



**INSTITUTO POLITÉCNICO NACIONAL  
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**

No. 24    Serie: VERDE    Fecha: Julio 99

**Proyectos y Temas de Tesis  
en Computación**

**Adolfo Guzmán Arenas <sup>1</sup>**

**RESUMEN**

Se describen varios trabajos en las áreas de computación, informática, software, comunicaciones, que podrían servir, según su nivel de dificultad, como trabajos finales de cursos, trabajos de laboratorio, trabajos de investigación, tesinas, tesis de licenciatura, tesis de maestría, y quizá (algunos cuantos) tesis de doctorado. El autor está en la mejor disposición de proporcionar mayor información o descripción de algunos de estos trabajos propuestos, y de dirigir unos cuantos de ellos.

Antecedentes. En 1980-85, el autor publicó un informe técnico parecido. Esta nueva versión es muy distinta de la anterior, dados los avances de nuestra disciplina.

**Palabras Clave:** Investigación original, proyectos, temas de tesis, tesinas, tesinas de doctorado, tesinas de maestría, trabajos de investigación.

<sup>1</sup>Centro de Investigación en Computación del I.P.N.  
aguzman@pollux.cenac.ipn.mx

Copyright © 1999  
Instituto Politécnico Nacional  
Centro de Investigación en Computación  
Av. Juan de Dios Bátiz, s/n, s/n  
México, D.F. México, D.F.  
México 07380-00  
ISBN 970-18-3334-1

## ADVERTENCIA

*“Este reporte contiene información desarrollada por el Centro de Investigación en Computación del Instituto Politécnico Nacional a partir de datos y documentos con derechos de propiedad y por lo tanto su uso queda restringido a las aplicaciones que explícitamente se convenga.*

*La aplicación no convenida exime al Centro de su responsabilidad técnica y da lugar a las consecuencias legales que para tal efecto se determinen.*

*Información adicional sobre este reporte podrá obtenerse recurriendo a la Unidad de Publicaciones y Reportes Técnicos del centro de Investigación en Computación del I.P.N. Av. Juan de Dios Bátiz s/n, teléfono 729-60-00 ext. 46103, 56608 y 56610”.*

## Proyectos y temas de tesis en Computación

Adolfo Guzmán Arenas  
Centro de Investigación en Computación del I. P. N.<sup>1</sup>  
aguzman@pollux.cic.ipn.mx

Fecha de modificación: 22 de febrero de 1999.

**RESUMEN.** Se describen varios trabajos en las áreas de computación, informática, software, comunicaciones, que podrían servir, según su nivel de dificultad, como trabajos finales de cursos, trabajos de laboratorio, trabajos de investigación, tesinas, tesis de licenciatura, tesis de maestría, y quizá (algunos cuantos) tesis de doctorado. El autor está en la mejor disposición de proporcionar mayor información o descripción de algunos de estos trabajos propuestos, y de dirigir unos cuantos de ellos.

**Antecedentes.** En 1980-85, el autor publicó un informe técnico parecido.<sup>2</sup> Esta nueva versión es muy distinta de la anterior, dados los avances de nuestra disciplina.

### 1. Clasitex. Análisis de documentos escritos en español para encontrar sus temas principales

- OTROS NOMBRES (títulos alternos).** Clasificador automático de textos según su contenido.
- ÁREA (dentro de la Computación).** Inteligencia Artificial, Procesamiento de lenguaje natural, procesamiento de textos en español.
- DESCRIPCIÓN.** En qué consiste el trabajo. Continúe desarrollando el árbol de conceptos que permite que Clasitex encuentre los temas principales a los que se refiere un artículo escrito en español. Haga el desmembrador de cada palabra en raíz, prefijo, sufijo. Haga el propagador de votos. Haga el módulo de votación a la francesa o de dos vueltas. Desarrolle las otras sugerencias que aparecen en mi artículo de Clasitex en el *Journal of Expert Systems*, enero 1998. Haga un programa en C que utilice *dispersión al azar* o hashing para encontrar rápidamente los conceptos en el árbol de conceptos. Pero sobre todo, *agregue unos cincuenta mil nodos* al árbol de conceptos.
- ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Búsqueda inteligente de texto. Se buscan los conceptos, no las palabras. Un artículo que habla de defensa delantero, portero, gol, tiro de esquina, será correctamente clasificado como en "fútbol soccer", aún cuando este par de palabras no aparezca en el artículo.

<sup>1</sup> Centro de Investigación en Computación, Instituto Politécnico Nacional. Av. Juan de Dios Bátiz esq. Con Miguel Othón de Mendizabal Ote. Unidad Profesional "López Mateos" del I. P. N., Col. Zacatenco. Edificio C. I. C. Dirección postal: apartado 75-476. 07738 México, D. F. Teléfonos: +52 (5) 586 2990, 729 6155. Facsímil: 586 2936. Correo electrónico: aguzman@pollux.cic.ipn.mx, aguzman@amiac.org.mx. HTTP: //www.cic.ipn.mx

<sup>2</sup> "Ciento veinticinco proyectos y temas de tesis en Computación." Reporte AHR 80 7, IIMAS UNAM. 1980. Segunda edición: Reporte AHR 80 7a., 1981, IIMAS. 3a. edición: Reporte AHR 81 19, 1981, IIMAS. 4a. edición: en el CINVESTAV, IPN.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Use el programa Clasitex ó el Clasitex++ como punto de partida.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

- 98.<sup>3</sup> Guzmán A. Hallando los temas principales en un artículo en español. *Soluciones Avanzadas*. Vol. 5, núm. 45, pág. 58. I parte, 15 de Julio de 1997, II parte vol. 5, núm. 49, pág. 66, 15 de septiembre de 1997. También en: *Simposium Internacional de Computación*. Centro de Investigación en Computación. Instituto Politécnico Nacional. Noviembre 12-14, 1997. México, D.F., páginas 36-51.
99. Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.
101. Beatriz Beltrán Martínez, Adolfo Guzmán Arenas, Francisco Martínez Trinidad, José Ruiz Shulcloper. Clasitex++: una herramienta para el análisis de textos. Memorias del *Tercer Taller Iberoamericano de Reconocimiento de Patronos*, TIARP-98, Centro de Investigación en Computación, Instituto Politécnico Nacional, México, D. F. Marzo 1998. Páginas 369-379

## 2. Agentes móviles

OTROS NOMBRES (títulos alternos). Ayudante del Inyector de Agentes.

ÁREA (dentro de la Computación). Inteligencia Artificial, Agentes, Computación distribuida.

DESCRIPCIÓN. En qué consiste el trabajo. Se inyecta de alguna manera en una máquina huésped, un agente que lleva a cabo una tarea específica (por ejemplo, la sumarización de datos que lleva a cabo el Sumarizador de *Anasin*). Por algún motivo, el agente detecta una de las siguientes situaciones:

- a) los datos han cambiado de morada, ya se mudaron a otro sub-directorio o a otra máquina, por lo que el agente debe emigrar.
- b) la máquina huésped presenta paros frecuentes (está enferma) lo que amenaza con que el agente se muera y no resucite. Es necesario *eternizar* al agente (hacerlo inmortal, o sea, resistente a fallas o caídas de la máquina huésped).
- c) parte de los datos se han ido, el archivo se ha duplicado o replicado, por lo que es conveniente que el agente haga una *mitosis* y se divida en dos (o más), y sus clones emigren y persigan a los nuevos datos.

Diseñe e implemente usted un sistema donde haya agentes móviles capaces de ser inyectados, de detectar las condiciones (a) a (c) (y algunas otras), y de responder a ellas con las características de migración, mitosis, inmortalidad, etc., señaladas.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Para definir procesos distribuidos débilmente acoplados que sean más robustos, adaptables a los cambios que inevitablemente ocurren en un ambiente de explotación de información.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Quizá el *Inyector de Agentes* del Ing. Gilberto Martínez Luna (Lab. de Sistemas de Información, C. I. C.) sea un punto de partida.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

---

<sup>3</sup> Estos números se refieren a la lista de trabajos de A. Guzmán que aparece en su curriculum vitae.

- 1). Adolfo Guzmán. Sumarizador-Transmisor de *Anasin*. SoftwarePro International, Austin, Texas.
- 2). Gilberto Martínez Luna. Inyector de agentes. Tesis de M. en C., Depto. de Ing. Eléctrica, CINVESTAV-I. P. N., 1998.
- 3). 94. Guzmán A. Tecnología de Agentes. *Simposium Internacional de Computación*. Centro de Investigación en Computación, Instituto Politécnico Nacional, Octubre 21-23, 1996, México, D.F., páginas 25-36. También en: *I. P. N. Ciencia, Arte y Cultura*, vol. 2, número 10, Nov-Dic. 1996, páginas 3-10.

### 3. Clasificadores tipo árboles k-d

OTROS NOMBRES (títulos alternos). Clasificación supervisada usando árboles de decisiones. ÁREA (dentro de la Computación). Inteligencia Artificial, Reconocimiento de patrones, clasificación supervisada.

DESCRIPCIÓN. En qué consiste el trabajo. Se puede hacer un clasificador que aprenda de una matriz de aprendizaje y cuyo resultado sea un programa en C, un árbol de decisiones (If's anidados, árbol k-d) que es el clasificador apropiado para lo que se aprendió. Aquí hay varios trabajos relacionados:

1. Aprenda de varias matrices de aprendizaje, y de alguna manera combine los clasificadores (árboles k-d) obtenidos en cada aprendizaje, para obtener un clasificador (árbol) óptimo, es decir, la mejor combinación de los árboles individuales.
2. Utilice árboles balanceados («b-trees») en vez de árboles k-d para tales clasificadores.
3. Utilice archivos indexados o funciones *hash* (de dispersión al azar) en vez de árboles k-d, para la implementación de tales clasificadores. El uso de archivos hará lento el sistema, pero hará posible la clasificación con clasificadores enormes (por ejemplo, el clasificador ALVOT requiere todos los datos de la matriz de aprendizaje [MA] a tiempo de ejecución; si la MA es muy grande, habrá que guardarla en disco).
4. Utilice strip-trees (árboles rebanados) en vez de árboles k-d.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Acoplar las herramientas modernas de tecnología de software a la ciencia de Reconocimiento de Patrones.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Para cada una de las proposiciones (1) a (4), habrá que usar métodos de codificación distintos, creo yo. Es demasiado pedir un solo algoritmo unificado, tal vez no sea deseable o posible.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

A. Guzmán. Generador de árboles k-d. SoftwarePro International, 1993.

89.<sup>4</sup> Guzmán, A. Árboles k-d como clasificadores supervisados y para la sustitución de sistemas expertos. Congreso Internacional sobre Reconocimiento de Patrones. *ICI-MAF*, Habana, Cuba. 1995. (Este congreso fue el TIARP I, o Taller Ibero americano sobre Reconocimiento de Patrones).

100. Ruiz Shulcloper, José; Guzmán, A. y Díaz de León, Juan Luis. Enfoque lógico combinatorio al Reconocimiento de Patrones: Clasificación Supervisada. Libro a ser editado por Editorial Politécnica, 1998.

<sup>4</sup> Estos números se refieren a la lista de trabajos de A. Guzmán que aparece en su curriculum vitae.

A. Guzmán. Exploración de distintos árboles para clasificación supervisada. Proyecto DE-PI-I. P. N., 1997-98.

Ricardo Avilés. Tesis de Ingeniero en Comunicaciones y Electrónica, ESIME-I. P. N., 1978? Utiliza árboles rebanados para clasificación rápida, pero los genera a mano.

#### 4. Generador de clasificadores

OTROS NOMBRES (TÍTULOS ALTERNOS). Búsqueda automática de clasificadores que tengan una eficiencia dada.

ÁREA (dentro de la Computación). Inteligencia artificial, reconocimiento de patrones, clasificación supervisada.

DESCRIPCIÓN. En qué consiste el trabajo. Utilice algún método para obtener un clasificador supervisado, el cual depende de ciertos parámetros: qué testores típicos se usaron, cuál es la *granularidad* o precisión de las variables (por ejemplo, una variable continua, como *peso en kilogramos*, puede engrosarse o convertirse a una variable con los intervalos [de 0 a 10 kg.], [de 11 a 20 kg.], etc.) Utilícese una matriz de control para obtener la eficiencia (porcentaje de aciertos) de tal clasificador. Ahora bien, supongamos que la eficiencia obtenida no es la adecuada. Hay que buscar un clasificador mejor (con mayor eficiencia).

El trabajo consiste en generar de manera automática muchos clasificadores, buscando uno (o varios) que tengan eficiencia alta. El generador modificará los parámetros (granularidad de las variables, escoger otros testores típicos, eliminar objetos de la matriz de aprendizaje, ...) de suerte que se vayan generando clasificadores con eficiencia (quizá) mayor. El método obvio es el de fuerza bruta, pero es muy lento (por ser exhaustivo y estar trabajando con una gran cantidad de variaciones de los parámetros). La idea es guiarse por la eficiencia, por la cantidad de información de los rasgos, por la *confusión* entre objetos, etc., para guiar el método de fuerza bruta hacia uno de "fuerza semi-bruta", que proponga parámetros razonables y encuentre mejores clasificadores sin tanta exploración.

Adicionalmente, pueden combinarse los 3 (o n) mejores clasificadores, para obtener un clasificador aún mejor que los clasificadores que se combinan. *Idea*: utilice la *sensibilidad* (precios sombra) de cada parámetro del clasificador, para saber en qué sentido modificar el parámetro k.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.

Es engorroso estar buscando manualmente un clasificador mejor que el que ya se tiene. A menudo este trabajo se hace por una persona que es especialista en el campo de aplicación (es especialista en préstamos), pero no es especialista en Reconocimiento de Patrones. El método manual o del tanteo no garantiza encontrar rápido un clasificador deseado.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Desarrolle un algoritmo de fuerza bruta para generar clasificadores, y simultáneamente estudie (teóricamente) los efectos de los diferentes parámetros del clasificador en su eficiencia. Consecuentemente, agregue heurísticos a su generador bruto para guiar la generación inteligentemente. *Bono adicional*: demuestre que su generador es monotónico (la eficiencia de un nuevo clasificador generado, con respecto al anterior, aumenta siempre o no disminuye).



Para el algoritmo de fuerza bruta, el clasificador a usar será uno que (a) sea el más robusto o eficiente posible, y (b) se conozca bastante bien su fundamento teórico, a fin de entender el papel de cada parámetro del clasificador. Por ejemplo, use ALVOT. *Bono adicional*: tenga una tabla que diga: para tales y cuales condiciones, el mejor clasificador a usar, *a priori*, es la recta de Fisher, para esto otro hay que usar diagramas de Voronoi, etc.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

100. Ruiz Shulclóper, José; Guzmán, A. y Díaz de León, Juan Luis. Enfoque lógico combinatorio al Reconocimiento de Patrones: Clasificación Supervisada. Libro a ser editado por Editorial Politécnica, 1998.

A. Guzmán. Generación de clasificadores con una eficiencia dada. Trabajo en progreso, incompleto. SoftwarePro International. (tengo un borrador de este artículo)

## 5. Transformador de expresiones en SQL para minería de datos

OTROS NOMBRES (títulos alternos). Virtualizador del cubo de minería de datos.

ÁREA (dentro de la Computación). Sistemas de información, minería de datos.

DESCRIPCIÓN. En qué consiste el trabajo. Un minero utiliza expresiones en SQL (o en un lenguaje especial, según el trabajo # 18, convertible en expresiones en SQL) para acceder un cubo de datos. El cubo de datos es isomorfo a la base de datos original (contiene la misma información, guardada en una forma especial que facilita la búsqueda al minero). Cuando se utiliza un cubo real, se duplica la información: se tiene en la base de datos original, y se tiene en el cubo.

El trabajo consiste en eliminar el cubo de datos, pero haciendo creer al minero que aún existe, que aún continúa accediéndolo. El truco es convertir las expresiones en SQL que el minero usa para acceder al cubo, en otra expresión equivalente en SQL, pero que ahora accesa *la base de datos real* u original. Es decir, la expresión original en SQL accesa al cubo y obtiene los mismos datos que la expresión *nueva* en SQL obtendrá al acceder la base de datos original. Por consiguiente, si “engañamos” al minero y le cambiamos su expresión en SQL por la *nueva expresión en SQL*, y usamos ésta para acceder la base de datos original, se obtendrán los mismos resultados, por lo que el cubo de datos ya no nos es útil y puede eliminarse, lo que ahorra espacio en disco.

Usted hará el traductor de SQL a nuevo SQL. Este traductor estará parametrizado por la definición de la base de datos original, y por el árbol que nos dice quiénes son los hijos geográficos de Jalisco, los hijos temporales de 1998, etcétera (ver trabajo # 18).

Variante: no haga un traductor, haga un intérprete que *atrape* la expresión original en SQL que el minero usa para acceder al cubo, y la *interprete* y entienda y haga los accesos adecuados a la base de datos original, arrojando la misma información que se obtendría si el cubo (ahora inexistente) se hubiese accedido.

*Bono adicional*: ¿Cómo se manejarán los campos (totalizadores) que no existen, que son la mayoría de los totalizadores? Ahora que el cubo es real, tenemos la opción de *cachear* un totalizador cuyo valor era desconocido y acaba de ser computado, guardándolo en el cubo real (valor obtenido bajo demanda, o evaluación perezosa), o desechar tal valor, teniendo que recomputarlo después si se vuelve a necesitar. La rapidez de procesamiento sugiere que los totalizadores se vayan cacheando para evitar recomputarlos, pero ¿dónde

guardarlos? ¡No me atrevo a sugerir que se guarden en la base de datos real, la que está en producción!

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. La minería de datos es un campo reciente, muy productivo, con productos comerciales primitivos que sin embargo han alcanzado cierto éxito (y mucho ruido). Sirven para descubrir anomalías, situaciones interesantes, desviaciones y tendencias en un mar de datos, de manera automática. Apoyan la toma de decisiones tácticas y estratégicas en los mandos medios y superiores de una empresa.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Utilice técnicas de expansión de macros en texto ASCII (después de todo, eso es lo que es una expresión en SQL).

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

92. Guzmán A. *Mineros de Datos*. En *Soluciones Avanzadas*, México D.F., 1996. También en: *ARCHIPIÉLAGO*. Vol. 2 núm. 9, página 19. Nov.-Dic. 1996.

97. Guzmán A. Estado del Arte y de la Práctica en Minería de Datos, Análisis y Crítica. *II Taller Iberoamericano de Reconocimiento de Patrones*. Marzo 24-28, 1997. La Habana Cuba, páginas 367-376.

## 6. Eliminador Universal de Baches (E U B)

**OTROS NOMBRES (TÍTULOS ALTERNOS).** Rueda que se adapta a las irregularidades de la carretera.

**ÁREA (dentro de la Computación).** Inteligencia Artificial, Visión, Control en tiempo real.

**DESCRIPCIÓN.** En qué consiste el trabajo. Diseñar el hardware y software para el EUB.

Una cámara de imágenes, conectada a una computadora, montada en la parte delantera de un automóvil, va detectando los baches (agujeros) y topes (agujeros negativos) que se encuentran en el camino adelante. Las ruedas están formadas por rayos (como las de las bicicletas, aunque no tantos rayos) y cada rayo termina en un pedazo de hule, que es el que hace contacto con el terreno. Es decir, la circunferencia de la rueda no es una sola pieza, sino que está formado de quizá doce rayos que terminan en pedazos de hule (como si fueran zapatos), cada uno al final de un rayo. Cuando cada rayo de la rueda tiene una misma longitud (radio), los doce pedazos de hule forman una circunferencia.

Ahora bien, los rayos pueden cambiar su longitud, controlados por la computadora. El objeto del EUB es (1) detectar los baches positivos y negativos sobre los que pronto andará el automóvil; (2) calcular qué rueda y qué rayos harán contacto con el bache, y (3) alargar o acortar los rayos correspondientes, a fin de que el bache no se sienta. Es un sistema que trabaja en tiempo real.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad.

La ventaja del EUB es que permite que un automóvil no "sienta" los baches, porque los detecta y modifica la superficie de contacto, adaptándola a la superficie de la carretera, de tal manera que el vehículo avance a una altura constante sobre la superficie, aunque ésta sea irregular. Es un amortiguador automático adaptivo en tiempo real.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.



## 7. Sección Amarilla electrónica

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Sistemas de Información, Procesamiento de lenguaje natural, Procesamiento de textos en español.

DESCRIPCIÓN. En qué consiste el trabajo. Usando un árbol de conceptos como el que usa CYC ó Clasitex, construir una sección amarilla electrónica (podría venderse en un CD rom) que supiera sinónimos, padres e hijos de un concepto, y otras relaciones sobre el árbol de conceptos. Por ejemplo, una persona está buscando "consultores", pero la sección amarilla los tiene almacenados como "asesores" o "consejeros". Por ejemplo, buscamos "bisagras de 3 pulgadas" y encontramos que hay una fábrica de bisagras (pero no dice explícitamente que sean de 3 pulgadas), y hay ferreterías, y hay tiendas para carpinteros. El sistema debe ser capaz de señalar "lo más cercano" a lo que se pidió, midiendo la distancia en el árbol de conceptos.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.

Mejoraría la sección amarilla actual, ya que uno pierde tiempo buscando "autobuses de pasajeros" y está guardada la información como "transportes foráneos de personas".

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Use el árbol de conceptos de CYC ó Clasitex.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

98. Guzmán A. Hallando los temas principales en un artículo en español. *Soluciones Avanzadas*. Vol. 5, núm. 45, pág. 58. I parte, 15 de Julio de 1997, II parte vol. 5, núm. 49, pág. 66, 15 de septiembre de 1997. También en : *Simposium Internacional de Computación*. Centro de Investigación en Computación. Instituto Politécnico Nacional. Noviembre 12-14, 1997. México, D.F., páginas 36-51.

99. Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.

## 8. Sistema para manejo de llegadas y salidas de viajeros en los aeropuertos internacionales

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Sistemas de Información.

DESCRIPCIÓN. En qué consiste el trabajo. En México, los aeropuertos reciben y dejan salir a las personas (el Instituto Nacional de Migración, para ser más exactos) de/a otros países en forma manual, llenando unas hojitas y verificando los documentos. La propuesta es captar esta información en la computadora, por el agente de migración, al momento en que tiene al viajero frente a él. En ese momento la computadora accesa una base de datos para verificar su estado, si no se le busca, etc. Nota 1: las aerolíneas ya tienen en su base de datos los nombres de los pasajeros que están llegando o van a salir, por lo que esta información se puede usar para "pre-cargar" la información y hacer más rápida la búsqueda (aunque hay que tomar en cuenta que los nombres que aparecen en los boletos de avión no siempre coinciden con los nombres del pasaporte).

Nota 2: Muchos países tienen ya sus pasaportes con código de barra y caracteres OCR (fáciles de leer por la computadora).

Extensión: a buques, puentes (paso con automóviles, paso a pie).

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Modernizar el manejo de visitantes internacionales en aeropuertos y puertos de entrada.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Use una base de datos local con acceso rápido, probablemente en modo cliente-servidor y no a través de páginas web, y acceso remoto a otras bases de datos exógenas.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Los países avanzados cuentan ya con este servicio.

## 9. Servicios (inscripción, etc.) de estudiantes vía Intranet

**OTROS NOMBRES (títulos alternos).**

**ÁREA (dentro de la Computación).** Sistemas de Información, Intranet.

**DESCRIPCIÓN.** En qué consiste el trabajo. Muchos trámites de control escolar (inscripciones, cambio de grupo, darse de baja, consulta de calificaciones, etc.) y administrativos se realizan con papel, y en forma presencial. El estudiante debe acudir a una ventanilla, llenar un papel, firmar, etc. Diseñar un sistema donde estos trámites se puedan hacer vía Intranet/Internet, de modo que los estudiantes que tienen acceso a ella no tengan que ir necesariamente a la escuela a hacer sus trámites. (También pueden hacer los trámites en la escuela, donde probablemente los dirijan hacia una terminal con Intranet).

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Modernizar los servicios a estudiantes. Evitar viajes a la escuela para realizar trámites.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Haga un experimento con un centro de enseñanza pequeño.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Muchas universidades modernas cuentan ya con este servicio. En el C. I. C., el M. en C. César Guzmán está trabajando en estas líneas.

Sistema para actualización de datos de contaminación, seguridad y protección industrial.

Proyecto del C. I. C. patrocinado por Pemex-Refinación, ASIPA, 1997.

## 10. Centro de acopio de donaciones

**OTROS NOMBRES (títulos alternos).**

**ÁREA (dentro de la Computación).** Sistemas de Información.

**DESCRIPCIÓN.** En qué consiste el trabajo. Hay muchas donaciones de ropa, medicina, cobijas, baterías, ..., para distintos damnificados. Hay, por otro lado, desastres. Y hay damnificados. Se trata de hacer la mejor distribución posible de las donaciones, a través de un sistema de programas que usted diseñará. Una de las finalidades es determinar o demostrar que tal donación se entregó o regaló a tal damnificado. El problema se parece al de asignación de personas (con ciertas habilidades y destrezas) a puestos de trabajo, o el de encontrar parejas de hombre-mujer, donde cada uno declara ciertas preferencias (busco muchacha de 25 a 35 años, católica, blanca, ....)

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Mejor administración y entrega de donaciones.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Programa para encontrar parejas hombre-mujer. Programa para asignar trabajadores a puestos de trabajo. Programa para asignar trabajos a distintas máquinas en un taller.

### 11. Servidor de imágenes de personas extraviadas; buscadas; de objetos robados; recuperados, via Intranet

**OTROS NOMBRES (títulos alternos).**

**ÁREA (dentro de la Computación).** Sistemas de Información, Manejo de imágenes, Intranet.

**DESCRIPCIÓN.** En qué consiste el trabajo. Una gran cantidad de objetos robados han sido recuperados y esperan a que su dueño los identifique y se los lleve. Las personas tienen que viajar actualmente a donde se encuentran los objetos. La idea es que el sistema que usted va a diseñar ponga a disposición de las personas interesadas, en distintos lugares, vía Internet, una imagen (o varias) de los objetos que se pueden reclamar. Lo mismo para fotos y descripciones de personas extraviadas.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Mejor servicio a la ciudadanía.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Utilice una base de datos que pueda almacenar imágenes.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

Sistema integral de información que muestra a través de Intranet instalaciones, funcionarios, y otras imágenes. Proyecto del C. I. C. patrocinado por Pemex-Refinación. 1997.

### 12. Purificador para detectar y resolver problemas de software sobre el año 2,000

**OTROS NOMBRES (títulos alternos).**

**ÁREA (dentro de la Computación).** Tecnología de Software.

**DESCRIPCIÓN.** En qué consiste el trabajo. Haga un programa que analice programas de Cobol, Basic, Fortran (lenguajes viejos) buscando declaraciones y usos de variables, constantes y campos que representen años y tengan 2 caracteres de precisión. Su programa pintará de verde los postulados fuente que cumplan con ese criterio, utilizando heurísticos tales como: declaraciones o nombres tipo DD MM AA, YY MM DD, variables con nombres FECHA, FECHA-DE-..., FECH, TIEMPO, etc. Analizando los comentarios para tratar de dilucidar si tal variable o campo se refiere a una fecha de dos dígitos. Y usando otros criterios que a usted se le ocurran.

Restricción: escoja un solo lenguaje, y restrinjase a él. No lo haga (a menos que usted prefiera) multilingüe.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Adecuar los programas para eliminarles fallas al cambio de siglo. Ayudar en tal adecuación. Nota: apúrese. Ya no servirá su solución si la termina después del año 2,000.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Haga un experimento manual analizando con la vista varios programas de Cobol (o de algún otro lenguaje) que padezcan del “año de dos dígitos.”

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

### 13. Diseño de marcas electrónicas en imágenes electrónicas

**OTROS NOMBRES** (títulos alternos).

**ÁREA** (dentro de la Computación). Procesamiento de Imágenes.

**DESCRIPCIÓN.** En qué consiste el trabajo. Las imágenes en papel se protegen utilizando marcas de agua, que son dibujos tenues que se sobreponen (o están contenidos o pre-impresos en el papel) sobre el dibujo original, de tal manera que no se noten o se noten poco. De esta forma es posible demostrar que tal dibujo fue sacado de la Biblioteca del Vaticano, pues tiene su marca de agua. Como no se nota mucho, es posible que pase desapercibido por el “infractor.” Las marcas de agua normalmente resisten las operaciones normales que se realizan con imágenes, como copiado, amplificación, reducción, mostrar una parte de la imagen, distorsiones comunes, reflexión en espejo.

Usted va a diseñar un sistema que ponga marcas electrónicas sobre imágenes electrónicas (o sea, que se muestran en una pantalla; también se pueden imprimir) con la misma finalidad de las marcas de agua normales. Piense qué marcas usará. Diseñe el sistema. Describa el método de autenticación de una imagen de usted que Juan Pérez alega que es de él.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Para identificar el dueño u origen de un documento.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

(falta aquí una referencia).

### 14. Prueba de programas mediante inyectores de errores

**OTROS NOMBRES** (títulos alternos).

**ÁREA** (dentro de la Computación). Tecnología de software.

**DESCRIPCIÓN.** En qué consiste el trabajo. El efecto de un error en el código puede medirse o deducirse mediante la introducción artificial de ciertos errores. Por ejemplo, puedo introducir un error en la variable X, añadiendo una línea (inyectando el error) que diga X := random (X). Describa un método (no diseñe ni haga programas, solo el método) para utilizar estas técnicas de inyección de errores. Describa su método, dé un ejemplo. ¿Cómo se compara contra métodos ya existentes (ver referencia)? ¿Cómo los mejora o supera? ¿Qué se gana con su método? ¿Y qué se gana con su método versus con otros métodos ya existentes?

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad.

En la etapa de prueba, para determinar la robustez de programas.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.  
TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

*Computer Dec. 97.*

### 15. Compiladores para paralelizar varias instrucciones

OTROS NOMBRES (títulos alternos). Instruction level parallelism.

ÁREA (dentro de la Computación). Arquitectura de computadoras, Tecnología de Software.

DESCRIPCIÓN. En qué consiste el trabajo. Tenemos un programa en ensamblador (o en binario) y tenemos un procesador que tiene una ALU (unidad aritmética y lógica) capaz de ejecutar en paralelo algunas instrucciones, amén de la paralelización que se obtiene en el canal (bus) de memoria en el ciclo de obtención (fetch), el de cómputo de la dirección efectiva, el de ejecución (escritura o lectura de datos a o desde memoria), etc. Podemos suponer, por ejemplo, que la computadora tiene una palabra larga (60 bits, como la CDC Cyber, R.I.P.), cuyas instrucciones podrían ejecutarse en paralelo. Diseñe un compilador que detecte las paralelizaciones posibles y las efectúe, ensamblando las instrucciones correspondientes, y estableciendo candados para cuando cierta paralelización no sea posible.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Aumentar la velocidad de ejecución.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

*Computer, Dec. 97, p. 63 ff.*

### 16. Tramitel electrónico: informes de trámites en español, por teléfono

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Síntesis de voz.

DESCRIPCIÓN. En qué consiste el trabajo. Existe una oficina que da información telefónica, mediante personas, de trámites en oficinas de gobierno. La idea es representar a cada trámite por un flujograma (red dirigida, los nodos son transacciones atómicas o trabajos que realiza un individuo, los arcos indican los nodos o transacciones "corriente abajo" que son consecuencia de las transacciones que ya concluyeron; hay nodos condicionales o "if": si es menor de edad, vaya a la ventanilla 1, si no, vaya primero a la caja "antecedentes penales"), y tener un "sintetizador de voz" que va "leyendo el trámite" por teléfono.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Dar información veraz por teléfono.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Usar tarjetas que se conectan a la PC para manejar (hablar) menús de voz, y leer (mediante tonos) los dígitos pulsados en el microteléfono.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

86bis. Guzmán, A. *Diseño de una Sistema General de Seguimientos*. Simposium Internacional de Computación, CENAC-IPN, Nov. 10-13, 1993, México, D.F.



Cecilia Palomino. Construcción de un sistema de flujo de documentos (work flow) con múltiples servidores. Tesis de M. en C., Centro de Investigación en Computación, 1998. Trabajo en progreso.

## 17. Búsquedas en bases de datos con llaves inexactas

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Reconocimiento de Patrones, Sistemas de Información.

DESCRIPCIÓN. En qué consiste el trabajo. Un "objeto" (persona) o situación ("conflicto entre campesinos") se describe por varios atributos o rasgos (sueldo, edad, sexo, colonia donde vive, altura, peso, religión ...), que poseen valores numéricos o simbólicos (católico, mahometano, budista, ...). [para simplificar, un objeto de éstos (no es un objeto de C++) puede representarse como un registro de un archivo o un renglón de una tabla de una base de datos]. Un método de hallar si un objeto dado está en un archivo grande o colección grande de objetos es calcular, a partir de sus atributos, una llave o índice (si se usan archivos, se denominan archivos indexados. Si se usan tablas en memoria, se llaman llaves, y se usa el método de dispersión al azar o *hashing*). Dado un objeto "nuevo", se desea averiguar si está o no en tal archivo. Entonces se computa, a partir de los atributos de tal objeto, una llave, y con ella se *indexa* el archivo. Si el índice encuentra registro, sí está el objeto en el archivo, y el registro puede contener información adicional sobre el objeto: su dirección, cuánto nos debe de renta, o a qué *clase* pertenece: beisbolista, nadador, levantador de pesas, corredor de fondo, etc. Si el índice es *inválido*, significa que no está en el archivo, el archivo no contiene al objeto nuevo, o no contiene un objeto con las características del nuevo. Se puede entonces tomar algún tipo de decisión: decir que no está, incorporarlo, etc.

Ahora bien, pudiera ser que los atributos sean más o menos difíciles de computar. Por ejemplo, si estamos mirando caras, un atributo puede ser "tamaño de los ojos" con valores chico, mediano, grande. Pero pudiera haber confusión o inexactitud, de manera que medimos "chico" cuando en realidad debía de ser "mediano". Se habla de llaves inexactas. La búsqueda con valores inexactos o imprecisos puede proceder así: se miden (lo mejor que se pueda) los atributos del objeto nuevo. Se computa la llave para tal objeto. Se mete la mano al archivo (se *indexa* con tal llave). Si se halla un objeto, *hemos hallado un objeto que casa perfectamente* o con cero error con el objeto nuevo. Si no hallamos objeto alguno, entonces "no se encuentra un objeto con error = 0". Se procede a calcular todas las llaves que distan 1 [o sea, que tienen error = 1] de la llave anterior (por ejemplo, la llave puede ser llave **a** = {nariz chata, ojo chico}, donde los valores para nariz son chata, regular, afilada, aguileña, y para ojo son chico, mediano, grande. Entonces una llave que dista 1 de la llave **a** es {nariz regular, ojo chico}, otra llave que también dista 1 de la llave **a** es {nariz chata, ojo grande}, una llave que dista 2 de la llave **a** es {nariz regular, ojo grande}, etc. Hállense todas las llaves con error = 1 y los objetos (accesando varias veces el archivo) con error = 1, y así también los objetos con error = 2 y quizá los que tienen error = 3, y preséntense en forma ordenada (primero los de error más bajo), como "objetos encontrados que se parecen al objeto nuevo buscado".

Por alguna razón, la distancia o error en un atributo (forma de nariz) puede ser más grande (más importante) que en otro atributo.

También, los valores de algún atributo pudieran formar árbol. Por ejemplo, pudiera ser fácil detectar con robustez que una nariz es grande, afilada o chata. Ahora bien, las narices grandes se dividen en triangulares, rectangulares y pentagonales. Las narices afiladas se dividen en rectas, cóncavas y convexas o aguileñas. Las narices chatas se dividen en chata-chica y chata-aplastada. Ahora bien, dado que un objeto tiene {nariz triangular, ojo chico}, cuando tal objeto no se halle en el archivo, se puede tratar de buscar {nariz grande, ojo chico} en vez de {nariz rectangular, ojo chico} + {nariz pentagonal, ojo chico}. Es decir, usar al "papá" del atributo en vez de a los "hermanos" del atributo. De esta manera es posible "generalizar" un objeto, al generalizar (subir en el árbol) uno o más de sus atributos. La generalización es una operación que nos protege de la imperfección al medir o computar atributos. El programa que trata de acceder un "objeto generalizado" (aquél que posee un atributo no expresado al máximo detalle, como hoja del árbol) debe poder computar las "llaves específicas" derivadas de la llave general, y tratar de acceder tales objetos específicos. Por ejemplo, el objeto (generalizado) {nariz grande, ojo chico} se particulariza en (representa a) los objetos {nariz triangular, ojo chico}, {nariz rectangular, ojo chico}, {nariz pentagonal, ojo chico}. La búsqueda del objeto {nariz grande, ojo chico} se lleva a cabo mediante cuatro accesos al archivo, con las llaves {nariz grande, ojo chico}, {nariz triangular, ojo chico}, {nariz rectangular, ojo chico}, {nariz pentagonal, ojo chico}. El archivo puede contener objetos específicos y objetos generalizados. Idea: que los objetos generalizados en el archivo contengan apuntadores a los objetos menos generalizados que sean sus hijos. Esto evitaría una gran cantidad de accesos inútiles al disco, originadas por llaves inválidas (que representan a objetos inexistentes): cada vez que se guarde un objeto específico en el archivo, compútense los objetos generalizados (o sea, los ascendientes de tal objeto), y regístrense éstos (teniendo cuidado de marcarlos como 'virtuales' o artificiales, ya que en realidad no han sido encontrados en la naturaleza, sino se han insertado en el archivo solo para poder hacer referencia a sus hijos) también en el disco, cuidando también que cada objeto apunte a sus descendientes inmediatos. *Los objetos en el archivo están organizados de acuerdo al árbol de atributos.* Por ende, la búsqueda es inmediata: cuando se tenga un objeto nuevo, mídanse sus atributos, y, ya sea un objeto generalizado o no, úsese su llave para acceder el archivo. Si no existe, *tampoco existen hijos suyos*, y la búsqueda solo puede proceder hacia arriba (generalizando más tal objeto). Si existe, puede ser real (imprimase como "encontrado") o virtual, pero en todo caso léanse de él las llaves de sus hijos inmediatos. Accéense éstos, y repítase el proceso hacia abajo del árbol.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Para comparar objetos complejos (imágenes de rostros, por ejemplo) contra una base de datos enorme (con muchos objetos).

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. El trabajo de comparar objetos complejos puede dividirse en dos partes, la primera (que no vamos a hacer en esta propuesta) es hallar los rasgos descriptivos o atributos del objeto, y medirlos; la segunda parte, contenida en esta propuesta, es *hallarlos rápidamente* utilizando un archivo organizado en árbol, según el árbol que forman los atributos.

Nótese que aunque no se van a desarrollar los identificadores o medidores de atributos, es menester tener una buena idea de los mismos, la que es imprescindible para entender cómo va a estar organizado el árbol de atributos. *Podría ser preferible*, en conse-

cuencia, desarrollar *también* los identificadores de atributos, en paralelo con el archivo arborizado.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

89. Guzmán, A. Árboles k-d como clasificadores supervisados y para la sustitución de sistemas expertos. Congreso Internacional sobre Reconocimiento de Patrones. *ICI-MAF*, Habana, Cuba. 1995. (Este congreso fue el TIARP I, o Taller Ibero americano sobre Reconocimiento de Patrones).

## 18. Lenguaje para expresar “situaciones interesantes” en minería de datos

OTROS NOMBRES (títulos alternos). Lenguaje que describe búsquedas generalizadas.

ÁREA (dentro de la Computación). Sistemas de Información, Minería de Datos

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.

La minería de datos es un campo reciente, muy productivo, con productos comerciales primitivos que sin embargo han alcanzado cierto éxito (y mucho ruido). Sirven para descubrir anomalías, situaciones interesantes, desviaciones y tendencias en un mar de datos, de manera automática. Apoyan la toma de decisiones tácticas y estratégicas en los mandos medios y superiores de una empresa.

Una base de datos se interroga utilizando SQL. Para interrogar el “cubo de datos” que usan los mineros, se ha usado un lenguaje sencillo, que nos permite expresar “brincos parametrizados” (de semana en semana, de mes en mes, de año en año; de tienda en tienda, de ciudad en ciudad, de estado a estado, de nación a nación). También nos debe permitir expresar los siguientes conceptos:

- los hijos geográficos de una entidad geográfica, por ejemplo, los hijos geográficos de Veracruz, que son: Jalapa, Minatitlán, Coahuila, ...
- Mis tíos de producto. Si estamos en el eje de productos, y yo soy “bisagras”, entonces mi papá es “ferretería”, de donde los hermanos de mi papá, o sea mis tíos, son “ropa”, “abarrotes”, “carnes y salchichonería”, etc.
- Los descendientes geográficos de mi papá temporal. Si yo soy “Jalisco, 15 de agosto de 1997”, entonces mi papá temporal es “Jalisco, agosto de 1997”, y los descendientes geográficos buscados son: “Guadalajara, agosto 1997”, “Chapala, agosto 1997”, “Ciudad Guzmán, agosto 1997”, etc.

DESCRIPCIÓN. En qué consiste el trabajo.

La idea es desarrollar completamente el lenguaje, para poder expresar de una manera sencilla, utilizando una notación amigable y entendible, todos o muchos de los conceptos (búsquedas, situaciones) que se usan o sean interesantes para minería de datos. Además del lenguaje, desarrollar su intérprete o su compilador (traductor a SQL + algunas rutinas que no están en SQL). Estos traductores están parametrizados por la estructura del cubo de datos, es decir, por un árbol que dice: Los hijos geográficos de Jalisco son Chapala, Ciudad Guzmán, ...; los hijos temporales de 1998 son Enero 1998, Febrero 1998, ..., Diciembre 1998.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Haga un intérprete del lenguaje. Otra: haga un traductor a expresiones SQL (así lo tenemos implementado ahora).

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

92. Guzmán A. Mineros de Datos. En *Soluciones Avanzadas*, México D.F., 1996. También en: *ARCHIPIÉLAGO*. Vol. 2 núm. 9, página 19. Nov.-Dic. 1996.

97. Guzmán A. Estado del Arte y de la Práctica en Minería de Datos, Análisis y Crítica. *Memorias del II Taller Iberoamericano de Reconocimiento de Patrones*. Marzo 24-28, 1997. La Habana Cuba, páginas 367-376.

## 19. Métodos para evaluar relevancia y utilidad de temas en artículos escritos en español

OTROS NOMBRES (títulos alternos). Sensor de los problemas de una empresa, de una comunidad.

ÁREA (dentro de la Computación). Inteligencia Artificial, Procesamiento de Lenguaje Natural, procesamiento de textos en español.

DESCRIPCIÓN. En qué consiste el trabajo. Mientras más se mencione un tema, es más importante, más relevante, de más actualidad. Utilizando Clasitex (refs. 98 y 99), analice las noticias de los periódicos, o el boletín informativo de la empresa, y determine cuáles son los temas que más acaparan la atención del lector (o de los escritores).

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Manejo inteligente de textos en español. Entender qué preocupa a una comunidad, a una empresa.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos. Ver también trabajo #20.

98. Guzmán A. Hallando los temas principales en un artículo en español. *Soluciones Avanzadas*. Vol. 5, núm. 45, pág. 58. I parte, 15 de Julio de 1997, II parte vol. 5, núm. 49, pág. 66, 15 de septiembre de 1997. También en : *Simposium Internacional de Computación*. Centro de Investigación en Computación. Instituto Politécnico Nacional. Noviembre 12-14, 1997. México, D.F., páginas 36-51.

99. Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.

## 20. Detectando tendencias y haciendo extrapolaciones en informes escritos en español

OTROS NOMBRES (títulos alternos). Series de tiempo en datos simbólicos.

ÁREA (dentro de la Computación): Minería de datos simbólicos. Minería de textos en español. Área: Inteligencia Artificial. Sub-área: procesamiento de lenguajes naturales.

DESCRIPCIÓN. En qué consiste el trabajo. Defínase una tendencia como una de las siguientes condiciones: la mención del tema "béisbol" va bajando (a través de los meses, digamos, en un determinado periódico), contando por ejemplo el número de artículos que lo mencio-



nan "significativamente" (un artículo menciona "significativamente" a un tema cuando lo menciona más de  $x$  veces, o más del  $z$  % de los conceptos mencionados en el artículo corresponden al tema. Ejemplo: un artículo menciona "béisbol" cuando más del 6 % de los conceptos del artículo se refieren a béisbol). Otra tendencia: ese número o porcentaje va subiendo "considerablemente" (se incrementa más del 8 por ciento mensual). Otra: ha llegado a un máximo. Otra: a un mínimo. La idea es leer con Clasitex un gran número de artículos, mes a mes, y ver si alguna de estas tendencias ocurre. Es como un "minero simbólico" que encuentra "situaciones interesantes" o tendencias en artículos en español.

Úsese Clasitex para contar el número de ocurrencias de conceptos que "votan" o se refieren a "béisbol". Una vez reducido cada artículo a un número de ocurrencias, úsense los métodos comunes de minería de datos para determinar si hay incremento, decremento, salto, mínimo, máximo, o alguna otra situación interesante.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Para detectar situaciones interesantes, tendencias, en textos escritos en español. Manejo inteligente de textos en español. Para saber la dinámica de la atención de una empresa, de un público. Cómo varía la importancia de un suceso, por la frecuencia con que se menciona el tema. *Bono adicional:* haga estadísticas que reporten más o menos lo siguiente: lo normal es que un tema de cierta importancia dure una semana en las noticias, un tema más importante (la muerte de la princesa Diana) puede durar hasta 2 meses, etcétera. Esto nos permite comparar la "importancia relativa" de dos sucesos, midiendo el tiempo e intensidad con que permanecen en las páginas principales de una publicación. *Variante:* analice los distintos artículos publicados en, digamos, los últimos tres años de *Pattern Recognition*, y haga estadísticas que nos digan cómo el tema "fusión de sensores" ha arrancado, cuándo llegó a su cúspide, y cómo ha declinado, quiénes son los autores o grupos más destacados, los grupos seguidores imitadores, los grupos líderes o iniciadores, etc. Cómo ha languidecido el tema "aplicaciones cliente-servidor", el tema "repositorio de datos", etc.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Utilice Clasitex ó Clasitex++ como punto de partida.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Ver también trabajo #19.

98. Guzmán A. Hallando los temas principales en un artículo en español. *Soluciones Avanzadas*. Vol. 5, núm. 45, pág. 58. I parte, 15 de Julio de 1997, II parte vol. 5, núm. 49, pág. 66, 15 de septiembre de 1997. También en : *Simposium Internacional de Computación*. Centro de Investigación en Computación. Instituto Politécnico Nacional. Noviembre 12-14, 1997. México, D.F., páginas 36-51.

99. Guzmán, A. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.

## 21. Cuidador electrónico de automóviles

**OTROS NOMBRES** (títulos alternos). Detector de salidas no autorizadas.

**ÁREA** (dentro de la Computación). Electrónica digital.

**DESCRIPCIÓN.** En qué consiste el trabajo. Utilizando gafetes activos (son unas antenas escondidas dentro de un gafete, con un chip, reciben una señal de radio y transmiten otra --



su identificación), diseñar un sistema para que un transmisor de baja potencia "cuide" a los automóviles que están en el estacionamiento privado de un edificio. Los autos entran, y el gafete los identifica ("ya llegó el Chevrolet verde"). Al salir, el gafete se identifica ("ya se va el Chevrolet verde", ya no se está recibiendo la señal). El dispositivo radía con baja potencia, a una antena de mayor ganancia que se coloca en la azotea del edificio, alumbrando al estacionamiento a proteger. Cuando una persona se aleja en su automóvil, tiene que avisar de alguna manera a la computadora a fin de que sea "salida normal". Por ejemplo, teleándole al chip (o a la computadora en el edificio) una contraseña. La falta de este aviso hará que la computadora notifique a los guardias del estacionamiento "salida anormal del Chevrolet verde placas xxx", o dé algún aviso parecido de alarma. *Variante:* no ponga la antena en el techo, póngala cerca de la puerta de salida del estacionamiento. Necesitará menos potencia.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Evita robos de vehículos estacionados. Eliminación o detección de robos en lugares confinados. Reducción del personal de vigilancia. Aumento de seguridad. Protección.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Hay que diseñar el hardware, el software y las comunicaciones. Se parece a los pagers y a las comunicaciones de los teléfonos inalámbricos (no de los celulares). Hay que ver qué frecuencias o técnicas usar, a fin de no sufrir interferencias indeseadas.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

El Ing. Marco Antonio Reyes García estuvo desarrollando antenas que formaban parte de gafetes activos, en el Instituto Mexicano de la Comunicación, S. C. T.

## 22. Transparentador de profesores

**OTROS NOMBRES (títulos alternos).** Despliegue inteligente de imágenes recibidas.

**ÁREA (dentro de la Computación).** Procesamiento de imágenes. Sistemas en tiempo real.

**DESCRIPCIÓN.** En qué consiste el trabajo. Cuando se está televisando una clase o conferencia, la cámara enfoca al pizarrón o a la pantalla. En eso el profesor se atraviesa y con su cuerpo bloquea parte de la imagen. Y no se ve a dónde apunta o qué está señalando. La idea es con la computadora alterar la imagen transmitida de manera que el cuerpo del profesor se borre o se haga transparente, y vuelva a aparecer el texto que se encuentra oculto por él.

*Variante 1:* Hacer la imagen del profesor no del todo transparente, sino que se vea apenas, como fantasma o silueta.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Transmisión completa de imágenes, a pesar de obstrucciones indeseables.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. En la computadora receptora, guarde la imagen sin interferencia, y cuando detecte que hay una "interferencia de profesor", conmute y despliegue sucesivamente una y otra imagen, esto nos dará el efecto de transparentar a la obstrucción y semi-removerla. Regrese a la imagen nueva cuando la obstrucción cese.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

### 23. Juntas electrónicas para la Comisión de Honor del I. P. N.

OTROS NOMBRES (títulos alternos). Reuniones electrónicas.

ÁREA (dentro de la Computación). Sistemas de información, *groupware*.

DESCRIPCIÓN. En qué consiste el trabajo. Para dictaminar sobre varios casos, los miembros de la Comisión de Honor (y de muchos otros comités dentro y fuera del I. P. N.) se reúnen presencial y sincronamente, para discutir sobre tal o cual caso, examinando para cada uno varios documentos.

- a) Tal examen pudiera ser *no presencial* pero sí sincrónico, es decir, a determinada hora cada uno de los miembros de la Comisión se sientan frente a una PC en su oficina, y se integran a la junta.
- b) Otra variante: pudiese ser no presencial, como (a), y también *asíncrona*: en realidad la "junta" dura dos días, con un horario largo: tienen los integrantes hasta las 10 de la mañana para enviar candidatos; hasta las 6 p.m. para enviar objeciones, hasta las 10 p.m. para votación preliminar, etc.

Existe ya software comercial para juntas electrónicas, el proyecto podría ser imitar uno, tratando de mejorarlo o de agregarle algunas características útiles adicionales. Con este software es posible hacer votaciones de distintas maneras, conservar el anonimato de los emisores de opiniones o votos, opinar en paralelo, hacer consensos sobre las cinco ideas o candidatos mejores, los 5 peores, etc.

### 24. Construcción de una computadora paralela como la SP2 utilizando procesadores Pentium (no Power PCs) y un switch de fibra óptica

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Arquitectura de computadoras, Sistemas Digitales.

DESCRIPCIÓN. En qué consiste el trabajo. Hay dos tipos de máquinas paralelas (fuertemente acopladas), aquéllas que intercambian datos mediante *memoria compartida* (ejemplo comercial: Origin, de Silicon Graphics) y las que los intercambian mediante conmutación de mensajes o líneas o switches (ejemplo: IBM SP2). Ahora bien, una SP2 está formada por varias Power PCs interconectadas por un switch rápido. La idea es usar Pentiums (u otras PCs con chips de la familia 80X86) e interconectar estas máquinas mediante un switch óptico (de fibra óptica), el que se compraría. Se comprarían también las Pentiums. El trabajo consiste en hacer el diseño, las interfaces, la interconexión, adaptarle algún sistema operativo (Unix en paralelo o para varios procesadores, digamos), y dejar la máquina andando, operacional.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Diseño y construcción de una máquina paralela barata. Entender las técnicas de conexión a un switch de fibra óptica.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Compre el switch óptico.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

## 25. Traductor de un lenguaje "estándar" 4GL hacia SQLs de Informix, Progress y Oracle

**OTROS NOMBRES** (títulos alternos). Accesador universal a bases de datos

**ÁREA** (dentro de la Computación). Tecnología de software, Sistemas de Información.

**DESCRIPCIÓN.** En qué consiste el trabajo. La idea es programar en un 4GL más o menos común (por decir algo, en el de Progress), y poder conectarlo hacia otro manejador de base de datos, haciendo (en la variante 1) un pre-procesador que substituye las llamadas que realmente accesan la base de datos (de Progress), con llamadas a otras subrutinas que lo que hacen es accesar (variante 1) la base de datos de Oracle, y devolver los buffers (o lo que sea) llenos de datos tal como lo hubiera hecho Progress.

*Variante 1:* Traductor de Progress hacia Oracle. Compilador o macro-expansor que convierte los comandos de lectura de Progress, reemplazándolos por llamadas (en Progress) hacia una rutina que es la que llama al comando correspondiente de Oracle, el cual lee los datos del motor de Oracle y los devuelve (llenando el buffer de memoria) tal como los hubiera devuelto o recobrado el comando de Progress. Y similarmente para los comandos que escriben, los que verifican si la llave es inexistente, etc.

*Variante 1.b.* Intérprete (caja de datos) de comandos de Progress cuando el motor se ha reemplazado con uno de Oracle. Es un intérprete que intercepta los comandos, en vez de un preprocesador que convierte comandos en un lenguaje (Progress) hacia otro (Oracle).

*Variante 2.* Traductor de Progress hacia varios lenguajes, no solo hacia Oracle.

Cuando en las variantes anteriores decimos "Progress", en realidad estamos pensando en un lenguaje sencillo o cómodo (por eso pensé en Progress) para escribir nuestros programas, y de ahí hacer la traducción (o interpretación) automática hacia otros motores de bases de datos cuyos lenguajes 4GL no deseamos aprender. Es decir, pudiera ser que el lenguaje cómodo en que programamos rutinariamente no sea "Progress", sino alguno otro.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Para proporcionar independencia del motor de la base de datos. Para usar un único lenguaje de programación.

Para convertir sistemas que accesan a una base de datos relacional, a otra base de datos.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

## 26. Traductor de los diferentes front-ends que accesan una base de datos

**OTROS NOMBRES** (títulos alternos).

**ÁREA** (dentro de la Computación). Tecnología de software.

**DESCRIPCIÓN.** En qué consiste el trabajo. De Internet-Intranet, de Visual Basic, de Delphi.

O sea, entre ellos. Se puede acoplar con el "accesador universal de bases de datos", que es el proyecto 25.

La idea es programar el "front end" (que accesa a una base de datos) en un "lenguaje cómodo" (Delphi, digamos), y de ahí traducir hacia Internet-Intranet, hacia Visual Basic, etc. Tenemos aquí las mismas variantes que en el proyecto 25.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Proporcionar independencia del "front-end" que accesa una base de datos.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.  
TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

## 27. Manejador de documentos electrónicos (workflow) y convencionales

OTROS NOMBRES (títulos alternos). Manejador híbrido. Flujo de documentos

ÁREA (dentro de la Computación). Sistemas de información.

DESCRIPCIÓN. En qué consiste el trabajo. Existen manejadores de documentos electrónicos (workflow) que tienen un solo servidor. La Tesis de M. en C. de Cecilia Palomino (C. I. C.-I. P. N.) trata del diseño y construcción de un manejador con varios servidores, de manera que los trámites pueden emigrar de un servidor a otro, sin perder ilación o continuidad. Pro siga usted a lo largo de estas ideas, para hacer un manejador híbrido que combine los documentos electrónicos con los documentos de papel (convencionales).

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Un manejador de documentos electrónicos sirve para: Eliminación del papel, automatización de oficinas, agilización de trámites, seguimiento de asuntos, control de asuntos. Uno híbrido tiene el atractivo adicional de hacer una transición más suave o fluida entre la cultura de documentos de papel y la de documentos electrónicos.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Básese en la parte ya existente.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos. Cecilia Palomino, tesis de M. en C., C. I. C.-I. P. N. 1998. (en terminación)

## 28. Construcción de un pizarrón de mensajes ("object request broker") para aplicaciones de inter-operabilidad de software ya existente y en producción

OTROS NOMBRES (títulos alternos). Pizarrón de distribución de mensajes que solicitan servicios, pero cuyos servidores no se conocen. Conectando sistemas ya instalados que no se diseñaron para interactuar. Cómo darle longevidad a software ya existente y operacional, mediante su interconexión.

ÁREA (dentro de la Computación). Computación distribuida. Inter-operabilidad de sistemas, integración de aplicaciones

DESCRIPCIÓN. En qué consiste el trabajo. En una empresa con cierta antigüedad, existen sistemas de información aislados, que no se comunican entre sí. Fueron creados independientemente, en distintas plataformas, para distintos usos. En una empresa grande, existe software que ya trabaja, por ejemplo leyendo datos de unos archivos o pantallas de captura, haciendo una operación (venta al contado, cobranza del pago del recibo telefónico, inscripción a un examen a título de suficiencia). Estos sistemas se diseñaron aislados, de manera que, por ejemplo, un paquete le informa a un alumno que reprobó, pero no le produce un "recibo de pago" para que vaya a pagar su examen extraordinario. Él tiene que usar otro sistema, que le vuelve a preguntar los datos, porque el sistema dos no está conectado con el sistema uno. Otro ejemplo: en su banco Juan Pérez tiene una cuenta de ahorros y otra de



cheques. Cada uno lo maneja un software distinto, el primero está en una IBM AS 400 y el segundo en una HP con Unix. Juan Pérez cambia de dirección, y le toman sus datos para su cuenta de ahorros. Le tienen que volver a tomar los mismos datos para su cuenta de cheques, porque los sistemas no están conectados. Esto origina duplicación de información, lentitud, captura doble (con la consiguiente introducción de errores de captura o de teclado) porque la conexión entre sistemas se hace a mano. Después del tiempo, se ve la conveniencia de que el sistema A le pase información al sistema B, en ciertos casos. Y el problema se complica si hay varias decenas de sistemas que podrían, en ciertos casos, comunicarse con otro(s).

El problema a resolver es interconectar este software, haciendo que los resultados de un paquete vayan y avisen al otro paquete que tal o cual cosa debe suceder. (Una solución que se descarta en este trabajo es "volver a hacer todo de nuevo" usando una misma base de datos, es decir, volver a hacer desde cero un sistema integral de información, o de manejo de alumnos, o de transacciones en bancos. En este proyecto suponemos que tal solución es muy costosa y lenta, por lo que preferimos tender puentes entre los sistemas ya existentes).

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Aumenta la longevidad de software ya existente y en producción. Aumenta la productividad de la empresa. Ahorra conexiones manuales, por ejemplo, volver a recapturar los datos. Por ejemplo, tener que revisar manualmente 30 bases de datos (una por cada estado de la república) cuando buscamos un número de placas de automóvil.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

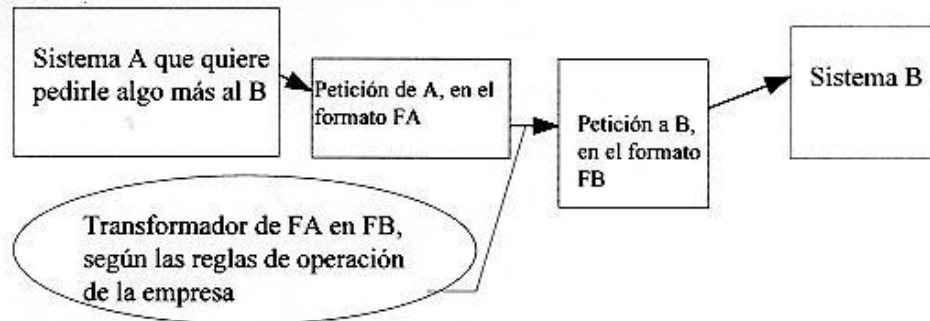
1. Abra las aplicaciones. Migrar a sistemas abiertos. Constrúyales una IPA (interfaz programática de aplicación, en inglés API) para que otros programas puedan usarlas.
2. Construya un pizarrón central para manejar (recibir y redirigir) los mensajes que distintas aplicaciones le dirijan a otras "que tal vez pudieren interesarle", algo así como el mensaje "a quien corresponda."
3. Construya puentes desde este pizarrón central a cada una de las aplicaciones, utilizando las IPAs previamente diseñadas en (1).
- 3bis. *Variante.* Si no son muchas las aplicaciones a interconectar, construya directamente puentes entre cada par de ellas (que realmente *haya* que comunicar, obviamente).
4. Considere la infraestructura. Redes. Software para convertir. El problema de acceder a una IBM AS 400. El problema de acceder a Informix desde el web.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

1. John J. Donovan. *Business Re-engineering with information technology*. PTR Prentice Hall. 1994. Es un libro con una metodología para establecer puentes entre sistemas ya existentes, haciendo que colaboren entre sí.
2. En Pemex PEP, el Ing. Cadena conectó los principales sistemas de administración, haciendo juntas entre expertos, y preguntándoles cosas como: cuando se hace un pedido con el sistema A, ¿qué información hay que pasarle al sistema B de almacenes? Los expertos (en almacenes, no necesariamente en informática) dicen qué debería suceder lógicamente, por ejemplo "agregar a los registros de los productos pedidos, la información de que tal mercancía debe llegar, a tal precio, de tal proveedor, aproximadamente en tal fecha." Luego se consulta a otros expertos (en informática) en los sistemas A y B que nos explican qué cosa sucede en realidad, o sea, dónde se queda corto A y B. Luego, Cadena o alguien diseña dos archivos, uno de salida de A



que se considera como "peticiones a B". Otro de entrada de B que se considera como "atenderé peticiones que se encuentren en el formato B". Y un convertidor de texto a texto que pasa los datos del formato A al formato B.



### 29. Transformación de funciones en tiempo real (en línea), a otras de tipo store & forward, en una bodega de datos

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Computación distribuida, bodegas de datos.

DESCRIPCIÓN. En qué consiste el trabajo. Cómo hacer que ciertas funciones que normalmente se ejecutan en tiempo real en un sistema de data warehousing, ahora se ejecuten con información obtenida en forma diferida, con los mismos resultados, o con resultados equivalentes, pero sin necesidad de utilizar conexiones en tiempo real, y usando en cambio información diferida. Por ejemplo, en vez de actualizar inmediatamente la réplica de una base de datos o tabla, acumulo mis transacciones de actualización y luego las envío en bloque.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Mitiga la necesidad de comunicaciones de gran disponibilidad. Reduce el ancho de banda (se pueden usar tiempos nocturnos para realizarlas).

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

### 30. Presentador de reportes dinámicos, reorganizables

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Interfaces de entrada y salida, sistemas de información.

DESCRIPCIÓN. En qué consiste el trabajo. Haga un desplegador de reportes, pero escóndale dentro una pequeña base de datos (mejor: esconda unos archivos indexados) a fin de que los reportes que se muestran sean interactivos, y el usuario que va a ver los informes que pidió, pueda por ejemplo, ver nadamás las mujeres, o nadamás los que tienen saldo mayor a mil pesos.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Consultas no planeadas, con cierto grado de inteligencia o generalidad.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Utilizando un extractor de datos como los que posee Anasin, cargue la mini-base del des-

plegador con los datos solicitados. Invente (genere) vistas o reportes comunes, y un menú sencillo para que el usuario pueda ver otros. Bono adicional: agréguele drill-down (la capacidad de ver detalles de información global, por ejemplo, ventas en Coahuila = 580, le pico al 580 y me da una pantallita donde me da detalles: Torreón 300, Saltillo 280).

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. La base de datos Progress tiene un presentador de este tipo. El M. en C. Rafael Domínguez tiene experiencia e interés en este tipo de desarrollos, y puede dirigir o asesorar en el trabajo.

### 31. Digitalizador en tres dimensiones utilizando rayos láser

**OTROS NOMBRES (títulos alternos).** Captura de imágenes en tres dimensiones.

**ÁREA (dentro de la Computación).** Dispositivos de entrada y salida.

**DESCRIPCIÓN.** En qué consiste el trabajo. Construya usted un dispositivo que permita captar una imagen en tres dimensiones. Este dispositivo le permitirá montarlo en el centro de una plaza pública, por ejemplo, para fotografiar (tener un modelo tri-dimensional) de los edificios, los árboles, los monumentos, etc., que rodean o forman parte del recinto. Es como si tomara una fotografía panorámica, o con un lente "ojo de pescado", que toman una fotografía amplia del lugar. Con la ventaja de que la imagen captada es en tres dimensiones.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Para tomar fotografías tridimensionales estáticas.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

El dispositivo es un lanzador de rayos láser que mide la distancia a donde llega el láser. Es similar al que tenemos en las "reglas láser" que permiten medir la distancia entre dos puntos, más específicamente, entre el instrumento y el punto donde rebota o toca el rayo.

La idea consiste en montar este lanzador en una mesa que va girando lentamente de manera horizontal, de manera que se van grabando los ángulos de giro y las distancias a donde el rayo llegó. Esto nos produce una "rebanada" o plano del espacio. Para capturar otra rebanada, se levanta un poco el rayo, digamos 5 grados. Y se vuelve a girar la mesa, hasta completar 360 grados. Esto produce otra "rebanada" del espacio (más bien un cono) que rodea al dispositivo fotográfico. Para capturar otra rebanada, se levanta otros cinco grados el rayo, y se vuelve a girar la mesa hasta completar 360 grados. De esta manera, en coordenadas polares, se van leyendo las distancias a que los distintos objetos yacen, con respecto al origen del rayo.

El dispositivo registra (teta, fi, erre), en coordenadas polares, donde erre es la distancia a la que se encuentra el punto de rebote del rayo, fi es la elevación del rayo, y teta es su azimut.

*Bono adicional:* Una vez hecho un levantamiento (fotografía en 3 D) desde un cierto origen, mueva el láser a otro punto, haga otro levantamiento, y con cálculos geométricos, mejore la calidad (resolución) de su fotografía 3-D.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

### 32. Buscador inteligente de información en la web, de ejecución diferida

OTROS NOMBRES (títulos alternos). Búsqueda diferida de documentos en una red.

ÁREA (dentro de la Computación). Análisis de texto y lenguaje natural, computación distribuida, quizá agentes.

DESCRIPCIÓN. En qué consiste el trabajo. Utilizando Clasitex, lance un agente que busca artículos (en la red, en web) que hablen de algún tema específico.

Variante: Hágalo para que ejecute en tiempo real.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Hallazgo inteligente de textos en español (y, utilizando Clasitex+, en inglés).

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

El grupo de Computación del INAOE ha realizado y está realizando buscadores de este tipo.

En el C. I. C., el M. en C. Norberto Medina tiene un proyecto en esta línea.

Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.

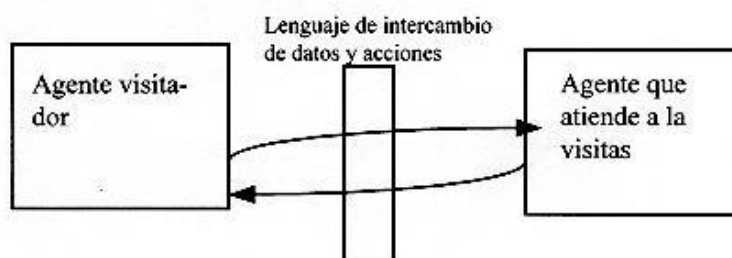
101. Beatriz Beltrán Martínez, Adolfo Guzmán Arenas, Francisco Martínez Trinidad, José Ruiz Shulcloper. Clasitex++: una herramienta para el análisis de textos. Memorias del *Tercer Taller Iberoamericano de Reconocimiento de Patrones*, TIARP-98, Centro de Investigación en Computación, Instituto Politécnico Nacional, México, D. F. Marzo 1998. Páginas 369-379

### 33. Interacción segura entre agentes sospechosos

OTROS NOMBRES (títulos alternos). Lenguaje de comunicación para intercambio de información entre agentes.

ÁREA (dentro de la Computación). Agentes.

DESCRIPCIÓN. En qué consiste el trabajo. Es peligroso dejar que agentes externos entren a una computadora a realizar trabajos. Es más fácil recibirlos "en la puerta de entrada", preguntarles qué información o ejecución desean, entregarles lo pedido (o realizarlo), y despedirlos. Se puede inclusive grabar una bitácora con el intercambio de información. Para esto, se requiere un lenguaje de comunicación (y su procesador correspondiente) para solicitar datos, o para solicitar cierta acción. La interacción se lleva a cabo a través de un muro de seguridad, parecido pero no sustituto del "muro de contención" («firewall») que separa una Intranet de la Internet.



**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Permite que las transacciones entre agentes se lleven a cabo de manera segura. El agente que atiende puede llevar una bitácora de la transacción, con "constancias" de lo que se pidió y lo que se hizo o se entregó. El lenguaje debe ser de tal naturaleza claro o sencillo, para que las expresiones en el mismo puedan ser analizadas por el agente receptor y no haya lugar a dudas de que las acciones solicitadas son válidas y no involucran peligro o acciones indebidas. Por ejemplo, no contendrá apuntes ni índices no acotados, por el riesgo que implica de alterar datos indebidamente.

La principal aplicación es en comercio electrónico entre agentes.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Ver algunas conferencias entre agentes en las Memorias del 4th World Congress on Expert Systems, ITESM México Mar 98.

Algunos proyectos de esta lista están relacionados. El lenguaje probablemente tenga una parte para identificar al agente visitante, otra para especificar datos, otra para especificar comandos. Además, el visitador se referirá a los datos que necesita expresado en un diccionario de datos que no necesariamente concuerde con el diccionario del lugar visitado. Concéntrate en la parte del lenguaje (y su intérprete o procesador) que maneja intercambio de información, y quizá también en la de ejecución "segura" de comandos. Ver Proyecto 38.

#### 34. Replicador automático de archivos, con actualización diferida o en tiempo real

**OTROS NOMBRES (títulos alternos).** Transmisor automático de información de un lugar a otro.

**ÁREA (dentro de la Computación).** Sistemas de información.

**DESCRIPCIÓN.** En qué consiste el trabajo. Desarrolle un software al que se le declaran dos archivos (con la trayectoria completa de directorios, incluyendo el nombre de la máquina donde residen), uno fuente y otro destino. Su software mantendrá actualizada (en sincronía) la copia destino, cada vez que se hagan cambios en el archivo fuente. La actualización podrá ser en tiempo real o en forma diferida, según las facilidades de comunicación con que se cuente. Normalmente, a la copia (archivo destino) no se le podrán hacer modificaciones, será solamente de lectura. *Bono adicional:* que se le puedan hacer modificaciones al archivo destino, mismas que se propagarán al archivo fuente. Full duplex. Desaparece entonces la distinción entre archivo fuente y archivo destino.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Transmisión automática de información y de sus cambios.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

- 1) Lotus notes.
- 2) Ver Proyectos 34 y 38.

### 35. Explicaciones automáticas

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Inteligencia Artificial.

DESCRIPCIÓN. En qué consiste el trabajo. Desarrolle un programa que ofrezca explicaciones útiles. Monte su programa sobre otro (por ejemplo, un sistema experto) ya hecho, y concéntrese en la parte de la explicación.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Mejora la aceptación y credibilidad de ciertas recomendaciones hechas por la computadora, digamos a través de un sistema experto.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Típicamente, los sistemas que explican dan información (a) sobre la línea de cómputo que ejecutaron, vaciando la pila de acciones o traza que condujo a la "solución"; es decir, explican la causa o razón de la salida o resultado; o (b) con frases cortas predeterminadas. Esto es poco útil. Dado que el sistema "base" (cuyas salidas tratamos de explicar) acaba de proporcionar cierta salida o información I, agregue usted formas de ofrecer algunas de las siguientes explicaciones:

(c) Detalles acerca de la nueva información I.

(d) Cómo es que la información I se relaciona con lo que ya se sabía.

(e) Por qué es necesaria la información I.

(f) La *estrategia* de la solución o recomendación I.

(g) Justificación. Cómo y por qué se llegó a la información I. Contestar preguntas del tipo ¿cómo?, ¿por qué?, ¿qué?

Quizá sea útil tener un modelo de las expectativas del usuario, incluyendo su estado de conocimiento, a dónde quiere llegar, qué está haciendo en este momento.

¿Cómo saber o medir la *aceptación* de una explicación? ¿Cuál es el *método* de explicación?

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

Andrew Lyner, Amelia Baldwin, Jan Scott. Explanation generation in accounting expert systems. *Proc. 4th World Congress on Expert Systems*, ITESM, Mexico, March 1998, pages 25-32.

### 36. Anotaciones que ayudan al mantenimiento de programas grandes o ya existentes

OTROS NOMBRES (títulos alternos). Mantenimiento de código fuente ayudado por la computadora.

ÁREA (dentro de la Computación). Tecnología de software.

DESCRIPCIÓN. En qué consiste el trabajo. Desarrolle un sistema que le ayude a mantener programas codificados ya hace tiempo por otras personas.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.



POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

- (1) Conforme vaya estudiando el código existente, vaya haciendo anotaciones sobre pedazos relevantes del código: aquí se actualiza el pago; aquí se validan los datos del cliente, etc. J. K. Lee [1] utilizó palabras clave, en tanto que usted puede usar un árbol de conceptos (no de palabras clave) [2]. Esta labor es manual, pero la hará usted una sola vez.
- (2) Usando el mismo lenguaje o taxonomía de (1), vaya clasificando las variables del código, sobre todo aquéllas que no son temporales o de trabajo.
- (3) Diseñe un machote (una ventana en Delphi, por ejemplo) para que el usuario especifique las modificaciones que requiere. El machote hace uso de verbos o acciones (“aumentar el monto de ...”, “borrar los ...”) sobre de los conceptos de (1).
- (4) Escriba un programa que busque en el código fuente, usando las anotaciones de (1) y (2), aquéllas partes que mejor casan con las especificaciones de (3). Es un programa que hace casamiento o careo de (1+2) contra (3), sobre el árbol de conceptos.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

- [1] Jae K. Lee, Byoung Y. Lee, Wooju Kim. A knowledge based maintenance of large scale legacy systems: Metasoft. *Proc. 4<sup>th</sup> World Congress on Expert Systems*, ITESM, Mexico, March 1998, pages 65-72.
- [2] Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.

### 37. Base de datos que se llena por agentes

OTROS NOMBRES (títulos alternos). Llenado automático de una base de datos vacía por agentes oferentes («push agents») y agentes buscadores.

ÁREA (dentro de la Computación). Agentes. Sistemas de Información

DESCRIPCIÓN. En qué consiste el trabajo. Diseñe un método para llenar tablas que contienen descripciones de lo que cada campo debe contener. El llenado se hará en forma diferida, utilizando agentes que toman esa descripción y la comparan contra descripciones parecidas de bases de datos ya existentes (y que, por ende, pudiesen tener la información solicitada), generalmente geográficamente distantes.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Recopilación automática de información. Sincronización de bases de datos (Proyecto 34). Resúmenes automáticos de información.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

- A. Diseñe una base de datos vacía (cree las tablas).
- B. en cada registro (y campo) coloque una fórmula que indique cómo se debe llenar ese campo. Como si estuviéramos describiendo una macro de una hoja de cálculo. La descripción usará como taxonomía o diccionario de datos (a) uno pre-determinado (caso sencillo), ó (b) uno local, que se casará con otro diccionario de datos exógeno (Ver proyecto 33) usando la ayuda de un diccionario común (árbol del conocimiento común) de conceptos.
- C. Llene la base (A) utilizando agentes de uno de los dos siguientes tipos.

- (1) agentes que usted mismo diseñará (agentes buscadores), que viajarán en la red de la empresa (o en alguna otra red) buscando información que case con la descripción de cada celda; o bien,
- (2) agentes visitadores («push agentes»), que son agentes que otras personas han diseñado, y que buscan bases de datos vacías conteniendo (en el lenguaje de (B)) descripciones de productos o información que casa con la que ellos pueden proporcionar. Un agente visitador que encuentre una celda vacía que él puede llenar, procederá a llenarla. Como en la actualidad no existen esas otras personas que diseñen los agentes visitadores, diseñelos usted mismo. Nótese que los agentes visitadores no se disparan por la creación de la base vacía (A), sino que de manera asíncrona se enterarán de la presencia de una nueva base de datos vacía, misma que tratarán de llenar. Si un agente posee información que no casa con la descripción de las celdas, no las llenará, y continuará su viaje.

D. Tanto los agentes buscadores de C.1 como los visitadores de C.2 requieren un lenguaje que describa los datos de la base nueva, los datos de las bases ya existentes, y un método de comparación o casamiento para ver si una base ya existente puede contener información vertible a la base de datos nueva. Este lenguaje es el problema del proyecto 38. No se dedique a hacer el proyecto 38, busque una alternativa sencilla del lenguaje (es decir, haya una versión simple del proyecto 38).

E. Si se va a cobrar por la información, entonces el lenguaje de (B) se complica como sigue:

1. Qué información ofrece. (esta parte ya está considerada). A qué precio.
2. Descripción del proceso de oferta y contraoferta, peticiones. («bidding process»).
3. Un modelo formal del proceso de regateo o negociación. ¿Cuándo tenemos ofertas que son obligatorias (legalmente ya no las podemos retirar o rehusar)? ¿Cuándo se convierte en legalmente obligatoria una petición de compra? Por ejemplo, este lenguaje debe ser capaz de describir un proceso de subasta. Ver también proyecto 38.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

- [1] Adolfo Guzmán. ANASIN. SoftwarePro International, Austin, Texas, 1994. Anasin es un producto que envía agentes a buscar información (descrita sobre un diccionario de datos ya conocido (caso B.a)) a lugares remotos, para que llenen una base de datos vacía.
- [2] Jae Kyu Lee, KAIST, Corea. Opportunities of Artificial Intelligence in Electronic Commerce. *Proc. 4<sup>th</sup> World Congress on Expert Systems*, ITESM, Mexico, March 1998, page 8 (abstract only).

### 38. Lenguaje para intercambiar información entre agentes

OTROS NOMBRES (títulos alternos). Casamiento de descripciones que utilizan taxonomías (diccionarios de datos) distintos.

ÁREA (dentro de la Computación). Sistemas de Información. (El lenguaje tiene poco que ver con agentes).

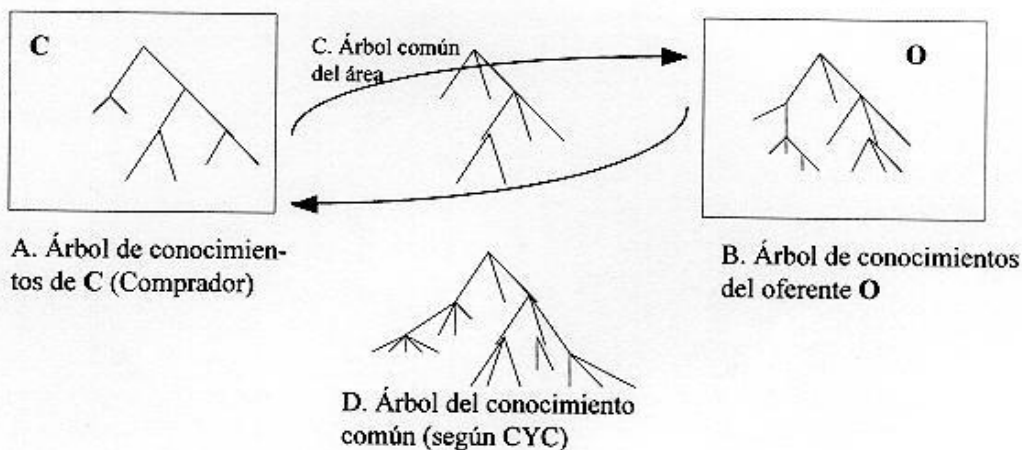
**DESCRIPCIÓN.** En qué consiste el trabajo. Sea **O** un agente que ofrece información, vende servicios, o renta artículos. **O** describirá su oferta en un lenguaje que se refiere a un diccionario de datos local, suyo. Por ejemplo, "vendo computadoras personales marca Compaq que son de multimedia, ..." Sea **C** un agente que desea comprar o adquirir cierta información, bien o servicio. **C** describe sus necesidades en el mismo lenguaje que **O**, pero refiriéndose a un diccionario de datos propio de **C**. "Deseo comprar una PC pequeña de buena marca, con 2 Gby en disco duro, ..." Para que haya una transacción, es necesario que ambos agentes descubran si hablan de lo mismo. El proyecto consiste en diseñar el lenguaje, y el método de casar una expresión en  $L_O$  contra otra en  $L_C$ .

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Comercio electrónico. Acceso a bases de datos poco familiares.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

Utilice cuatro árboles de conceptos [3], como sigue:

- A. El árbol de conocimientos del área del agente **C** (el comprador, el sumidero de datos).<sup>5</sup>
- B. El árbol de conocimientos del área del agente **O** (el vendedor, el oferente o fuente de datos).
- C. Un árbol de conocimientos del área pero comunes (por ejemplo, el área puede ser Ferretería. Entonces estamos hablando de un lenguaje común a los ferreteros).
- D. El árbol de conocimientos comunes, según CYC [4].



**Trucos:** (1) utilice el código de barras de productos que ya existe en industrias organizadas (industria refresquera, por ejemplo). El lenguaje se refiere entonces a productos cuya descripción existe en el estándar de la industria. (2) Utilice el hecho de que la jerarquía misma contenida en los árboles descarta pronto grandes ramas de productos

<sup>5</sup> Con un poquito de suerte, los árboles de los incisos A y B estarán basados tan solo el diccionario de datos de la base local, que fue la variante que utilicé en [1]. Empero, el árbol de conocimientos comunes de [1] es el del inciso C.

(por ejemplo, cerca del nodo raíz se descartan productos tales como "Comida" o "Eventos musicales", si lo que estoy buscando es "Ferretería").

Si complicamos más el lenguaje, éste debe poder describir:

1. El tipo de agente que soy. Comprador. Vendedor. Rentero. Consignación. Es independiente del dominio.
2. Qué información ofrece. (esta parte ya está considerada, es el lenguaje de descripción del producto, o Capa de Especificación del Producto.)
3. Descripción del proceso de oferta y contraoferta, peticiones. («bidding process»). Si los agentes que visitan no solo van a intercambiar información, sino que le van a solicitar a la máquina visitada que realice cierto proceso, hay que poder describirlo (Knowledge Query and Manipulation Language, ver proyecto 33)
4. Un modelo formal del proceso de regateo o negociación. Ver Proyecto 37.

Nótese que el lenguaje debe poder permitir la interacción ente comprador **C** y vendedor **O** no importa si uno de ellos (o ambos) es una persona (en vez de un agente). En este caso, la persona debe identificarse como tal "soy persona, no agente", y el otro lado que lo atiende debe poder pasarse a un modo de interacción basado en menús gráficos y formas de captura interactivas, guiadas (¿construidas al vuelo?) por el lenguaje

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

- [1] Adolfo Guzmán dirigió una tesis de maestría de un estudiante de Computación del Cinvestav (≈1994, no se recibió) que hacía búsqueda en bases de datos poco familiares, utilizando el mapeo de un árbol de conocimientos comunes en una cierta área, a las definiciones específicas de la base de datos a la cual se quería extraer información.
- [2] Jae Kyu Lee, KAIST, Corea. Opportunities of Artificial Intelligence in Electronic Commerce. *Proc. 4<sup>th</sup> World Congress on Expert Systems*, ITESM, Mexico, March 1998, page 8 (abstract only).
- [3] Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.
- [4] Lenat, D. B., and Guha, R. V. (1989) *Building large knowledge-based systems*. Reading, MA: Addison Wesley.
- [5] S. K. Lee, Jay Kyu Lee, K. J. Lee. *Journal Expert Systems with Applications*, Vol. 11 No. 4, pages 431-441, 1997.
- [6] Jesús Olivares está desarrollando en el C. I. C. una tesis de doctorado que involucra agentes y sistemas evolutivos (que quizá aprendan), y probablemente se usen también técnicas como las de Clastex para análisis de lenguaje natural.

### 39. Uso de modelos digitales del terreno para estudiar erosión

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). GeoProcesamiento.

DESCRIPCIÓN. En qué consiste el trabajo. Utilice los modelos digitales del terreno, por ejemplo, los que utilizan vóxeles, para estudiar la erosión real de agua y viento. Para esto, agregue a sus modelos información sobre el tipo de suelo, dureza, precipitación, vientos



dominantes, etc. Haga simulaciones o predicciones de qué pasaría en cierto tiempo  $x$ , en  $2x$ , ...

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. Idea: use las técnicas de erosión y dilación de morfología matemática, convenientemente cambiadas para tomar en cuenta la dureza del suelo y los efectos de la lluvia (o viento). Es decir, cambiar la métrica de las operaciones morfológicas de dilación y erosión (matemática) para modelar la erosión real (hídrica, eólica).

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

35. Gómez, Dora y Guzmán, A. A digital Model for Three-dimensional Surface Representation. *Journal of Geoprocessing* 1, 1979, 53-70. Elsevier Publishing Co. Also in: *Proceedings of an International Conference "Computer Mapping for Resource Analysis"*, a CoGeoData Conference. Kansas Geological Survey, University of Kansas, and Instituto de Geografía de la UNAM. Mexico, 1978, pages 183-204.

#### 40. Números de forma en tres dimensiones

**OTROS NOMBRES** (títulos alternos).

**ÁREA** (dentro de la Computación). GeoProcesamiento, reconocimiento de formas.

**DESCRIPCIÓN.** En qué consiste el trabajo. Generalice usted los números de forma [33, 34, 39<sup>6</sup>] para que describan formas (superficies) en tres dimensiones.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Para describir la forma de una superficie tridimensional.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

33. Bribiesca, E., y Guzmán, A. Shape Description and Shape Similarity Measurement for Two-dimensional Regions. *Proceedings of the 4th International Conference on Pattern Recognition*, Kyoto, Japón, 1978. 608-612. Also available as Technical Report PR 78 18 (Orange Series 9, 166), IIMAS UNAM, Also in: *Journal of Geoprocessing*, Vol. 1, No. 2, 129-144 (1980).

34. Bribiesca, E. y Guzmán, A. Números de Forma: una notación para describir formas puras y para medir semejanzas y diferencias en formas. *Reporte Técnico PR 78 20* (Serie Naranja 178), IIMAS UNAM, 1978.

39. Bribiesca, E. y Guzmán A. How to Describe Pure Forms and how to Measure Differences in Shapes using Shape Numbers. Invited paper to the IEEE Conference on Pattern Recognition and Image Processing. Chicago, USA. Also in *Pattern Recognition*, Vol 12, No. 2, 1980, 101-112. This article won the Seventh Annual Pattern Recognition Award, awarded by the Pattern Recognition Society in November 1981, as the best article of the year.

---

<sup>6</sup> La numeración de los artículos de A. Guzmán sigue el orden en que aparecen en su curriculum.

## 41. Análisis de imágenes médicas

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación). Procesamiento de imágenes.

DESCRIPCIÓN. En qué consiste el trabajo. Tome usted un conjunto o tipo (de los muchos que puede haber) de imágenes médicas: cortes del riñón, del hígado, imágenes de venas, de eritrocitos, ..., pregúntele al médico o especialista respectivo para qué se tomó la foto, qué se le busca, qué información aporta. Haga un programa de cómputo que halle las respuestas de manera automática o semi-automática.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Automatización del análisis y reconocimiento de imágenes médicas.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

## 42. Lenguaje de simulación de efectos y consecuencias (en poblaciones de afectados), de políticas, reglas y burocracias

OTROS NOMBRES (títulos alternos). Análisis de las consecuencias de decisiones sobre conjuntos de poblaciones.

ÁREA (dentro de la Computación). Simulación.

DESCRIPCIÓN. En qué consiste el trabajo. Construya un lenguaje y su procesador (intérprete, quizá) que permita estudiar las consecuencias o efectos que ciertas disposiciones, reglamentos, acciones, normas, leyes, costumbres, sobre todo de tipo administrativo o burocrático, tienen sobre personas. Para esto el lenguaje debe ser capaz de:

1. Definir los objetos (personas, digamos), sus propiedades, sus variables (salarios, estado de descontento, tendencia a cambiar de empleo, tendencia a votar por el partido de oposición). Probablemente los objetos tengan una jerarquía: estudiantes, profesores, burócratas, funcionarios, jefes de compras, ... Probablemente haya otros entes (objetos) que no sean personas estrictamente, pero que funcionen como tales; comité de compras, comité de becarios, comité de viajes, comité de prestaciones, comité del año sabático, comité de lo que usted mande y guste.
2. Las interacciones, la manera en que estos objetos o personas interaccionan entre sí.
3. Una forma de visualizar el progreso de la simulación. Graficación, despliegue de dibujos que visualmente muestren lo que va ocurriendo, de manera dinámica, conforme la simulación avanza. Cómo cambia o se agudiza el estado de descontento, frente a una decisión ("suspender la beca del profesor, mientras lo evalúo").
4. Maneras de detener la simulación, cambiar parámetros, etc.
5. *Bono extra*. Maneras de dar marcha atrás.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Estudiar los efectos de decisiones burocráticas, vicios, tendencias, sobre poblaciones confinadas (investigadores del C. I. C., por ejemplo).

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Use un lenguaje para hacer simulación.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

### 43. Administrador del conocimiento de una empresa

**OTROS NOMBRES (títulos alternos).** Programa bibliotecario que ha leído los manuales técnicos, patentes, y otros documentos textuales de una organización, y conoce los temas de los que trata cada uno.

**ÁREA (dentro de la Computación).** Procesamiento de texto y lenguaje natural, inteligencia artificial.

**DESCRIPCIÓN.** En qué consiste el trabajo. Mucho del valor de una empresa es su conocimiento, guardado primero dentro de las cabezas de sus empleados, y luego en los manuales, directorios, catálogos, patentes, informes finales, semestrales, ... que se generan continuamente. Un nuevo empleado puede andar buscando cierta información, pero tiene que preguntar a otras personas dónde puede encontrar esa información. Es probable que no se le refiera al documento existente, por ignorancia.

Usted desarrollará un bibliotecario, basado en Clasitex, que lea todos los documentos escritos en español de una empresa, y los indexe y catalogue según los temas que cada uno abarca, usando el árbol de conceptos del sentido común de CYC (o el simplificado de Clasitex) o el árbol de conceptos especializados de la empresa (árbol de pinturas, recubrimientos e impermeabilizantes, por ejemplo). Este bibliotecario podrá contestar preguntas provenientes de empleados o usuarios nuevos o poco familiarizados, del tipo "¿dónde puedo encontrar información sobre pinturas alquídicas?" El bibliotecario hallará los documentos pertinentes, aunque uno de ellos se refiera, digamos, a "recubrimientos fenólicos", debido a que posee el árbol de conocimientos específicos.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Mucha de la inversión valiosa de una empresa de alta tecnología yace en su "conocimiento", que se encuentra en reportes finales, catálogos, descripciones de productos, artículos técnicos, manuales de usuario, y otros documentos escritos en español. Una persona poco familiarizada con los términos o los artículos de una empresa puede tener dificultades para acceder información valiosa, debido a que no sabe dónde buscar, hay mucho dónde buscar, y no le orientan adecuadamente. El bibliotecario que usted desarrollará pondrá fin a este problema, y le ahorrará mucho dinero a la empresa, al sistematizar el acceso y uso a la información técnica generada en lenguaje natural.

Many enterprises downsize to adapt to more competitive environments. But unless they have captured the knowledge of their employees, downsizing can result in a loss of critical information. Similarly, as employees leave, organizations are likely to lose access to large quantities of critical knowledge. And as companies expand internationally, geographic barriers can affect knowledge exchange and prevent easy access to information. These and other forces are pushing enterprises to explore better methods for knowledge management.

Enterprise knowledge management entails formally managing knowledge resources, typically by using advanced information technology. KM is formal y that knowledge is classified and categorized according to a prespecified –but evolving– ontology into structured and semistructured data and knowledge accessible and reusable to the enterprise.

The business world is becoming so concerned about knowledge management that, according to one report, over 40 percent of the Fortune 1000 now have a chief knowledge officer, a senior-level executive responsible for creating an infrastructure and cultural envi-

ronment for knowledge sharing. This article surveys some components of this young field [1].

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. Utilice Clasitex. Vea los diferentes bibliotecarios de mi artículo [La computación en la inter-ciencia].

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

[1] Daniel E. O'Leary. Enterprise knowledge management. *Computer*, March 1998, pp 54-61.

98.<sup>7</sup> Guzmán A. Hallando los temas principales en un artículo en español. *Soluciones Avanzadas*. Vol. 5, núm. 45, pág. 58. I parte, 15 de Julio de 1997, II parte vol. 5, núm. 49, pág. 66, 15 de septiembre de 1997. También en: *Simposium Internacional de Computación*. Centro de Investigación en Computación. Instituto Politécnico Nacional. Noviembre 12-14, 1997. México, D.F., páginas 36-51.

99. Adolfo Guzmán. Finding the main themes in a Spanish document. *Journal Expert Systems with Applications*, Vol. 14, No.1/2, Jan/Feb 1998, pages 139-148.

101. Beatriz Beltrán Martínez, Adolfo Guzmán Arenas, Francisco Martínez Trinidad, José Ruiz Shulcloper. Clasitex++: una herramienta para el análisis de textos. Memorias del *Tercer Taller Iberoamericano de Reconocimiento de Patrones*, TIARP-98, Centro de Investigación en Computación, Instituto Politécnico Nacional, México, D. F. Marzo 1998. Páginas 369-379

Adolfo Guzmán. La computación en la inter-ciencia. Reporte técnico No. xx, Centro de Investigación en Computación, Instituto Politécnico Nacional, abril 1998.

#### 44. Expansor de materiales

OTROS NOMBRES. Cálculo de costos globales a partir de costos unitarios.

ÁREA (dentro de la Computación). Estructuras de datos. Aplicaciones.

DESCRIPCIÓN. En qué consiste el trabajo. Un metro cuadrado de pared está compuesto de  $x$  ladrillos, de  $z$  gramos de mezcla, de  $w$  minutos de albañil, etc. Esta composición se define en una tabla. A su vez, un cuarto de baño tipo "A" está formada de 4 paredes, totalizando  $k$  metros cuadrados, de  $m$  metros cuadrados de azulejos, de un espejo, un lavabo, dos llaves, una regadera, etc. Hay otros tipos cuartos de baño "B", "C", etc., más lujosos o amplios. A su vez, un departamento tipo "Perla" está formado por dos recámaras de tal tamaño, una sala de  $x$  tamaño, un comedor de  $w$  tamaño, etc. Hay otros tipos de departamentos más lujosos o más económicos, llamados "Condorito", "Saúz", "Balancín", y "Soltero". A su vez, un edificio tipo "3" está formado de  $x$  pisos, y cada piso tiene dos departamentos tipo "Perla" y tres tipo "Soltero". Hay otros tipos de edificios, llamados "4" y "5", con composiciones distintas. Se requiere un software que ayude al arquitecto a costear los distintos tipos de edificio que él construye, tomando en cuenta los precios unitarios: de cada ladrillo, de cada minuto (u hora) de albañil, de cada azulejo, etc. Esto es importante porque los precios fluctúan. Las composiciones no varían.

<sup>7</sup> Estos números se refieren a la lista de trabajos de A. Guzmán que aparece en su curriculum vitae.



ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Para calcular propiedades globales (costos, pesos, volúmenes, ...) a partir de propiedades individuales, cuando se tiene el árbol de composición (el árbol que describe la relación "es parte de").

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. No pre-alambre el árbol dentro de su progama. Utilice archivos o tablas para almacenar quién es parte de qué. Las partes cambian. También use tablas para guardar las propiedades (costos, digamos) de cada parte atómica (indivisible).

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

#### 45. ¿Dónde venden más barato?

OTROS NOMBRES. Centro exportador virtual. Centro vendedor virtual. Centro distribuidor virtual.

ÁREA (dentro de la Computación). Sistemas distribuidos. ¿Comercio electrónico?

DESCRIPCIÓN. En qué consiste el trabajo.

Los vendedores de La Merced ya tienen un sistema para anunciar sus precios (verduras, frutas y abarrotes) por medio del web. Con él, un cliente puede visitar las páginas web donde se ven los precios de los distintos productos, tienda por tienda. (Las páginas son distintas de una tienda a otra, pero la base de datos es la misma, es el mismo diseño, el mismo software). Unas páginas tienen fotos de lechugas, tomates. Otras pueden tomar órdenes directamente (usted llena su pedido). La Unión de vendedores le ha pedido a usted construir un sistema S, que se dará o rentará a amas de casa, para:

1. Captar por Internet los pedidos del ama de casa. Ella puede decir: prefiero Ron Bacardí añejo botella de a litro, pero si no hay, que sea Blanco, y si no hay, entonces déme añejo de  $\frac{3}{4}$  de litro. Puede decir: quiero un kilo de jamón que esté en oferta. Puede decir: quiero un kilo de salchichas buenas pero que no pase de 40 pesos el kilo. Puede especificar por marca pero omitir detalles (2 paquetes de Pan Bimbo, pero no dije si blanco o integral, y no dije si mediano o grande).

2. Decidir dónde comprar cada renglón de la lista. El sistema S determina dónde venden más barato lo que a ella le interesa comprar. El modo normal es "compra donde venden más barato". Sin embargo, el ama de casa puede decir "compra siempre leche en Lechería Minatitlán, excepto si está 20% más caro que el más barato". También las existencias se van consumiendo y pudiera ser que a la tienda que vende más barato frijol negro ya se le acabó.

2. S parte un pedido en varios sub-pedidos a las diferentes tiendas: leche en el Local 14, tomates y naranjas en el Local La Margarita, jabón Octagón y Detergente Ariel en Abarrotes Manzur, ... Las tiendas tienen un contrato con una empresa de camionetas repartidoras, la que entrega a domicilio en las 2 horas siguientes (promete tardarse máximo 2 horas desde que llegó completo el pedido de Juana Pérez hasta que se lo entrega en la puerta de su casa).

3. Las tiendas imponen un "castigo" (10% de sobreprecio) si un sub-pedido no llega a 100 pesos, y se niegan a surtir sub-pedidos menores de 20 pesos. La forma de trabajar es: un ama de casa escribe por Internet su pedido, haciendo algunos señalamientos de dónde comprar, o dejando que sea lo más barato, o "súrtame del último lugar que me surtió", o "donde ustedes quieran", o "Primero con Don Pancho, luego con Don Luis, luego con La Luz". El sistema divide el pedido en varios sub-pedidos, a diferentes tiendas. Cada tienda llena una

bolsa con la mercancía, le pone la etiqueta (impresa por S, obviamente) "Entrega de la tienda La Luz para la Sra. Juana Pérez" y la va a dejar a un local donde salen camionetas repartidoras. Otras tiendas entregan en el mismo local otros sub-pedidos para Juana Pérez. Finalmente todos los sub-pedidos de Juana Pérez se han surtido (la empresa de entregas no se espera mucho, y prefiere entregar un pedido incompleto a tiempo que uno completo con varias horas de retraso), y la camioneta sale a repartir mercancía (obviamente, no nadamás la de Juana). El tiempo máximo de espera desde que Juana hizo su pedido hasta que sale la camioneta con su mercancía es de 3 horas. A esto hay que agregar el tiempo que se tarde en llegar a la casa de Juana Pérez.

El sistema S va a trabajar por Internet. El ama de casa desde su computadora escribe una lista de compras (su pedido), y S hace posible que se le entreguen. El ama de casa espera un rato y viene una camioneta repartidora con dos o tres cajas con su pedido, el ama le paga (un solo cheque, o efectivo), y la camioneta ya repartirá esa paga entre los distintos vendedores que le surtieron a ella. El ama de casa no se mueve de su computadora para efectuar sus compras semanales, excepto pararse para recibir la mercancía y pagarles.

No incluya cobranza ni manejo de efectivo o crédito. Ponga cuidado con las diferentes tablas de datos (por ejemplo, hay que recordar que Juana Pérez siempre ha comprado pan integral, por lo que si omite la clase de pan, desea integral. Si desea pan blanco, debe especificarlo) y cómo interactúa la información. Piense algo en la seguridad del sistema.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Para formar "centros comerciales" virtuales, con un solo punto de venta.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

#### 46. Compresión de imágenes fijas o de video para transmisión vía canales telefónicos angostos

OTROS NOMBRES. Hallar el lugar donde la imagen cambia más.

ÁREA (dentro de la Computación). Imágenes digitales.

DESCRIPCIÓN. En qué consiste el trabajo. Usted va a transmitir una fotografía (variante: una señal de video) en blanco y negro, tomada por una cámara pequeña, a través de un canal angosto (9,600 bauds, digamos). Debe transmitir tantas fotos (o cuadros, si es video) como es posible, aunque dudo que llegue a, por ejemplo, 16 por segundo. Para esto, debe usar alguna técnica para transmitir una gran cantidad de información por un canal estrecho. Sugiero uno de (a), (b) ó (c). (a) compresión de imágenes. (b) análisis de la imagen, para ver qué partes son fijas, y cuáles se mueven. Entonces transmita un código diciendo: las partes x, y, z son fijas, y solo transmito "que no han cambiado." Pero la parte w ya cambió, y aquí va la nueva versión de w. Problema con (b): como las cosas se mueven, w puede "invadir" b, w puede desaparecer o achicarse, etc. (c) hallazgo de los contornos. Procure transmitir los contornos de la imagen, y utilice el ancho de banda restante para transmitir las partes de la imagen que ya cambiaron mucho y que no se pueden recobrar de la versión anterior.

Usted puede suponer que cerca de la cámara hay una computadora para hacer sus compresiones, y cerca del receptor otra computadora, para recrear o regenerar la imagen.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Hará posible la transmisión de imágenes por canales baratos.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos.

1. Método del Dr. José Luis Marroquín para hallar contornos. <http://www.cimat.mx>.

#### 47. Cortador de mangos sazones

**OTROS NOMBRES (títulos alternos).**

**ÁREA (dentro de la Computación).** Hardware, metrología.

**DESCRIPCIÓN.** En qué consiste el trabajo. Una persona que corta mangos usa una vara de unos 5 m, en cuya extremidad está una canastilla y una tijera. Él usa su vista y criterio para decidir si el mango está 1/4, 1/2, 3/4 ó maduro. Él debe cortar solo los mangos con cierto grado de madurez (los 1/2, digamos), pero su vista y criterio le fallan y corta mangos en el árbol con distinto grado de madurez, originando que se envíen (a lugares remotos) mangos que unos llegarán buenos, otros ya muy maduros, ocasionando mermas. Usted va a diseñar un aparato que no pese mucho, que trabaje con pilas, colocado en la canastilla. Este aparato debe medir el grado de madurez del mango, sin tocarlo o tocándolo ligeramente (con la fuerza que se produce cuando un mango cae 10 centímetros), y debe señalar (con una luz verde, o un silbido) si el mango debe cortarse o no. Usted puede usar cualquier método que diferencie los mangos. Por ejemplo, pruebe con ultrasonido, medir la resistividad, la conductancia, el brillo de la cáscara, la resonancia a cierta frecuencia o vibración, la absorción de microondas, de rayos láser, etc. Pero: (a) debe discriminar, y (b) debe poderse montar en la vara (poco peso).

Variante: su primera versión puede ser un aparato que vaya montado en la línea de ensamble, donde existe un transportador que acarrea los mangos (para su separación a mano, digamos), pero ahora usted va a montar ahí su dispositivo para separarlos con él. No diseñe la parte mecánica de este separador, concéntrese en la parte electrónica o de detección.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Problema real, de gran aplicación.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo.

Experimente con muchas formas de medir propiedades, viendo cuál discrimina satisfactoriamente mangos con distinta madurez. Tenga a mano mangos con distinta madurez.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Ya hay medidores de grado de madurez de melones, de aguacates.

#### 48. Traducción de taquigrafía al español

**OTROS NOMBRES (títulos alternos).**

**ÁREA (dentro de la Computación).** Caso 1: Análisis de dibujos en tiempo real. Caso 2: análisis de imágenes.

**DESCRIPCIÓN.** En qué consiste el trabajo. Haga un programa que traduzca de taquigrafía Pittman (o Gregg) al español.

**ANTECEDENTES, MOTIVACIÓN.** Para qué sirve, cuál es su finalidad. Supongo que no tiene gran aplicación práctica.

**POSIBLE DESARROLLO.** Una de las maneras en que (tal vez) se pueda llevar a cabo. La taquigrafía Pittman tiene trazos claros y oscuros. Cada trazo representa un fonema, un sonido. Por ejemplo, una barra vertical representa el sonido “de” (o “te”, ya no me acuerdo). También es importante la colocación de los trazos, si están encima de la raya del block de taquigrafía, o sobre ella, o debajo de ella. Tenemos dos casos para pasar de Pittman al español:

**CASO 1.** Se toma la taquigrafía en tiempo real, y en ese momento hay que producir el correspondiente texto en español (para simplificar las cosas, digamos que hay que producir un archivo con caracteres Ascii que contiene las palabras o dictado en español correspondiente a los trazos que alguien va escribiendo en la pantalla). Es decir, en el momento en que alguien está escribiendo los trazos (palabras) en Pittman, la computadora va observando cómo se van escribiendo o dibujando estos trazos. Por ejemplo, el escritor de taquigrafía podría usar el ratón (o un dispositivo similar, una pluma electrónica) para irle dibujando a la computadora los trazos en taquigrafía. La computadora tiene acceso a la secuencia de cómo se fueron dibujando los trazos sobre el papel. En este caso, la solución es más o menos sencilla. Se divide el área donde la persona está escribiendo los trazos, en una retícula de digamos 3 por 4 (o sea, 12 cuadrados o rectángulos) y, conforme va dibujando sus trazos, la computadora observa que el trazo es vertical (porque va del cuadrado A1 al b1 y luego al C1, por ejemplo). Es decir, la hoja de papel queda dividida así:

A1 A2 A3 A4  
B 1 B2 B3 B4  
C1 C2 C3 C4

En cambio, un trazo horizontal comprendería los cuadros B1, B2 y B3, digamos. Luego, hay que determinar si el trazo fue dibujado sobre la raya roja del block de taquigrafía, o debajo de la raya. Esto es fácil, porque digamos que la raya aparece en medio de B1 B2 B3 B4, por lo que los rectángulos superiores son A1 A2 A3 A4, y los que aparecen debajo de la raya son C1 C2 C3 C4. Nótese que no estamos usando una cámara de televisión para ver cómo va dibujando el escribano sus trazos, sino que estamos utilizando el dispositivo de dibujo (el ratón, o una pluma que funciona como el ratón) como dispositivo de entrada a la computadora.

Luego, hay que determinar si el trazo es ligero (claro) o fuerte (oscuro, o grueso). Aquí, habría que ver si el dispositivo de entrada tiene manera de medir la presión que se hace sobre él (el ratón no tiene esa manera). Si se tiene, entonces el dispositivo de entrada nos dirá: trazo ligero o trazo oscuro (con bastante presión). Si no se tiene, sospecho que los trazos fuertes son más lentos que los trazos ligeros, se dibujen con mayor lentitud. Habría que hacer mediciones. Y entonces, como el dispositivo de entrada nos da la dinámica del dibujo, nos dice si el trazo B1 B2 B3 B4 fue dibujado con rapidez (tardó 120 milisegundos en pasar de B1 a B4) y por consiguiente es trazo ligero, o fue lento (se tomó 200 milisegundos de pasar de B1 a B4) y por consiguiente es oscuro.

Una vez hecho lo anterior, tenemos un algoritmo que nos pasa de dibujos hechos en Pittman a “palabras” o sílabas del tipo “de” “te” “te” en vez de “detente” o “deste-



te". Hecho lo cual, podemos pulir el algoritmo aplicando otras reglas de la taquigrafía, que nos dice que si un trazo termina en un caracolito hacia arriba, significa "ente", si es hacia abajo, significa "ando", etc. [Estoy inventando estas reglas, ya se me olvidó la taquigrafía Pittman].

Por lo que vamos puliendo nuestro traductor. Finalmente, podemos ayudarlo teniendo un diccionario ya armado de palabras válidas en español, conteniendo "andando", "andaba", pero no "andante", de manera que una vez que el reconocedor de Pittman hace su trabajo, el diccionario puede decirnos qué palabras en él son las que más se parecen a lo que el reconocedor encontró. Inclusive, se pueden usar técnicas de índices para meter la mano de manera rápida al diccionario. Es decir, armarlo en forma de un archivo donde la llave es una llave compuesta formada por los pedazos "de", "te", "ante", etc., que logra reconocer el reconocedor Pittman.

Luego de haber pasado por el filtro del diccionario de palabras válidas, podemos pulir más nuestro algoritmo utilizando otras reglas del español. Por ejemplo, el adjetivo casa en número y género con el sustantivo. De manera que si no estamos seguros que un trazo diga "blanco" o "blancas", nos fijamos en las palabras anteriores, y si decía "las palomas ..." podemos concluir que lo que sigue son "blancas." Es decir, se usa un programa que conozca la sintaxis y las principales reglas de sintaxis del español.

Podemos hacerle otros embellecimientos. Una vez que hemos encontrado cinco veces el trazo (taquigráfico) que corresponde a "caballo", podemos insertarlo en nuestro diccionario de palabras válidas en español, de manera que el programa aprende a reconocer la manera peculiar en que este escribano escribe (dibuja) la palabra caballo.

También ocurren en taquigrafía ciertas abreviaturas, que significan "atentamente" o "su seguro servidor". Estas formas peculiares de dibujar estas palabras se pueden meter al diccionario indexado.

**CASO 2.** La taquigrafía no es en tiempo real. Es decir, se muestran páginas escritas con anterioridad de taquigrafía Pittman, y la computadora debe traducirlas a archivos en Ascii. Aquí hay que usar una cámara de televisión o mejor, un scanner, para pasar los dibujos a imágenes, y analizar éstas. Se procede como en el Caso 1, pero se tiene menos información, ya que no se tiene la dinámica del dibujo. No se sabe si un trazo vertical se hizo de arriba hacia abajo o de abajo hacia arriba. Ni si se hizo rápida o lentamente. Pero los mismos procedimientos o algoritmos que en el Caso 1 se pueden usar, básicamente. Para notar si un trazo es negro (fuerte) o gris (ligero) se usa el color o intensidad de gris dado directamente por la imagen. Asugiero que en la imagen se utilicen tonos grises (no imágenes binarias simplemente), y que también se capten las rayas del bloc de taquigrafía.

**TRABAJOS RELACIONADOS Y REFERENCIAS.** Trabajos previos. Ver proyecto 49.

#### 49. Traducción de español a taquigrafía

**OTROS NOMBRES (títulos alternos).**

ÁREA (dentro de la Computación). Generación de gráficas, despliegue.

DESCRIPCIÓN. En qué consiste el trabajo. Pasar de mensajes escritos (caso 1) o hablados (caso 2) en español, a taquigrafía Pittman.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad. Supongo que no tiene grandes aplicaciones prácticas.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo. La taquigrafía Pittman tiene trazos claros y oscuros. Cada trazo representa un fonema, un sonido. Por ejemplo, una barra vertical representa el sonido "de" (o "te", ya no me acuerdo). También es importante la colocación de los trazos, si están encima de la raya del block de taquigrafía, o sobre ella, o debajo de ella. Ahora bien, distingo dos casos en el problema de pasar de español a Pittman.

CASO 1. Las oraciones en español están en un archivo de texto leíble por computadora, en Ascii. Este problema es sencillo. Si vemos en el archivo los caracteres "perro", podemos generar el dibujo en Pittman que corresponda a tal palabra, simplemente consultando el diccionario del caso 1 del problema 48. Es decir, usar un diccionario que tiene dos columnas, la primera es la palabra en español: "perro", "gato", "accidente", etc. La segunda columna es el dibujo o trazo en Pittman que corresponde a la palabra de la primera columna. Se va recorriendo el archivo de entrada (que está en ascii, en español), y se va viendo lo que contiene. "La locomotora echa humo." Nos vamos metiendo al archivo (diccionario) usando como llaves "la", "locomotora", "echa", "humo". Cada vez que nos metemos al archivo, sacamos el dibujo o trazo de la palabra correspondiente. En el ejemplo anterior, habríamos sacado cuatro trazos, que corresponden a cuatro palabras en Pittman. Y las dibujamos en un archivo (imagen) de salida.

CASO 2. Se tienen oraciones en español, son sonidos emitidos por una garganta humana. Este problema es difícil, pues hay que entender primero lo que ese sonido dice, a qué palabra de español se refiere. (Pasarlo a Ascii. Idea: use el paquete Dragon que hace esto). Una vez que se tiene el sonido convertido a palabra en español en ascii, nos pasamos al Caso 1. El problema de pasar de sonidos en español a texto escrito en español es que depende del orador (hay muchas maneras de decir o pronunciar la palabra "presidente", depende del acento del hablante, si es mujer, si es sureña, si tiene gripa, si está exaltada, si le falta el aire por venir corriendo, etc.)

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos. Ver proyecto 48.

## 50. Otro trabajo

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación).

DESCRIPCIÓN. En qué consiste el trabajo.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

## 51. Otro trabajo

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación).

DESCRIPCIÓN. En qué consiste el trabajo.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.

## 52. Otro trabajo

OTROS NOMBRES (títulos alternos).

ÁREA (dentro de la Computación).

DESCRIPCIÓN. En qué consiste el trabajo.

ANTECEDENTES, MOTIVACIÓN. Para qué sirve, cuál es su finalidad.

POSIBLE DESARROLLO. Una de las maneras en que (tal vez) se pueda llevar a cabo.

TRABAJOS RELACIONADOS Y REFERENCIAS. Trabajos previos.