

Data Warehouse and Data Mining

José A. Royo

<http://www.cps.unizar.es/~jaroyo>

email: joalroyo@unizar.es

Departamento de Informática e Ingeniería de Sistemas

¿Por qué DW y DM?

- Mayor poder de procesamiento y sofisticación de herramientas
- Demanda de mejora del acceso a datos
- Necesidad de información para la toma de decisiones
- Recopilación de información \Rightarrow Alto Coste

Data Warehouse

Información en las empresas

- La Información proviene de fuentes internas (sistemas de producción) y externas (hasta un 20%)
- Problemas
 - Saturación de información
 - Difícil de acceder
 - No selectiva
- La información se necesita para:
 - Competir (comparación con otros productos)
 - Personalizar (simular que cada cliente es único)

Data Warehouse

- **Definición:** colección de datos orientados al tema, integrados, no volátiles e historiadados, organizados para el apoyo de un proceso de ayuda a la decisión
- Se guarda toda la información útil (proveniente de varias fuentes) en un único lugar

Data Warehouse

- **Orientación al tema**
 - Disponer de toda la información sobre un tema
 - No organizar los datos según los procesos funcionales
 - La información común a varios temas no debe duplicarse
 - Los Data Mart apoyan la orientación al tema
 - BD orientada al tema puesta a disposición de los usuarios en un contexto de decisión descentralizado

Data Warehouse

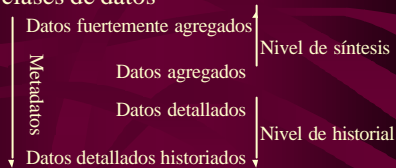
- **Datos integrados:**
 - Los datos deben formatearse y unificarse para llegar a un estado coherente
 - Ej.: consolidar todas las informaciones respecto a un cliente
- **Datos historiadados**
 - Los datos no se actualizan nunca
 - representan un valor en un momento concreto
 - Los datos se referencian temporalmente

Data Warehouse

- **Datos no volátiles:**
 - Consecuencia de la historización
 - = consulta en = periodo \Rightarrow = resultado
 - Sistemas de producción no volátiles
- **Infocentro:**
 - similar al Data Warehouse pero centrado en el sistema de producción

Data Warehouse: Estructura

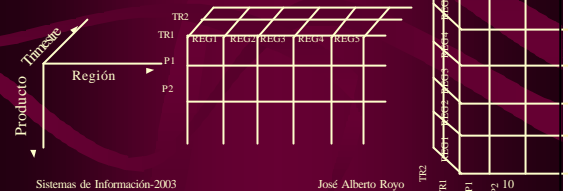
- Varia clases de datos



- Estructura multidimensional

Datos

- Matrices Multidimensionales o hipercubos
 - Ej.: periodo impositivo, producto y región
 - Pivotación: cambio de orientación de los ejes



Datos: Representaciones Jerárquicas

- Exploración ascendente (roll-up)
 - Agrupación de datos
 - Ej.: Agrupar datos mensuales en trimestrales
- Exploración Descendente (drill-down)
 - División de datos
 - Ej.: Ventas nacionales \Rightarrow provinciales

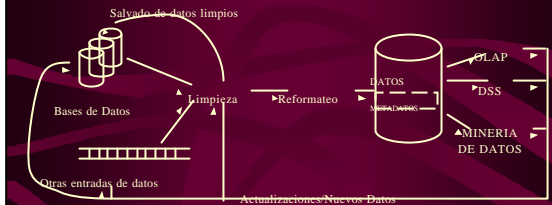
Datos: Representación física

- 2 Tablas
 - Tabla de dimensiones
 - Tabla de hechos
- Tipos de esquema
 - Esquema de estrella:
 - Tabla de hechos con una única tabla para cada dimensión
 - Esquema de copos
 - Tablas dimensionales organizadas jerárquicamente

Construcción de un Data Warehouse

- Adquisición: Recopilar información de varias fuentes y unificarla
 - Extracción
 - Preparación (formateo y limpieza)
 - Carga
- Almacenamiento: basado en un SGBD
 - El historial influirá en la estructura física
- Acceso: distintos grupos de usuarios requerirán distintas consultas

DW: Características



OLTP y OLAP

- OLTP (On-Line Transactionnel Processing): entorno donde las respuestas se darán en un tiempo aceptable y serán consistentes
 - Transacciones predeterminadas
 - Utiliza pocas tablas
- OLAP (On-Line Analytical Processing): entorno de ayuda a la decisión (análisis de datos)
 - Transacciones muy variadas
 - Manejan volúmenes grandes de datos (+tiempo)
 - Se relacionan datos aparentemente sin relación

DSS: *Decision Support Systems*

- EIS (*Executive Information Systems* o sistemas de información ejecutiva)
- Ayuda para la toma de decisiones
- Dan datos de nivel superior
- DW preparado para OLAP y DSS

Factores de éxito del DW

- Integra datos de producción con datos externos y gestiona historiales
- Contiene la información útil
- Los datos son coherentes, actualizados y documentados (calidad)
- Ofrece acceso directo a los usuarios
- Aumenta el número de accesos
- Da una flexibilidad que apoya el crecimiento
 - De usuarios, herramientas y volumen

Diferencias entre DW y vistas

- DW son un almacenamiento permanente
 - Vistas Construidas cuando es necesario
- DW son multidimensionales
 - Vistas suelen ser relacionales
- DW son indexados para optimizar su rendimiento
 - Vistas son indexadas dependiendo de la BD subyacente
- DW dan unas funcionalidades específicas
 - Las vistas no
- DW poseen grandes cantidades de datos integrados y temporales
 - Vistas son extractos de la BD

Errores a evitar

- Cargar datos solamente porque están disponibles (podrían ser no útiles)
- Crear el esquema de la BD de forma tradicional
- Crearlo pensando en la tecnología usada
- Concentrarse en los datos internos
- Creer que los problemas acaban una vez instalado el Data Warehouse

Implementación: Dificultades

- Alto coste
- Urgente ayuda para la toma de decisiones
 - *Data Mart*
- Gran coste de mantenimiento
 - Cambio de necesidades
 - Nuevas fuentes de datos
 - Cambio de la capacidad
 - Cambio de tecnología
- Control de calidad de los datos
- Heterogeneidad e Integración de datos

Bibliografía

- J.M. Franco. EDS-Institut Prométhéus, “El Data Warehouse. El Data Mining”, Eyrolles,1997.

Data Mining

Data Mining

- Búsqueda de información relevante (conocimiento) en grandes volúmenes de datos
- Descubrir de forma automática las reglas estadísticas y pautas de un conjunto de datos
- ¿Diferencia con *machine learning*? Grandes volúmenes de datos grabados en disco
- Objetivo obtener un conjunto de reglas

Data Mining: Visión General

- Data Mining y Data Warehouse
 - Búsqueda automática de relaciones
 - Extracción de patrones
- Descubrimiento de conocimiento (Fases)
 - Selección de datos
 - Limpieza de datos
 - Transformación/Codificación de datos
 - Minería de Datos
 - Presentación visualización

Resultados del Data Mining

- Descubrir
 - Reglas de asociación
 - Patrones secuenciales
 - Árboles de clasificación
- Presentación de Resultados
 - Listas
 - Representaciones Gráficas
 - Tablas resumen

Objetivos del Data Mining

- Predicción
 - Ej.: que compraran los clientes bajo determinados descuentos
- Identificación
 - Ej.: secuencia de nucleótidos \rightarrow presencia gen
- Clasificación
 - Ej.: clientes que buscan descuentos, fieles y ocasionales
- Optimización
 - utilización de recursos limitados: tiempo, espacio, dinero, etc.

Conocimiento Descubierto

- Reglas de asociación
 - Ej.: Compra bolso \rightarrow Compra zapatos
- Jerarquías de clasificación
 - Ej.: Clasificación de los clientes de un banco
- Patrones secuenciales
 - Ej.: Cámara digital \rightarrow Memorias MMC
- Patrones de series de tiempo
 - Ej.: Aumento de ventas de automóviles antes del verano
- Categorización y segmentación
 - Ej.: niños, jóvenes, adultos y jubilados

Tipos de reglas

- $\forall X$ antecedente \Rightarrow consecuente
 - X: lista de una o varias variables con rangos asociados
 - Ej.: \forall transacción T, compra(T,pan) \Rightarrow compra(T,leche)
- Rango de las variables \rightarrow población
- Soporte: porcentaje de la población que cumple el antecedente o el consecuente
- Confianza: porcentaje con que el consecuente es cierto al serlo el antecedente

Data Mining Automático

- Descubrimiento automático de reglas
- Técnicas de *machine learning*, adaptadas para grandes volúmenes de datos
- Tres tipos de problemas:
 - Clasificación: reglas que dividan en grupos
 - Asociaciones: $X \rightarrow Y$
 - Correlaciones entre series

Clasificación

- Comienza con una muestra de datos de clasificación conocida
- Los datos se dividen según uno de sus atributos, sucesivamente
 - Atributos enumerados \Rightarrow un conjunto por valor
 - Atributos con rango numérico \Rightarrow intervalos
- Resultado: árbol de clasificación (taxonomía)
- Hasta usar todos los atributos o clasificar correctamente los datos

Asociaciones

- Se genera un mapa de bits para cada transacción (un bit para cada artículo a estudiar)
- Nos quedamos con los artículos más adquiridos
- Se generan todos los subconjuntos posibles de artículos y se cuenta el número de transacciones
- Los subconjuntos con un número alto de transacciones generan las reglas

Asociaciones negativas

- Ej.: el 60% de los clientes que compran patatas fritas no compran agua mineral
- Si no combinación de elementos \rightarrow asociación negativa
 - Tenemos muchísimas reglas sin ningún interés
- Utilizar conocimiento previo del problema
- Optimizaciones
 - Muestreo

Data Mining Guiado por el Usuario

- El usuario plantea hipótesis
- El sistema comprueba si se verifica o no
- Las hipótesis se pueden ir refinando
- La visualización gráfica de datos ayuda al usuario a examinar grandes volúmenes de datos

Otras Técnicas

- Regresión
 - Estadística
 - Establecimiento de probabilidades
- Redes neuronales
 - Entrenar la red
 - Reconocer los patrones según el entrenamiento
- Algoritmos genéticos
 - Algoritmos probabilistas
 - Población inicial
 - Sobreviven los mejores
 - Obtenemos datos en cada iteración

Aplicaciones

- Marketing
 - Comportamiento del consumidor basado en patrones de compra
- Finanzas
 - Análisis de rendimiento de operaciones
 - Solvencia de clientes
 - Valoración de opciones de financiación
- Fabricación
 - Optimización de recursos: maquinaria, mano de obra, materiales
 - Optimización del proceso de fabricación

Bibliografía

- J.M. Franco. EDS-Institut Prométhéus, “El Data Warehouse. El Data Mining”, Eyrolles,1997.
- R.A. Elmasri, S.B. Navathe, “Fundamentos de Sistemas de Bases de Datos, 3ª ed.”, Addison-Wesley,2000.

Sistemas de Información Geográfica

José Alberto Royo

joalroyo@unizar.es

Departamento de Informática e Ingeniería de Sistemas

Sist. de Inf. Geográfica (GIS)

- Sistemas que relacionan, almacenan, manipulan y visualizan información referenciada geográficamente
- Sistemas de información que manejan datos espaciales
 - Algunos datos son referencias espaciales o coordenadas geográficas
 - Poseen operadores para manejar dichos datos espaciales

GIS: Categorías

1. Aplicaciones Cartográficas
 - Variedad de datos: características del suelo, densidad de cultivo, calidad del aire
 - Representación basada en campos
 - Superposición de capas
2. Aplicaciones para el modelado digital de terrenos
 - Variedad de datos
 - Representación basada en campos
3. Aplicaciones de Objetos geográficos
 - Objetos físicos: centrales eléctricas, hospitales, etc.
 - Representación basada en objetos

Información espacial

- Multidimensional (x,y,z,t,...)
- Voluminosidad
- Naturaleza inexacta (no hay representaciones exactas de la Tierra)
- Las preguntas combinan topología, geografía y otros atributos, con información aproximada
- Combinan distintos contextos legales y económicos (varían de un país a otro)

Mapas

- Estructuración
 - Vectores (+versátil, -fácil de crear)
 - Creados con paneles digitalizadores
 - *Rasters* (-versátil, +fácil de crear)
 - Cada celda almacena el tipo de terreno
 - Creados mediante *scanning*
 - Se puede pasar de un formato a otro (con un cierto error)
 - Reconocimiento de formas en un raster → vectores
 - Pixelización de vectores → raster

GIS: Operaciones sobre los Datos

- Interpolación
 - Obtención de datos de elevación no obtenidos en la muestra
- Interpretación
 - Cierre de polígonos
 - Definición, reducción y mejora de detalles
- Análisis de proximidad
 - Cálculo de zonas de interés
- Procesamiento de imágenes en una matriz de puntos
 - Integrar características geográficas en distintas capas
 - Análisis digital de imágenes

Otras funcionalidades

- Extensibilidad
 - Continua evolución de los sistemas GIS
- Control de calidad de los datos
 - ¿dónde estoy?
 - Problema debido a la variedad de tipos de datos
- Visualización
 1. Contorneado
 - Isolíneas
 2. Sombreado de montañas
 - Método de iluminación empleado para representar relieves
 3. Visualizaciones de perspectivas
 - Imágenes tridimensionales

Preguntas a los GIS

- ¿Qué hay en cierta posición?
- ¿Dónde hay cierto elemento?
- Muestra zonas que cuyos atributos cumplen ciertas condiciones
- Generación de nuevos gráficos
 - mapas de elevación, densidad de población, etc.

Utilidad de los GIS

- Generación de mapas
- Selección de lugares
- Creación de planes de emergencia
 - Ante terremotos u otras catástrofes
- Simulación de transformaciones medioambientales
 - Cambio en paisajes ante túneles, obras, urbanizaciones, etc.

GIS: Trabajo Futuro

- Nuevas arquitecturas:
 - Distribución de datos
 - Separación de datos espaciales y no espaciales
- Versionado
 - Ej.: ¿Qué pasa si construimos una autopista?
- Estándares de Datos
 - Compartir de datos
- Aplicaciones específicas
 - Tipos de datos distintos
- Ausencia de semántica en las estructuras de datos
 - Vías de sentido único

Bibliografía

- Keith C. Clarke, “Getting Started with GIS”, Prentice-Hall, 1997, ISBN 0-13-294-786-2.
- <http://www.usgs.gov/research/gis/title.html>

Data Warehouse and Data Mining

José A. Royo

<http://www.cps.unizar.es/~jaroyo>

email: joalroyo@unizar.es

Departamento de Informática e Ingeniería de Sistemas

¿Por qué DW y DM?

- Mayor poder de procesamiento y sofisticación de herramientas
- Demanda de mejora del acceso a datos
- Necesidad de información para la toma de decisiones
- Recopilación de información ⇒ Alto Coste

Data Warehouse

Información en las empresas

- La Información proviene de fuentes internas (sistemas de producción) y externas (hasta un 20%)
- Problemas
 - Saturación de información
 - Difícil de acceder
 - No selectiva
- La información se necesita para:
 - Competir (comparación con otros productos)
 - Personalizar (simular que cada cliente es único)

Data Warehouse

- **Definición:** colección de datos orientados al tema, integrados, no volátiles e historiadados, organizados para el apoyo de un proceso de ayuda a la decisión
- Se guarda toda la información útil (proveniente de varias fuentes) en un único lugar

Data Warehouse

- **Orientación al tema**
 - Disponer de toda la información sobre un tema
 - No organizar los datos según los procesos funcionales
 - La información común a varios temas no debe duplicarse
 - Los Data Mart apoyan la orientación al tema
 - BD orientada al tema puesta a disposición de los usuarios en un contexto de decisión descentralizado

Data Warehouse

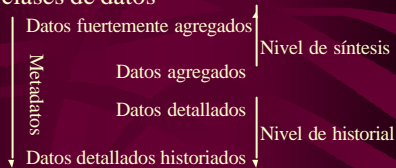
- **Datos integrados:**
 - Los datos deben formatearse y unificarse para llegar a un estado coherente
 - Ej.: consolidar todas las informaciones respecto a un cliente
- **Datos historiadados**
 - Los datos no se actualizan nunca
 - representan un valor en un momento concreto
 - Los datos se referencian temporalmente

Data Warehouse

- **Datos no volátiles:**
 - Consecuencia de la historización
 - = consulta en = periodo \Rightarrow = resultado
 - Sistemas de producción no volátiles
- **Infocentro:**
 - similar al Data Warehouse pero centrado en el sistema de producción

Data Warehouse: Estructura

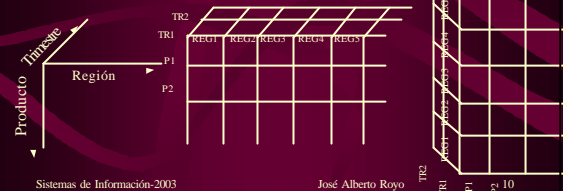
- Varia clases de datos



- Estructura multidimensional

Datos

- Matrices Multidimensionales o hipercubos
 - Ej.: periodo impositivo, producto y región
 - Pivotación: cambio de orientación de los ejes



Datos: Representaciones Jerárquicas

- Exploración ascendente (roll-up)
 - Agrupación de datos
 - Ej.: Agrupar datos mensuales en trimestrales
- Exploración Descendente (drill-down)
 - División de datos
 - Ej.: Ventas nacionales \Rightarrow provinciales

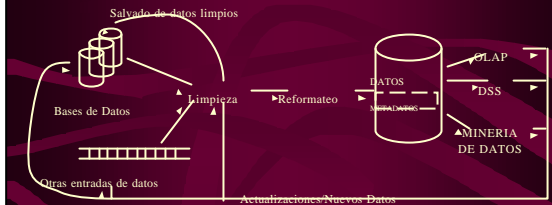
Datos: Representación física

- 2 Tablas
 - Tabla de dimensiones
 - Tabla de hechos
- Tipos de esquema
 - Esquema de estrella:
 - Tabla de hechos con una única tabla para cada dimensión
 - Esquema de copos
 - Tablas dimensionales organizadas jerárquicamente

Construcción de un Data Warehouse

- Adquisición: Recopilar información de varias fuentes y unificarla
 - Extracción
 - Preparación (formateo y limpieza)
 - Carga
- Almacenamiento: basado en un SGBD
 - El historial influirá en la estructura física
- Acceso: distintos grupos de usuarios requerirán distintas consultas

DW: Características



OLTP y OLAP

- OLTP (On-Line Transactionnel Processing): entorno donde las respuestas se darán en un tiempo aceptable y serán consistentes
 - Transacciones predeterminadas
 - Utiliza pocas tablas
- OLAP (On-Line Analytical Processing): entorno de ayuda a la decisión (análisis de datos)
 - Transacciones muy variadas
 - Manejan volúmenes grandes de datos (+tiempo)
 - Se relacionan datos aparentemente sin relación

DSS: *Decision Support Systems*

- EIS (*Executive Information Systems* o sistemas de información ejecutiva)
- Ayuda para la toma de decisiones
- Dan datos de nivel superior
- DW preparado para OLAP y DSS

Factores de éxito del DW

- Integra datos de producción con datos externos y gestiona historiales
- Contiene la información útil
- Los datos son coherentes, actualizados y documentados (calidad)
- Ofrece acceso directo a los usuarios
- Aumenta el número de accesos
- Da una flexibilidad que apoya el crecimiento
 - De usuarios, herramientas y volumen

Diferencias entre DW y vistas

- DW son un almacenamiento permanente
 - Vistas Construidas cuando es necesario
- DW son multidimensionales
 - Vistas suelen ser relacionales
- DW son indexados para optimizar su rendimiento
 - Vistas son indexadas dependiendo de la BD subyacente
- DW dan unas funcionalidades específicas
 - Las vistas no
- DW poseen grandes cantidades de datos integrados y temporales
 - Vistas son extractos de la BD

Errores a evitar

- Cargar datos solamente porque están disponibles (podrían ser no útiles)
- Crear el esquema de la BD de forma tradicional
- Crearlo pensando en la tecnología usada
- Concentrarse en los datos internos
- Creer que los problemas acaban una vez instalado el Data Warehouse

Implementación: Dificultades

- Alto coste
- Urgente ayuda para la toma de decisiones
 - *Data Mart*
- Gran coste de mantenimiento
 - Cambio de necesidades
 - Nuevas fuentes de datos
 - Cambio de la capacidad
 - Cambio de tecnología
- Control de calidad de los datos
- Heterogeneidad e Integración de datos

Bibliografía

- J.M. Franco. EDS-Institut Prométhéus, “El Data Warehouse. El Data Mining”, Eyrolles,1997.

Data Mining

Data Mining

- Búsqueda de información relevante (conocimiento) en grandes volúmenes de datos
- Descubrir de forma automática las reglas estadísticas y pautas de un conjunto de datos
- ¿Diferencia con *machine learning*? Grandes volúmenes de datos grabados en disco
- Objetivo obtener un conjunto de reglas

Data Mining: Visión General

- Data Mining y Data Warehouse
 - Búsqueda automática de relaciones
 - Extracción de patrones
- Descubrimiento de conocimiento (Fases)
 - Selección de datos
 - Limpieza de datos
 - Transformación/Codificación de datos
 - Minería de Datos
 - Presentación visualización

Resultados del Data Mining

- Descubrir
 - Reglas de asociación
 - Patrones secuenciales
 - Árboles de clasificación
- Presentación de Resultados
 - Listas
 - Representaciones Gráficas
 - Tablas resumen

Objetivos del Data Mining

- Predicción
 - Ej.: que compraran los clientes bajo determinados descuentos
- Identificación
 - Ej.: secuencia de nucleótidos \rightarrow presencia gen
- Clasificación
 - Ej.: clientes que buscan descuentos, fieles y ocasionales
- Optimización
 - utilización de recursos limitados: tiempo, espacio, dinero, etc.

Conocimiento Descubierto

- Reglas de asociación
 - Ej.: Compra bolso \rightarrow Compra zapatos
- Jerarquías de clasificación
 - Ej.: Clasificación de los clientes de un banco
- Patrones secuenciales
 - Ej.: Cámara digital \rightarrow Memorias MMC
- Patrones de series de tiempo
 - Ej.: Aumento de ventas de automóviles antes del verano
- Categorización y segmentación
 - Ej.: niños, jóvenes, adultos y jubilados

Tipos de reglas

- $\forall X$ antecedente \Rightarrow consecuente
 - X: lista de una o varias variables con rangos asociados
 - Ej.: \forall transacción T, compra(T,pan) \Rightarrow compra(T,leche)
- Rango de las variables \rightarrow población
- Soporte: porcentaje de la población que cumple el antecedente o el consecuente
- Confianza: porcentaje con que el consecuente es cierto al serlo el antecedente

Data Mining Automático

- Descubrimiento automático de reglas
- Técnicas de *machine learning*, adaptadas para grandes volúmenes de datos
- Tres tipos de problemas:
 - Clasificación: reglas que dividan en grupos
 - Asociaciones: $X \rightarrow Y$
 - Correlaciones entre series

Clasificación

- Comienza con una muestra de datos de clasificación conocida
- Los datos se dividen según uno de sus atributos, sucesivamente
 - Atributos enumerados \Rightarrow un conjunto por valor
 - Atributos con rango numérico \Rightarrow intervalos
- Resultado: árbol de clasificación (taxonomía)
- Hasta usar todos los atributos o clasificar correctamente los datos

Asociaciones

- Se genera un mapa de bits para cada transacción (un bit para cada artículo a estudiar)
- Nos quedamos con los artículos más adquiridos
- Se generan todos los subconjuntos posibles de artículos y se cuenta el número de transacciones
- Los subconjuntos con un número alto de transacciones generan las reglas

Asociaciones negativas

- Ej.: el 60% de los clientes que compran patatas fritas no compran agua mineral
- Si no combinación de elementos \rightarrow asociación negativa
 - Tenemos muchísimas reglas sin ningún interés
- Utilizar conocimiento previo del problema
- Optimizaciones
 - Muestreo

Data Mining Guiado por el Usuario

- El usuario plantea hipótesis
- El sistema comprueba si se verifica o no
- Las hipótesis se pueden ir refinando
- La visualización gráfica de datos ayuda al usuario a examinar grandes volúmenes de datos

Otras Técnicas

- Regresión
 - Estadística
 - Establecimiento de probabilidades
- Redes neuronales
 - Entrenar la red
 - Reconocer los patrones según el entrenamiento
- Algoritmos genéticos
 - Algoritmos probabilistas
 - Población inicial
 - Sobreviven los mejores
 - Obtenemos datos en cada iteración

Aplicaciones

- Marketing
 - Comportamiento del consumidor basado en patrones de compra
- Finanzas
 - Análisis de rendimiento de operaciones
 - Solvencia de clientes
 - Valoración de opciones de financiación
- Fabricación
 - Optimización de recursos: maquinaria, mano de obra, materiales
 - Optimización del proceso de fabricación

Bibliografía

- J.M. Franco. EDS-Institut Prométhéus, “El Data Warehouse. El Data Mining”, Eyrolles,1997.
- R.A. Elmasri, S.B. Navathe, “Fundamentos de Sistemas de Bases de Datos, 3ª ed.”, Addison-Wesley,2000.

Sistemas de Información Geográfica

José Alberto Royo

joalroyo@unizar.es

Departamento de Informática e Ingeniería de Sistemas

Sist. de Inf. Geográfica (GIS)

- Sistemas que relacionan, almacenan, manipulan y visualizan información referenciada geográficamente
- Sistemas de información que manejan datos espaciales
 - Algunos datos son referencias espaciales o coordenadas geográficas
 - Poseen operadores para manejar dichos datos espaciales

GIS: Categorías

1. Aplicaciones Cartográficas
 - Variedad de datos: características del suelo, densidad de cultivo, calidad del aire
 - Representación basada en campos
 - Superposición de capas
2. Aplicaciones para el modelado digital de terrenos
 - Variedad de datos
 - Representación basada en campos
3. Aplicaciones de Objetos geográficos
 - Objetos físicos: centrales eléctricas, hospitales, etc.
 - Representación basada en objetos

Información espacial

- Multidimensional (x,y,z,t,...)
- Voluminosidad
- Naturaleza inexacta (no hay representaciones exactas de la Tierra)
- Las preguntas combinan topología, geografía y otros atributos, con información aproximada
- Combinan distintos contextos legales y económicos (varían de un país a otro)

Mapas

- Estructuración
 - Vectores (+versátil, -fácil de crear)
 - Creados con paneles digitalizadores
 - *Rasters* (-versátil, +fácil de crear)
 - Cada celda almacena el tipo de terreno
 - Creados mediante *scanning*
 - Se puede pasar de un formato a otro (con un cierto error)
 - Reconocimiento de formas en un raster → vectores
 - Pixelización de vectores → raster

GIS: Operaciones sobre los Datos

- Interpolación
 - Obtención de datos de elevación no obtenidos en la muestra
- Interpretación
 - Cierre de polígonos
 - Definición, reducción y mejora de detalles
- Análisis de proximidad
 - Cálculo de zonas de interés
- Procesamiento de imágenes en una matriz de puntos
 - Integrar características geográficas en distintas capas
 - Análisis digital de imágenes

Otras funcionalidades

- Extensibilidad
 - Continua evolución de los sistemas GIS
- Control de calidad de los datos
 - ¿dónde estoy?
 - Problema debido a la variedad de tipos de datos
- Visualización
 1. Contorneado
 - Isolíneas
 2. Sombreado de montañas
 - Método de iluminación empleado para representar relieves
 3. Visualizaciones de perspectivas
 - Imágenes tridimensionales

Preguntas a los GIS

- ¿Qué hay en cierta posición?
- ¿Dónde hay cierto elemento?
- Muestra zonas que cuyos atributos cumplen ciertas condiciones
- Generación de nuevos gráficos
 - mapas de elevación, densidad de población, etc.

Utilidad de los GIS

- Generación de mapas
- Selección de lugares
- Creación de planes de emergencia
 - Ante terremotos u otras catástrofes
- Simulación de transformaciones medioambientales
 - Cambio en paisajes ante túneles, obras, urbanizaciones, etc.

GIS: Trabajo Futuro

- Nuevas arquitecturas:
 - Distribución de datos
 - Separación de datos espaciales y no espaciales
- Versionado
 - Ej.: ¿Qué pasa si construimos una autopista?
- Estándares de Datos
 - Compartir de datos
- Aplicaciones específicas
 - Tipos de datos distintos
- Ausencia de semántica en las estructuras de datos
 - Vías de sentido único

Bibliografía

- Keith C. Clarke, “Getting Started with GIS”, Prentice-Hall, 1997, ISBN 0-13-294-786-2.
- <http://www.usgs.gov/research/gis/title.html>