificación WIN en formato (txt). Más adelante se mostrará el trabajo con arreglo de textos (artículos sobre sociología y ecología). Estos textos se encuentran en el directorio HELP \ TEXTS.

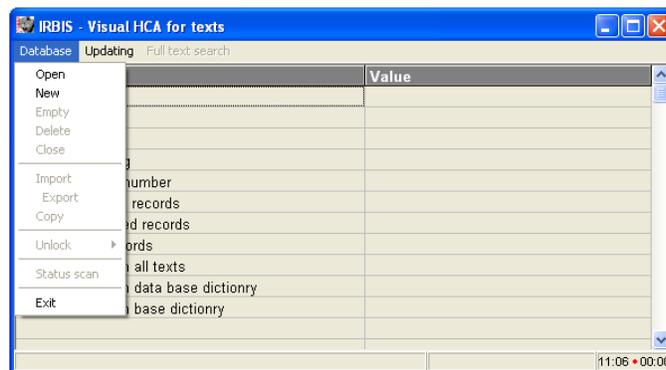


Figura C.2. Construcción de la base de datos de textos completos.

En la figura C.3 se muestra la ventana para construir el diccionario básico de palabras.

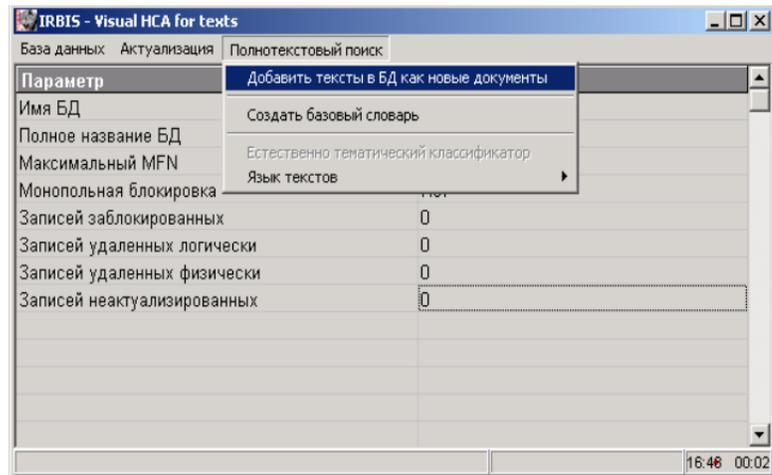


Figura C.3. Creación del diccionario básico.

Dentro de la opción para construir el diccionario básico se muestra la ventana para selección de textos de la base de datos como muestra la figura C.4.

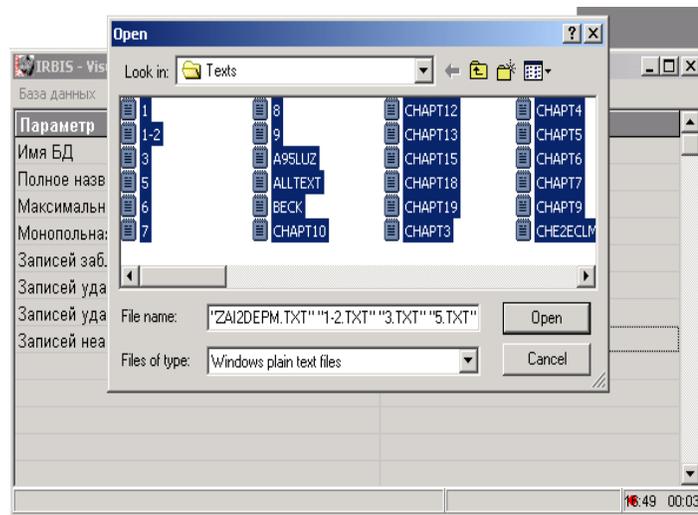


Figura C.4. Selección de textos de la base de datos.

2.- Creación de los índices de la base de datos (extracción de las palabras de los textos y creación del diccionario). Se crea simultáneamente el diccionario básico de la colección de palabras cuya frecuencia supera la frecuencia de uso común. El diccionario básico incluye palabras de todos los sub-lenguajes usados por los autores.

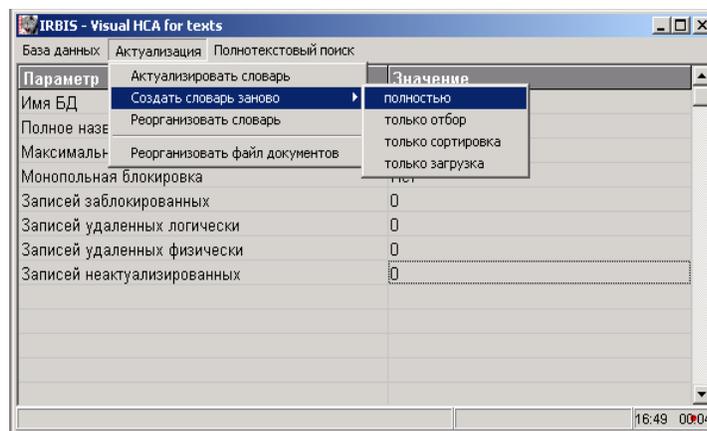


Figura C.5. Modificación del diccionario básico.

3.- Ingresar al régimen: «clasificación orientado a dominio».

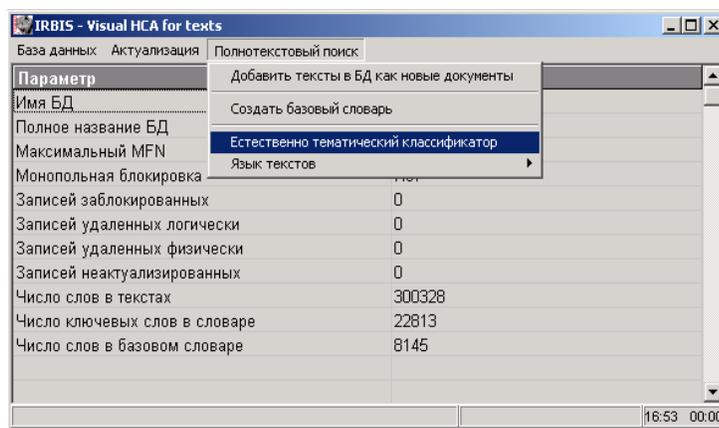


Figura C.6. Ventana para ingresar al régimen de clasificación orientado a dominio.

4.- Presionar el botón derecho del ratón sobre el panel superior o en el menú, y elegir «crear diccionario orientado a dominio».

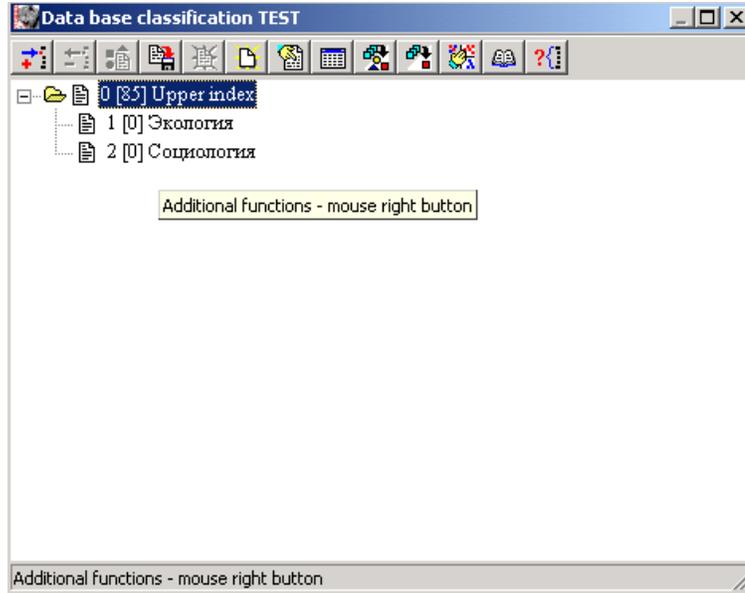


Figura C.7. Menú para distribución de textos por dominio y sub-dominio.

5.- Con la acción anterior aparece la ventana para elegir los criterios de la selección de las palabras del diccionario básico en DOD (con el botón derecho del ratón se selecciona el régimen (opción) que muestra el gráfico de la figura C.8. Los cursores establecen distintos criterios de selección después que se calcula el número de palabras que satisfacen los criterios y el número de las palabras en DOD) que satisfacen todos los criterios.

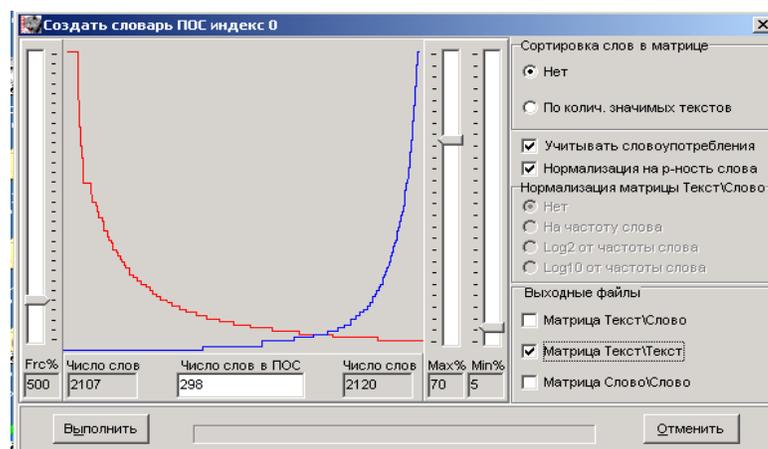


Figura C.8. Ventana para elegir los criterios de selección de las palabras del diccionario básico en DOD.

Los DOD's son herramientas básicas de clasificación de textos. Las palabras del diccionario básico pasan a través de una serie de criterios de selección. Así, DOD es parte de las palabras del diccionario básico que satisfacen los criterios dados por el experto. Los criterios de selección tienen el siguiente orden:

- el número máximo de textos significativos (el número máximo permitido de textos en el cual fue encontrada la palabra). Este criterio permite aislar las palabras;
- el número mínimo de textos significativos (el número mínimo de textos en el cual fue encontrada la palabra de los candidatos de DOD). Este criterio coincide en realidad con el límite estadístico inferior de la frecuencia de la palabra, rechazando palabras específicas para pequeñas cantidades de textos.
- Exceder la frecuencia relativa asignada.

Después se calcula la matriz numérica texto / palabra (TW, text/word), los elementos de esta matriz son el número de veces que una palabra dada de DOD es encontrada en el texto. La matriz TW contiene la cantidad de palabras de DOD (columna) en cada texto (fila) como elementos de ello y se usa para calcular las matrices cuadradas simétricas: palabra / palabra (WW) y texto / texto (TT). Los elementos de la matriz TT contiene la cantidad de las mismas palabras de DOD en cada par de textos como elementos de ello. El elemento de la matriz WW contiene la cantidad de textos que contienen los pares de palabras de columna y fila como elementos de ello, en otras palabras es la cantidad de textos en los que se usa un par dado de las palabras de DOD. La matriz WW se usa para revelar los DOD 's parciales, la matriz TT se usa para revelar grupos de textos con temas.

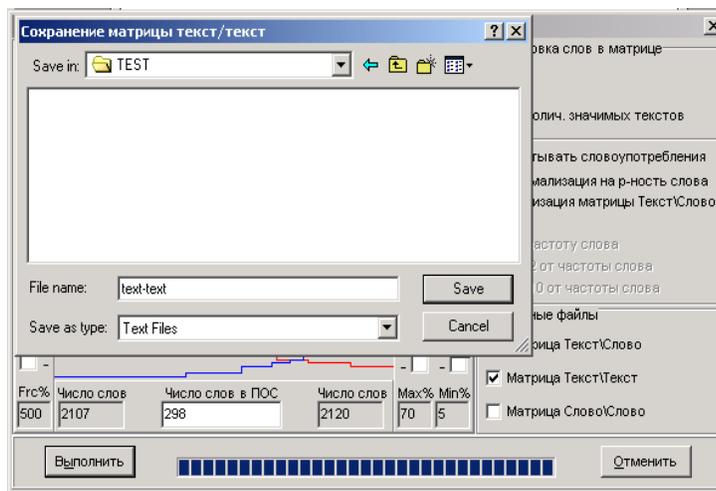


Figura C.9. Ventana para guardar las matrices TW, TT y TW obtenidas.

Después se crea DOD y se calculan las matrices en la opción «mostrar DOD». Para trabajar dentro de este régimen es necesario indicarle a WINDOWS la apertura de los ficheros con la extensión txt, a través del block de notas.

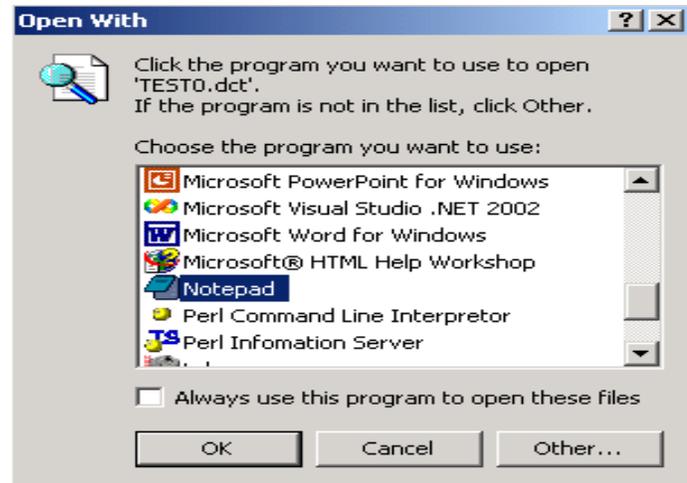


Figura C.10. Apertura de archivos con el block de notas.

El régimen «analizar DOD» se utiliza para el cálculo de las matrices.

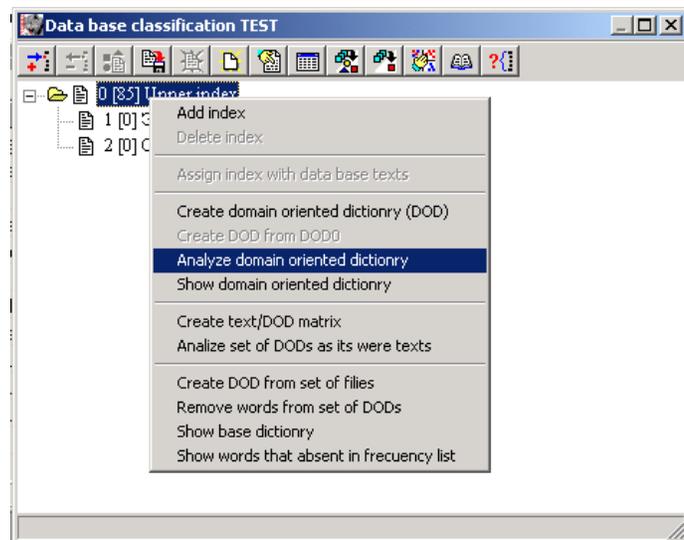


Figura C.11. Menú para analizar el Diccionario Orientado a Dominio.

6.- Análisis de la matriz texto \ texto con la ayuda de VHCA.

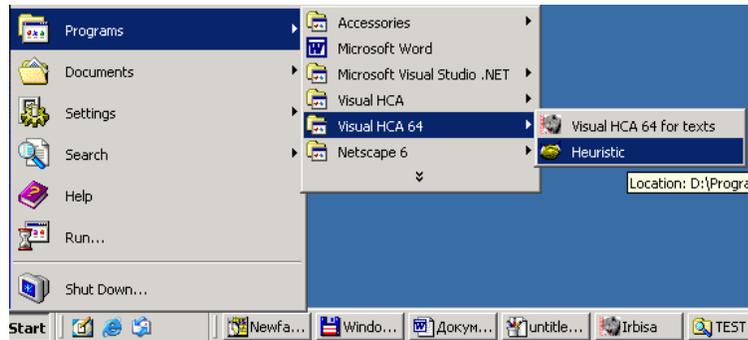


Figura C.12. Apertura de VHCA.

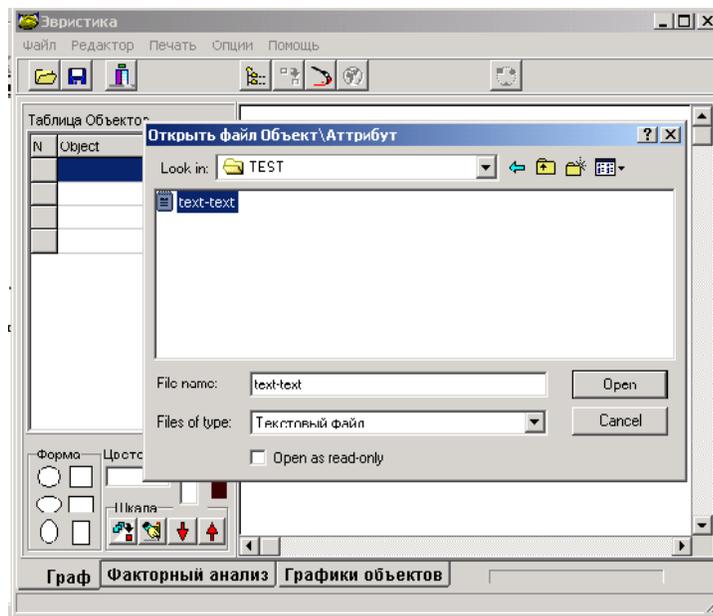


Figura C.13. Ventana que muestra la apertura de la matriz texto / texto.

Después de la apertura del fichero se muestra la ventana del «redactor» que muestra la matriz simétrica TT y la escala de color correspondiente en la parte izquierda. Las amplitudes se muestran en colores diferentes.

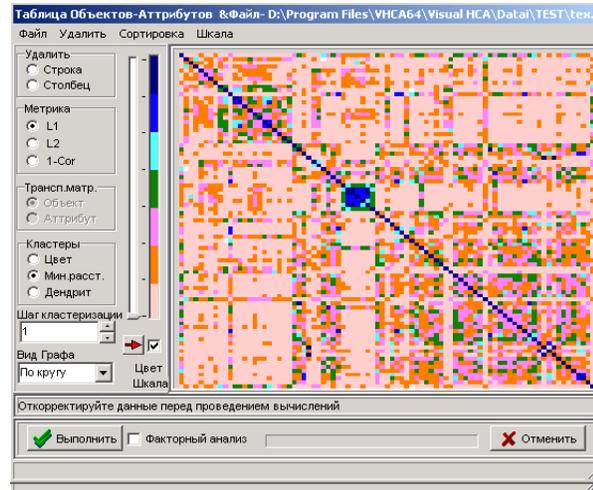


Figura C.14. Matriz simétrica TT en la ventana del redactor.

Presionando el botón «cumplir» se realiza un cálculo de las distancias entre los objetos (en este caso los textos). Presionando el botón izquierdo del ratón debajo de la escala de color se incluye el régimen automático de clustering; además los objetos son unidos en clústeres estables como muestra la figura C.15.

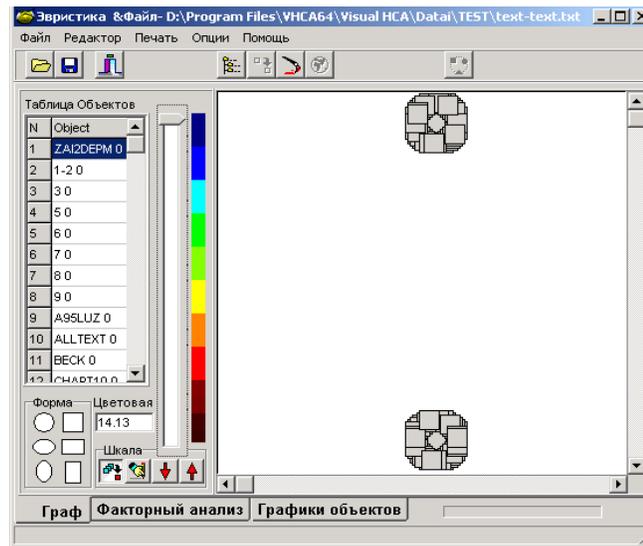


Figura C.15. Agrupación de los dos clústeres mas fuertes.

Presionando el botón izquierdo sobre el panel superior, traslada el programa al régimen de «representación en forma de árbol» figura C.16. Es necesario tomar el árbol de clústeres del gráfico correspondiente (para esto se presiona el botón «unir cluster» sobre el panel superior). Se reciben 2 clústeres los cuales son representados sobre la columna como se muestra en la figura C.15.

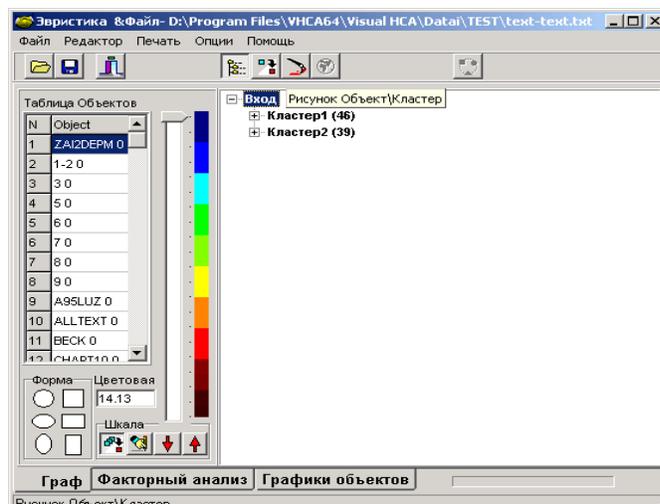


Figura C.16. Resultados de los clústeres fijados en forma de árbol.

Después se colorea el grupo de clústeres, seleccionando el icono que contiene el pincel, como muestra la figura C.17.

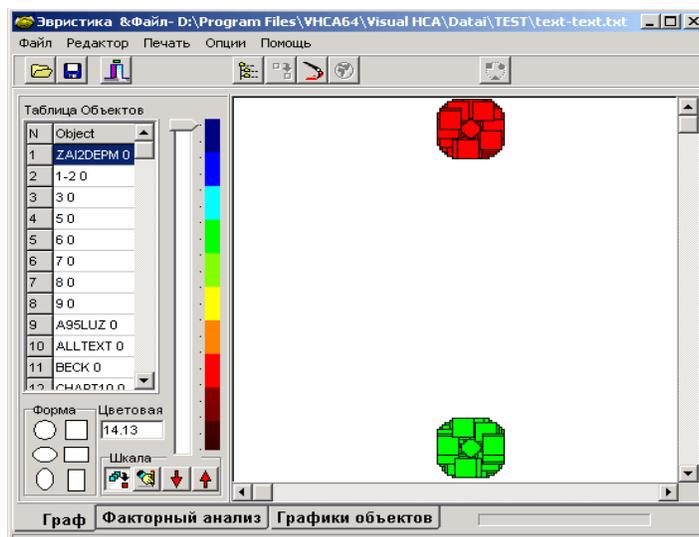


Figura C.17. Coloreado de los clústeres obtenidos.

Del menú principal se escoge el «redactor», como se muestra en la figura C.18.

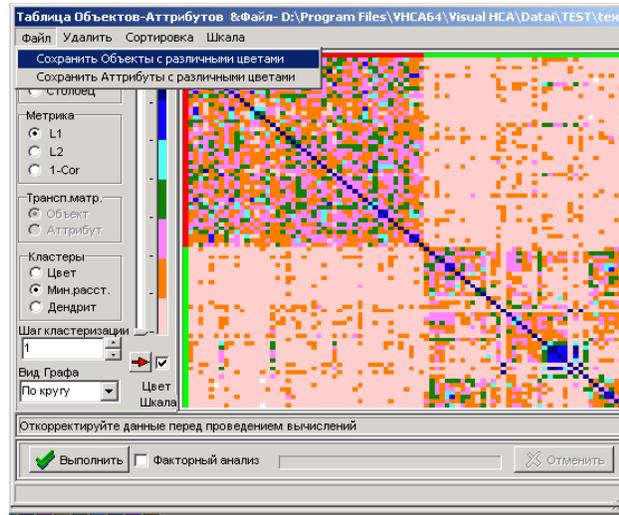


Figura C.18. Resultados del clustering.

En el cual se observan dos clústeres claramente identificados. Se conservan los objetos con distintos colores. Después de examinar varios archivos es posible determinar el nombre de los índices.

7.- Se crea el clasificador orientado a dominio. Después de agregar los sub-párrafos de índice 0 y 2 como muestra la figura C.19; es necesario ligar cada uno de ellos con los textos.

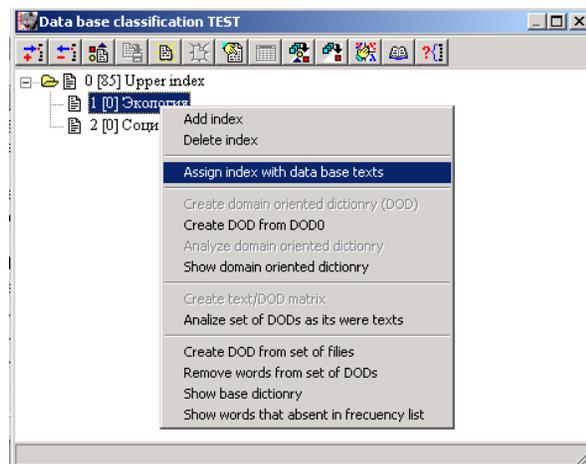


Figura C.19. Asignación de índices con la base de datos.

Y después se carga el primer cluster como se muestra en la figura C.20.

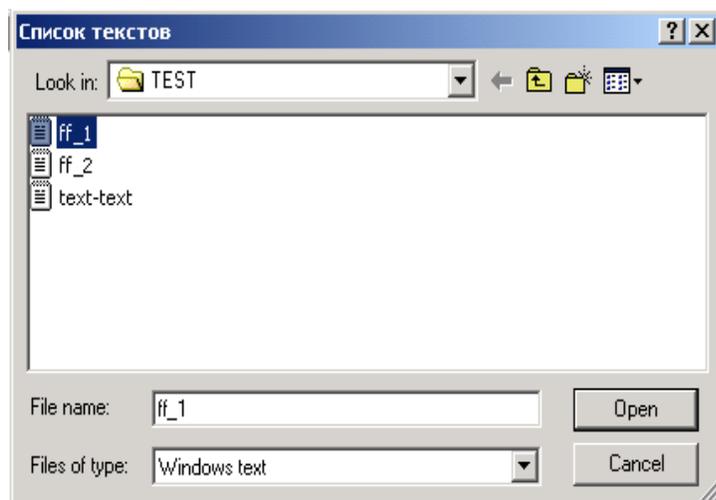


Figura C.20. Carga del primer cluster.

8.- Para cada índice temático se crea DOD realizando las acciones conocidas.

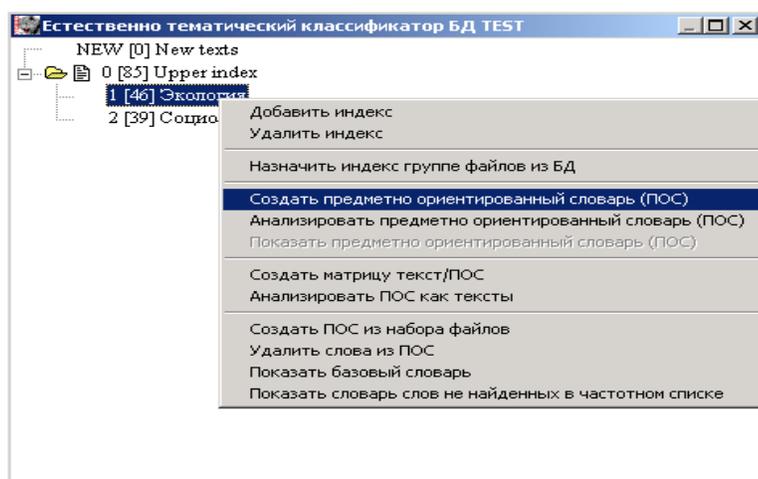


Figura C.21. Creación de DOD del primer cluster.

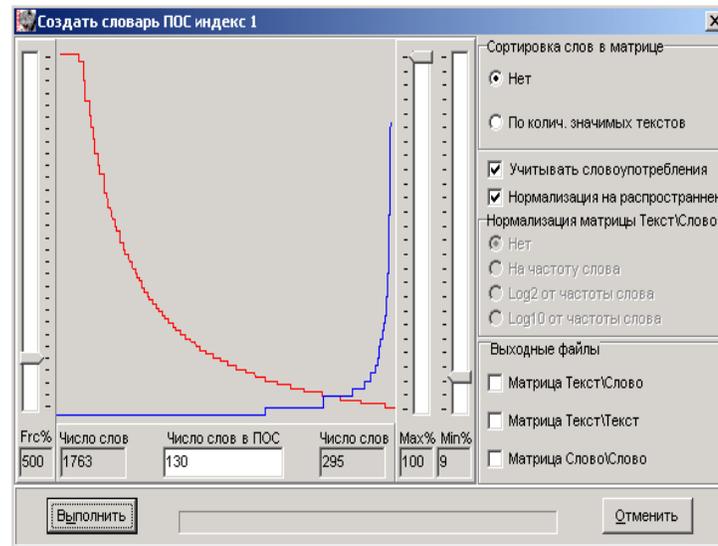


Figura C.22. Ventana para creación de matrices TT, WW, TW.

9. Se comprueba que los textos están correctamente identificados por los diccionarios DOD. Para esto se utiliza el régimen «crear matriz texto \ DOD». La matriz texto \ DOD es el porcentaje de las palabras de DOD en un texto dado.

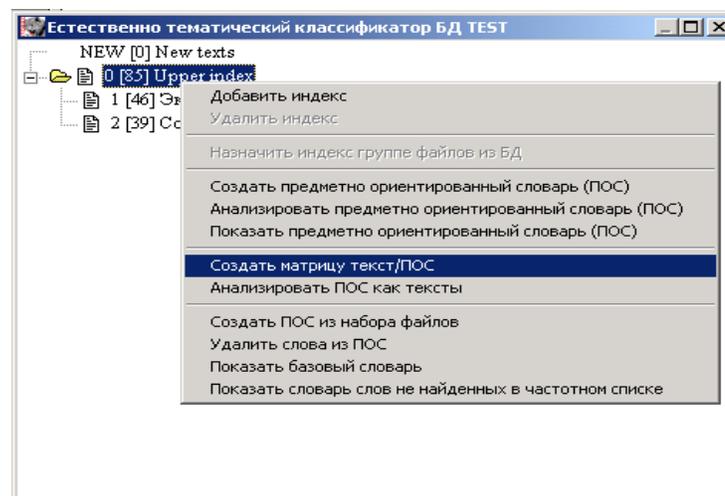


Figura C.23. Menú para crear la matriz texto/DOD.

En la figura C.24 se muestran los resultados del clustering

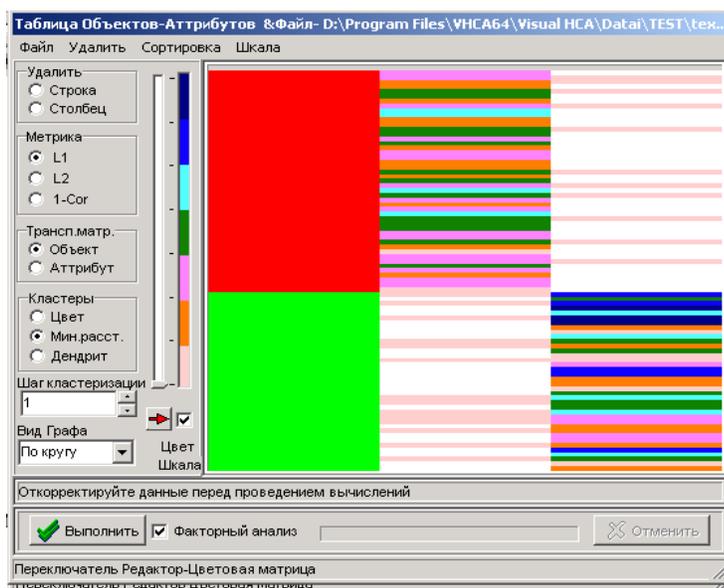


Figura C.24. Resultados del clustering.

Y finalmente en la figura C.25 se muestra el cálculo de los factores.

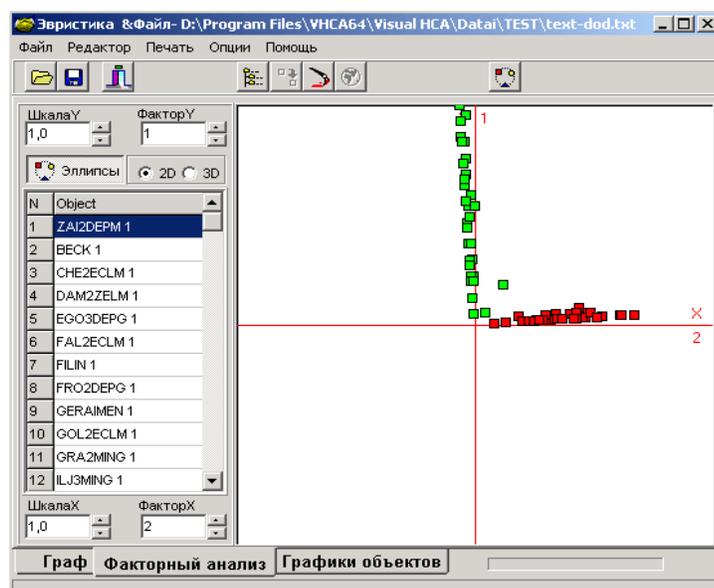


Figura C.25. Calculo de factores.

Apéndice D

D.1. Distribución temporal de resúmenes para VLSI y tolerancia a fallos para el periodo 1998-2010.

Considerando la información disponible gratuitamente en la Librería Digital de ACM Guide to Computing, se coleccionaron 5039 resúmenes para un período de desarrollo de 13 años. Este período se limitó de 1998 al 2010 debido a los resultados obtenidos en la ontología VLSI-FT del capítulo 3; los cuales revelaron dos etapas de desarrollo de 1990-1997 y 1998-2004, por lo cual se consideró necesario conocer el comportamiento para el periodo 1998-2010. Por otra parte se

excluyeron del análisis los artículos publicados durante el año 2011, debido a que únicamente se encuentra disponible el primer trimestre; incluir esta información en la investigación cambia la tendencia real del desarrollo de VLSI FT.

En las columnas 2, 5 y 8 de tabla D.1. se observa la cantidad de resúmenes recopilados de la Librería Digital de ACM Guide to Computing, mientras que en las columnas 3, 6 y 9 se muestra los datos acumulados de VLSI-FT.

| Año | Cantidad de resúmenes | Datos acumulados | Año | Cantidad de resúmenes | Datos acumulados | Año | Cantidad de resúmenes | Datos acumulados |
|------|-----------------------|------------------|------|-----------------------|------------------|------|-----------------------|------------------|
| 1997 | 141 | 863 | 2002 | 249 | 2016 | 2007 | 550 | 4530 |
| 1998 | 215 | 1078 | 2003 | 420 | 2436 | 2008 | 569 | 5099 |
| 1999 | 195 | 1273 | 2004 | 397 | 2833 | 2009 | 409 | 5508 |
| 2000 | 219 | 1492 | 2005 | 634 | 3467 | 2010 | 394 | 5902 |
| 2001 | 275 | 1767 | 2006 | 513 | 3980 | - | - | - |

Tabla D.1. Cantidad de resúmenes y datos acumulados para el periodo 1998-2010 para VLSI-FT.

La figura D.1 muestra la gráfica de las columnas 2, 5 y 8 de la tabla anterior; se observa que la tendencia es ‘creciente’ hasta el año 2005. En este año se tiene una colección de 634 resúmenes que se relacionan con temas de VLSI y tolerancia a fallos. En el año 2006 el interés por estos temas disminuye; una posible causa es que el interés de los investigadores se haya enfocado a nuevos temas de investigación o el apoyo económico se haya reducido. En el año 2008 se observa un pequeño incremento y nuevamente disminuye en los siguientes dos años.

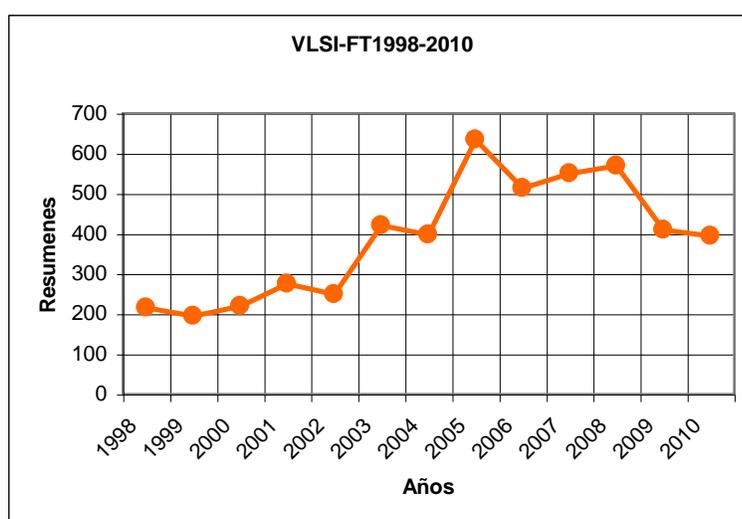


Figura D.1. Histograma de distribución de los resúmenes de la Librería Digital de ACM Guide to Computing Literature para VLSI y tolerancia a fallos para el periodo de 1998 al 2010.

El estudio de la ciencia normal para VLSI-FT permite conocer las etapas de inicio, máximo interés, así como la etapa cuando ésta será obsoleta. La gráfica color rojo de la figura D.2 muestra la curva acumulada de las columnas 3, 6 y 9 de la tabla D.1 para los 5039 resúmenes. Esta curva ascendente tiene un carácter exponencial en forma de **curva J** [37]. Según la **ley de Verhulst** [URL-13] ningún sistema con crecimiento exponencial puede seguir indefinidamente. Eventualmente, se alcanza un límite después de cierto período, y la curva J va tomando la forma de una curva S. Aprovechando la información de la curva J es posible restaurar completamente la curva S. Esta última curva se puede aproximar mediante un polinomio cúbico o la función logística.

La gráfica color azul de la figura D.2 muestra la línea de tendencia del polinomio cúbico obtenido por el método de mínimos cuadrados.

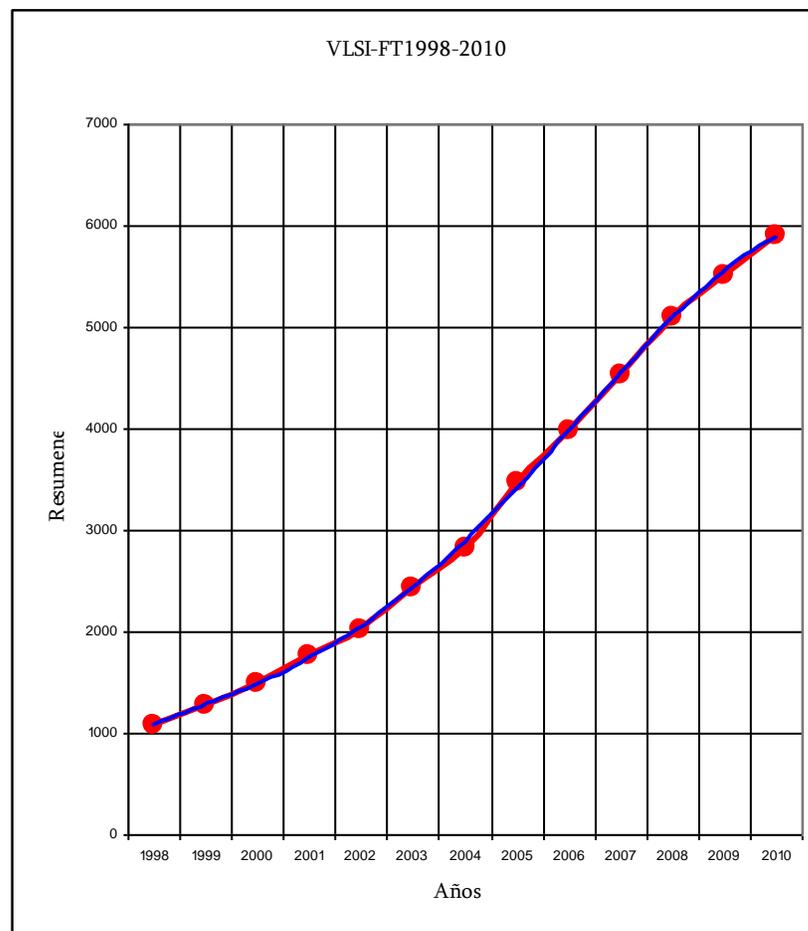


Figura D.2. Distribución acumulada (curva color rojo) y aproximación de la curva mediante un polinomio de tercer orden (curva color azul) para VLSI y tolerancia a fallos periodo 1998-2010.

D.2. Aproximación de la Curva S de VLSI-FT mediante un polinomio cúbico.

Para el cálculo de los coeficientes del polinomio cúbico se utiliza la metodología de la sección 4.1.3.2 y se tiene que:

$$\begin{aligned} a_0 &= 1417.1 \\ a_1 &= 236.3 \\ a_2 &= 46.6 \\ a_3 &= -2.5 \end{aligned} \quad (D.1)$$

sustituyendo estos coeficientes en la ecuación (4.5) del capítulo 4 se tiene el polinomio cúbico:

$$P_3(x)_{VLSI-FT} = -2.5x^3 + 46.6x^2 + 236.3x + 1417.1 \quad (D.2)$$

la grafica de la ecuación D.2 se muestra en la figura D.3, donde se observa el punto de inflexión $x_{\text{inf}(VLSI-FT)} = 6.21$ y representa al mes de febrero 2006. En esta fecha se manifestó el mayor progreso de VLSI-FT. Para fechas posteriores, el interés decrece hasta el punto $x_{\text{max}(VLSI-FT)} = 14.58$, el cual corresponde al mes de junio del 2014. Momento en que los temas VLSI-FT llegarán a la etapa de decadencia. El punto $x_{\text{min}(VLSI-FT)} = -2.15$ indica el inicio de VLSI-FT, el cual representa el mes de Enero de 1998. Los resultados obtenidos en la sección 4.1.7 contradicen este resultado pero se debe recordar que para este análisis se consideró el periodo 1998-2010 y en la sección 4.1.7 el periodo 1990-2004 por lo que el nacimiento real de VLSI y tolerancia fallos tuvo lugar en el mes de enero de 1987 el cual coincide con [41] y [42].

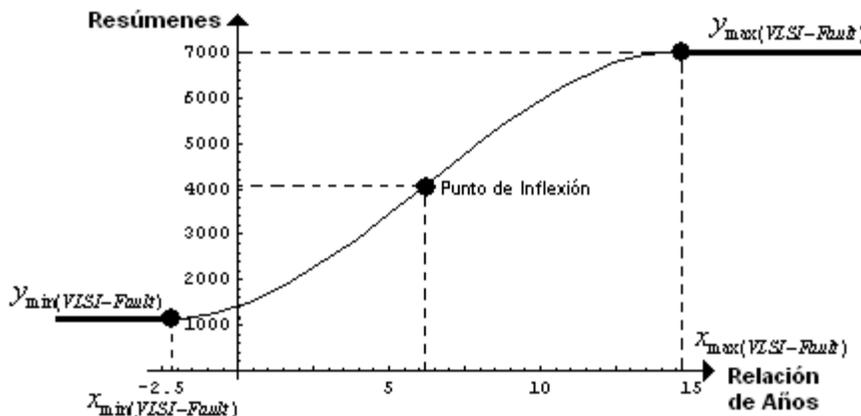


Figura D.3. Aproximación de la curva S de VLSI-FT con el polinomio cúbico de la ecuación D.13.

D.3. Aproximación de la Curva S de VLSI-FT mediante la función logística.

Así como en el capítulo 4 se obtienen los valores para α , β y γ ; la ecuación logística toma la siguiente forma:

$$l_{(VLSI-FT)} = \frac{8169.28}{1+e^{-0.2574(x-6.21)}} \quad (D.3)$$

En la figura D.4 se muestra la gráfica y se aprecia que la curva sigue un comportamiento en forma de S. En la parte central se aprecia el punto de inflexión, con coordenadas (6.21, 4084.64). Este punto representa la fecha de máximo desarrollo para VLSI-FT, así como la cantidad de resúmenes recopilados hasta esa fecha. Estos datos corresponden al mes de febrero de 2006. El desarrollo de computación VLSI-FT se encuentra limitada por el valor de $\alpha_{(VLSI-FT)} = 8169.28$; ésta indica el número de publicaciones (resúmenes) recopiladas, desde 1998 hasta el año 2014 y que sirve de asíntota a la función logística.

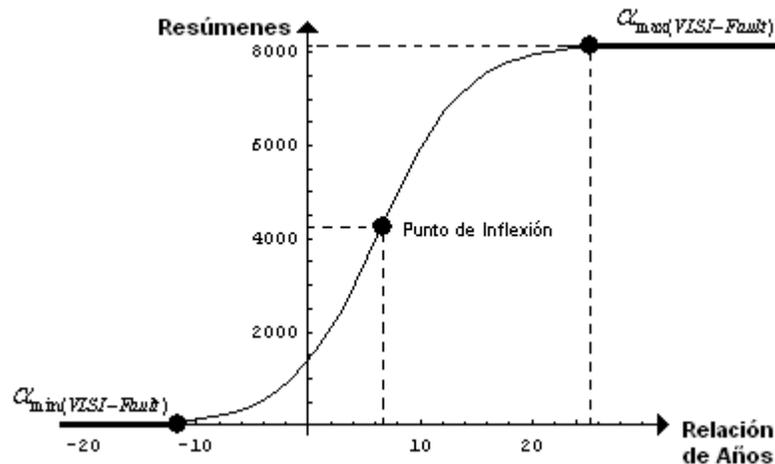


Figura D.4. Nacimiento, máximo desarrollo y muerte de computación PCDS, según la curva logística.

D.4. Superposición de las curvas cúbica y logística para VLSI y tolerancia a fallos para el período 1998-2010.

Después de realizar el análisis para ambas curvas es necesario conocer la discrepancia que existe entre ellas. Se debe recordar que tanto el punto de inflexión como la inclinación de la tangente en este punto, es la misma para ambas curvas. Esto se muestra en la figura D.5. La curva punteada corresponde al polinomio cúbico, mientras que la curva continua corresponde a la

función logística. La asíntota superior de la ecuación logística se sitúa en $\alpha_{\max(VLSI-FT)} = 8169.28$, mientras que la recta horizontal superior del polinomio cúbico, se encuentra en $y_{\max(VLSI-FT)} = 7020.01$. La asíntota inferior de la función logística se localiza en $\alpha_{\min(VLSI-FT)} = 0$, mientras que de la recta horizontal mínima del polinomio cúbico en $y_{\min(VLSI-FT)} = 1149.3$. Los puntos de inflexión coinciden en la parte central de la figura y tienen por coordenadas (6.21, 4084.64). En la figura D.5 se observa que el polinomio cúbico alcanza más rápidamente sus puntos mínimo y máximo, lo que demuestra que tiene un período mas corto de desarrollo, comparado con el modelo de la función logística.

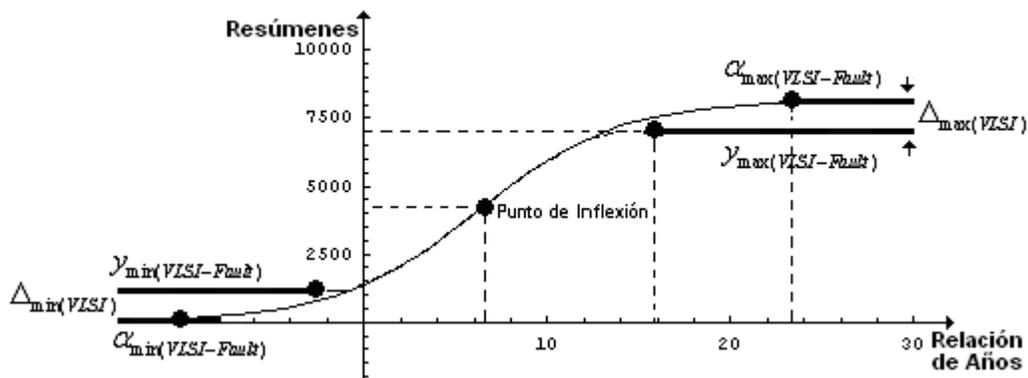


Figura D.5. Comportamiento de la función logística y el polinomio cúbico para VLSI-FT.

Para calcular la diferencia entre la asíntota superior del polinomio cúbico y la asíntota superior de la función logística, se utiliza la siguiente ecuación.

$$\Delta_{\max(VLSI-FT)} = \alpha_{\max(VLSI-FT)} - y_{\max(VLSI)} = 1149.27 \quad (D.4)$$

De igual forma, para calcular la diferencia entre la asíntota inferior del polinomio cúbico y la asíntota inferior de la función logística, se tiene:

$$\Delta_{\min(VLSI-FT)} = \alpha_{\min(VLSI-FT)} - y_{\min(VLSI-FT)} = 1149.27 \quad (D.5)$$

Se observa que el resultado de las ecuaciones (D.4) y (D.5) son iguales, por lo tanto

$$\Delta_{\max(VLSI-FT)} = \Delta_{\min(VLSI-FT)} \quad (D.6)$$

Lo anterior indica que las curvas son simétricas con relación al punto de inflexión, por lo que, la discrepancia es la misma, tanto en la parte superior como en la parte inferior de la curva, con un error equivalente al 14%. Los resultados obtenidos para el periodo 1998-2010 comprueban y validan los obtenidos en el capítulo 4.

Apéndice E

E.1. Material Generado.

Esta investigación ha sido parcialmente apoyada por la Universidad Tecnológica de la Mixteca, Universidad de la Sierra Sur y el Consejo Nacional de Ciencia y Tecnología a través del proyecto N-39011-A. Durante el desarrollo de esta investigación se realizaron las siguientes publicaciones:

- **Study of Knowledge Evolution in Parallel Computing by Short Texts Analysis.** In: Progress in Pattern Recognition, Image Analysis and Applications. Lecture Notes in Computer Science. LNCS 3287, Springer. 2004. ISBN 3-540-23527-2. Makagonov P, Ruiz Figueroa A. Berlin, Germany.
- **A Method of Rapid Prototyping of Evolving Ontologies** In: Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science. LNCS 3406, Springer. 2005. ISBN 3-540-24523-5. Makagonov P, Ruiz Figueroa A. Berlin, Germany.
- **Learning a Domain Ontology from Hierarchically Structured Texts.** Proceedings of the Workshop on Learning and Extending Lexical Ontologies by using Machine Learning Methods. ICML2005. August 2005. Makagonov P, Ruiz Figueroa A. Bonn, Germany.
- **Studying Evolution of a Branch of Knowledge by Constructing and Analyzing Its Ontology.** In: Natural Language Processing and Information Systems. Lecture Notes in Computer Science. LNCS 3999. Springer 2006, ISBN 3-540-34616-3. Makagonov P, Ruiz Figueroa A. Alexander Gelbukh. Klagenfurt, Austria.
- **Modelos de desarrollo del hardware y software basado en el estudio de Computación Paralela.** Revista Interciencia, vol. 32 num. 3. ISSN 0378-1844 Marzo 2007. Ruiz Figueroa A, Makagonov P. Caracas Venezuela.

Que se presentaron en congresos, talleres y seminarios nacionales e internacionales:

- **9th Iberoamerican Congress on Pattern Recognition. CIARP 2004.** Instituto Nacional de Astrofísica Óptica y Electrónica. 26 de Octubre del 2004. Puebla, Puebla.
- **6th International Conference on Intelligent Text Processing and Computational Linguistics. CICLing-2005.** Centro de Investigaciones en Computación. Instituto Politécnico Nacional. 13 de febrero 2005. México D. F.
- **The 22nd International Conference on Machine Learning.** Workshop on Learning and Extending Lexical Ontologies by using Machine Learning Methods. ICML2005. August 2005, Bonn, Germany.
- **Seminario Institucional.** Universidad Tecnológica de la Mixteca. 1 de Septiembre de 2005.
- **11th International Conference on Applications of Natural Language to Information Systems.** May 31, Klagenfurt, Austria.