

TRABAJO PRACTICO N° 1



Asignatura: BASES DE DATOS, Maestría en Ciencia de Datos, UDESA

Profesor: Gervasoni, Julio

Grupo: “BCD”

- Brosky, Brian
- Larrinaga, Carlos
- Suárez Gurruchaga, Carlos Roque
- Vigo, Daniel Eduardo

Tema: Modelado de una base de datos para la detección de deforestación para el Bosque Chaqueño Argentino.

Introducción y Objetivos

El objetivo de esta primera parte es que, dado un problema de mediana complejidad, los alumnos puedan implementar una solución utilizando las herramientas de modelado y diseño vistas desde el punto de vista lógico. La entrega deberá constar de la siguiente documentación:

- Introducción y explicación del problema a resolver.
- Modelo de Entidad Relación y Modelo Relacional derivado, utilizados para implementar la solución.
- Detalle de los supuestos asumidos para la resolución del problema.
- Conclusiones

Enunciado del Problema

Se debe modelar una base de datos para la detección de deforestación para el Bosque Chaqueño Argentino. Se debe registrar distinta información vinculada con este proceso.

El Gran Chaco es el segundo bosque más grande de América del Sur. En las últimas dos décadas este bosque ha experimentado un intenso proceso de deforestación, siendo una de las zonas más afectadas del planeta. Las consecuencias son múltiples y diversas: pérdida de biodiversidad, de servicios ecosistémicos como secuestro y almacenamiento de carbono, de provisión de agua, de la cultura indígena y de condiciones sanitarias.

Es una región geográfica ubicada principalmente en los territorios de Argentina, Paraguay y Bolivia, y una pequeña parte de Brasil. Se puede dividir en dos grandes ecorregiones, el Chaco Seco y el Chaco Húmedo. El Chaco Seco, a su vez, se puede dividir en Chaco Semiárido, Árido y Serrano.

Mediante información satelital se puede pixelar y obtener información de toda la región. El tamaño de píxel que se registra es de 6,25 hectáreas. Cada píxel debe ser asignado a su región correspondiente. Se debe guardar longitud y latitud del centro del píxel (coordenadas geográficas del píxel).

Para cada píxel se debe registrar su precipitación diaria y su temperatura acumulada diaria.

También el tipo de vegetación dominante. Estos pueden ser bosques mixtos xenófilos dominados por especies caducifolias o por especies semicaducifolias; *Aspidosperma* (quebracho-blanco); *Schinopsis lorentzii* (quebracho colorado) o humedal.

Cada región puede verse afectada por forzamientos antropogénicos y no antropogénicos. Esto también debe registrarse. El impulsor no antropogénico más importante es el clima, que puede perturbar las precipitaciones o las temperaturas. Los impulsores antropogénicos son la ganadería, la tala selectiva y la extracción de leña.

En la región hay una red de carreteras, compuesta por diferentes tipos de carreteras: camino principal, camino secundario, camino rural, etc. Etiquetando los píxeles que pertenecen a carreteras se puede tener también la información de la red vial.

Es importante registrar la cobertura del suelo. Las hay de distintas clases:

- Bosque: Terrenos que constituyen un ecosistema natural que presentan una cubierta arbórea.
- Otras tierras boscosas: Tierras que constituyen un ecosistema con palmeras y cañaverales.
- Otras tierras: Tierras no clasificadas como bosques u otras tierras boscosas
- Deforestada

Cada pixel además contiene un índice de vegetación.

Como las imágenes satelitales pueden no tener una calidad pareja, también se guarda la calidad general de los píxeles, agrupándolos en 5 categorías:

Sin Datos (Valor: -1), Buenos Datos (0), Datos Marginales (1), Nieve/Hielo (2) y Nublado (3).

Cada cierta cantidad de píxeles vecinos (una cantidad que puede ser arbitraria y muy variable) se le pueda etiquetar con lo que aparece más arriba como “consecuencias de la deforestación” (pérdida de biodiversidad, de provisión de agua, etc., etc.). Para cierto tratamiento es cómodo trabajar con estos conjuntos de píxeles como “regiones”, y de cada una de ellas guardar la info de a qué país pertenece, cuáles son las carreteras que las atraviesan y otro tipo de datos relevantes: si hay poblaciones, si tienen ríos o algún otro accidente geográfico relevante.

La información satelital puede proceder de distintos satélites que son gerenciados por empresas.

En toda la zona actúan diversas organizaciones de rewilding que buscan sanear los efectos de la deforestación introduciendo especies originales. Son ONGs que colaboran no solo con el trabajo de campo sino también aportando información de sus acciones en el bosque. Es decir que, si agregan especies vegetales, árboles lo notifican para que quede registro. Lo mismo si son testigos de forzamientos que impulsen la deforestación.

Importante: deben aparecer escritas en lenguaje natural todas las restricciones necesarias que no puedan ser expresadas en el diagrama entidad-relación.

Introducción

El Gran Chaco es el segundo bosque más grande de América del Sur. En las últimas dos décadas este bosque ha experimentado un intenso proceso de deforestación, siendo una de las zonas más afectadas del planeta.

Es por esto que es de suma importancia poder registrar todo acontecimiento relacionado al Gran Chaco para poder tener un conocimiento preciso de su estado, y llevar un seguimiento de su evolución a lo largo del tiempo, de las entidades que allí actúan y de las características presentes.

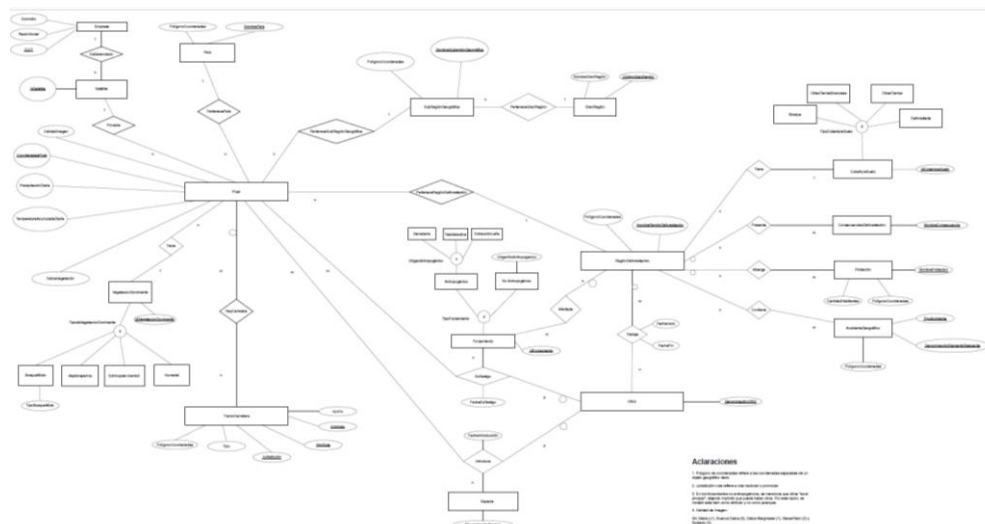
De acuerdo a la información planteada en el problema, en líneas generales se modelarán aspectos vinculados con regiones y con acontecimientos relacionados al proceso de deforestación. En cuanto a las regiones, la unidad mínima considerada es el "píxel", que surge de la información provista por satélites. Este píxel puede pertenecer a países, a ecorregiones, o a regiones definidas por criterios relacionados a las consecuencias de la deforestación. En cuanto a los acontecimientos, se considerarán la introducción de especies y los forzamientos atestiguados por ONG's que trabajan en la zona.

Así, el objetivo es modelar una base de datos para la detección de deforestación para el Bosque Chaqueño Argentino. Lo haremos primero diseñando un Modelo de Entidad Relación, que derivará en un Modelo Relacional.

Por último, listaremos las restricciones adicionales y supuestos asumidos.

Modelo de Entidad Relación

Se coloca una previsualización del mismo, para más detalle, ver archivo **draw.io** adjunto



Modelo Relacional

Empresa(CUIT, RazonSocial, Domicilio)

PK = CK = {CUIT}

Satelite(IdSatelite, CUIT)

PK = CK = {IdSatelite}

FK = {CUIT, CoordenadasPíxel}

Pais(NombrePais, PoligonoCoordenadas)

PK = {NombrePais}

CK = {NombrePais, PoligonoCoordenadas}

SubRegionGeográfica(NombreSubregionGeografica, PoligonoCoordenadas, CodigoGranRegion)

PK = {NombreSubregionGeografica}

CK = {NombreSubregionGeografica, PoligonoCoordenadas}

FK = {CodigoGranRegion}

GranRegion(CodigoGranRegion, NombreGranRegion)

PK = {CodigoGranRegion}

CK = {CodigoGranRegion, NombreGranRegion}

PIXEL(CoordenadasPíxel, CalidadImagen, PrecipitaciónDiaria, TemperaturaAcumuladaDiaria, IndiceVegetacion, IdSatélite, NombrePaís, NombreSubregiónGeográfica, NombreRegiónDeforestación, IdVegetacionDominante)

PK=CK={CoordenadasPíxel}

FK={IdSatélite, NombrePaís, NombreSubregiónGeográfica, NombreRegiónDeforestación, IdVegetacionDominante}

VegetacionDominante(IdVegetacionDominante, TipodeVegetacionDominante)

PK=CK={IdVegetacionDominante}

BosqueMixto(IdVegetacionDominante, TipoBosqueMixto)

PK=CK=FK={IdVegetacionDominante}

TramoCarretera ((NroRuta, Jurisdicción, KmInicio), PoligonoCoordenadas, Tipo, KmFin)

CK=PK={(NroRuta, Jurisdicción, KmInicio)}

HayCarretera ({CoordenadasPíxel, (Kminicio, NroRuta, NacProv)})

PK=CK={{CoordenadasPíxel, (Kminicio, NroRuta, NacProv)}}

FK={CoordenadasPíxel, (Kminicio, NroRuta, NacProv)}

RegionDeforestacion(NombreRegionDeforestacion, PoligonoCoordenadas, IdCoberturaSuelo)

PK = {NombreregionDeforestacion}

CK = {NombreregionDeforestacion, PoligonoCoordenadas}

FK = {IdCoberturaSuelo}

CoberturaSuelo(IdCoberturaSuelo, TipoCoberturaSuelo)

PK = CK = {IdCoberturaSuelo}

ConsecuenciasDeforestacion(NombreConsecuencia)

PK = CK = {NombreConsecuencia}

Presenta(NombreRegionDeforestación, NombreConsecuencia)

PK=CK={ (NombreRegionDeforestación, NombreConsecuencia) }

FK= {NombreRegionDeforestación, NombreConsecuencia}

Población(NombrePoblación, CantidadHabitantes, PolígonoCoordenadas)

PK = {NombrePoblación}

CK = {NombrePoblación, PolígonoCoordenadas}

Alberga(NombreRegionDeforestacion, NombrePoblacion)

PK=CK={ (NombreRegionDeforestación, NombrePoblacion) }

FK= {NombreRegionDeforestación, NombrePoblacion}

AccidenteGeografico((TipoAccidente, DenominacionElementoRelevante),
PoligonoCoordenadas)

PK = CK = { (TipoAccidente, DenominacionElementoRelevante) }

Contiene(NombreRegionDeforestacion, (TipoAccidente,
DenominacionElementoRelevante))

PK=CK={ (NombreRegionDeforestacion, (TipoAccidente,
DenominacionElementoRelevante)) }

FK= { NombreRegionDeforestacion, (TipoAccidente, DenominacionElementoRelevante) }

ONG (DenominaciónONG)

PK=CK= {DenominaciónONG}

Trabaja((NombreRegionDeforestacion, DenominaciónONG), FechaInicio, FechaFin)

PK=CK={ (NombreRegionDeforestacion, DenominaciónONG) }

FK= { NombreRegionDeforestacion, DenominaciónONG }

Especie(DenominaciónEspecie)

PK=CK={DenominaciónEspecie}

Introduce(DenominaciónONG, DenominaciónEspecie, CoordenadasPíxel,
FechaIntroducción)

PK = CK = { (DenominaciónONG, DenominaciónEspecie, CoordenadasPíxel) }

FK = {DenominaciónONG, DenominaciónEspecie, CoordenadasPixel }

Forzamiento(IdForzamiento, TipoForzamiento)

PK=CK={IdForzamiento}

NoAntropogénico(IdForzamiento, OrigenNoAntropogénico)

PK=CK=FK={IdForzamiento}

Antropogénico(IdForzamiento, OrigenAntropogénico)

PK=CK=FK={IdForzamiento}

Afectada(NombreRegiónDeforestación, IdForzamiento)

PK=CK={(NombreRegionDeforestacion, IdForzamiento)}

FK= { NombreRegionDeforestacion, IdForzamiento }

EsTestigo(IdForzamiento, DenominaciónONG, CoordenadasPíxel, FechaEsTestigo)

PK = CK = {(IdForzamiento, DenominaciónONG, CoordenadasPíxel)}

FK = { IdForzamiento, DenominaciónONG, CoordenadasPíxel }

Supuestos asumidos para la resolución del problema:

A partir de lo expuesto por el experto en dominio, tomamos los siguientes supuestos:

- Cada Píxel puede pertenecer únicamente a un país, no puede darse el caso que un Píxel esté dividido entre más de un país
- Cada Píxel puede pertenecer únicamente a una SubRegionGeográfica, no puede darse el caso que un Píxel esté dividido entre más de una SubRegionGeográfica
- Cada Píxel puede pertenecer únicamente a una RegionDeforestación, no puede darse el caso que un Píxel esté dividido entre más de una RegionDeforestación
- Cada RegionDeforestación sólo puede tener una CoberturaSuelo

Conclusiones:

Entendemos como **SubRegiónGeográfica** a una partición de un conjunto más grande (**GranRegión**) definida por ciertas características geográficas. En cambio, una **RegiónDeforestación** la definimos como un conjunto de píxeles que tienen características en común relacionadas con las consecuencias de la deforestación. En una analogía con mapas, un píxel podría pensarse como una Provincia (que pertenece a un país), una SubRegionGeográfica como una llanura (que puede incluir a varias provincias) y una RegiónDeforestación podría ser una región económica que también agrupa a varias provincias (que no tiene por qué ser las mismas que las agrupadas en la llanura).

Definimos **PolígonoCoodenadas** a una lista en donde cada elemento representa una coordenada de un vértice del polígono que encierra una determinada superficie. Por ejemplo, si tomamos la SubRegionGeográfica “Chaco SemiArido” y unimos todos los puntos de su PolígonoCoordenadas, la figura formada delimitara toda la subregión.

Un **TramoCarretera** va a verse definido en conjunto por su Jurisdicción (si es nacional o provincial), NroRuta y Kmlnicio. Esto es así porque entendemos que puede haber una ruta nacional que tenga el mismo número que otra ruta provincial, y al mismo tiempo que una misma ruta no tiene las mismas características en todo su trayecto (Figura 1).

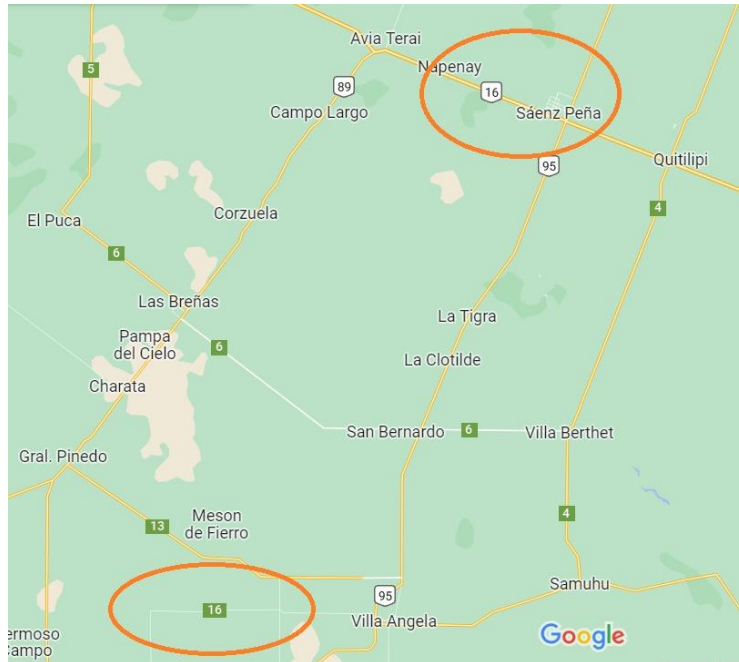


Figura 1. Tramos de carretera en la región de Chaco

Nótese la coexistencia de distintas calidades de camino para una misma carretera (RP13 y RP6) y la presencia de dos rutas con el mismo número (RP16 y RN16).

Las dos relaciones ternarias expuestas en el modelo se interpretan como:

i) ONG-Forzamiento-Píxel

- Una ONG que trabaja en un píxel → puede ser testigo de uno o más forzamientos
- Una ONG que es testigo de un forzamiento → trabaja en uno o más píxeles
- Un forzamiento en un píxel → es atestiguado por una o más ONG (pueden ser distinta fecha)

ii) ONG – Especie - Píxel

- Una especie introducida por una ONG → la introduce en uno o más lugares
- Una ONG que trabaja en un píxel → puede introducir una o más especies
- Una ONG que introduce una especie → la introduce en uno o más píxeles