**tfm contenido / proceso resumen**

* **- INTRODUCCIÓN ASIGNATURAS APLICADAS**
* Análisis de datos
  + Conocer la información
  + Entenderla
  + Seleccionarla
  + Procesarla
  + Resumen estadísticas básicas
* Base de datos
  + Preprocesado y depuración de datos
  + Uso de Docker para Neo4J
  + Procesado con BD NoSQL basada en grafos Neo4J
  + Modelado con Neo4J
  + Uso de grafos caso específico con Neo4J
* Visualización Interactiva y Herramientas de Visualización
  + Crear Historia en Tableau para Países y Cuotas de producción (CRM 2020)
  + Visualizaciones generales
  + Visualización especifica de Neo4J
  + Exportación SVG
    - Neo4j
    - Tableau
  + Visualización del SVG general con navegador
* Uso general en análisis de datos y tratamientos masivos
  + Uso de Python
  + Uso de librerías específicas Python
    - Librería Python para uso de Neo4J
    - Librería distancias entre países
* Gobierno del Dato (Protección de datos)
  + Disclaimer RGPD / GDPR
    - No uso de datos personales.
    - Fuente de datos oficial y abierta.
* **Búsqueda e identificación de fuentes de datos**
  + Oficiales o semioficiales
  + Privadas
    - Datos abiertos
    - Datos con suscripción
* **Valoración de las fuentes de datos**
  + Cualitativa
    - Alcance
      * Países
      * Materiales
        + Obtención
        + Aplicaciones
        + Consideraciones particulares
        + Cadenas de suministro
    - Periodos
    - Interrelación de los datos
    - Solvencia de las fuentes
    - Facilidad de acceso a los datos
  + Cuantitativa
    - Número de periodos de tiempo disponibles
    - Número de países
    - Número de materiales
    - Número de pasos cadena de suministro
    - Otros datos

**SELECCIÓN DE FUENTE DE DATOS**

FUENTES > CARACTERÍSTICAS EVALUADAS > VALORACIÓN > SELECCIÓN FINAL

**FUENTE SELECCIONADA: CRM DATOS OFICIALES EUROPEOS**

* Materiales críticos
* Materiales no críticos
* Informe Final

CRM Facsheet Critical

* contenido

CRM Non Critical

* contenido

CRM Final Report

* contenido

**DATOS SELECCIONADOS EN EL ESTUDIO:**

**SUMINISTRO GLOBAL**

PAISES ALCANCE MUNDIAL

**WGI - (de cada PAÍS)** – índice global mundial - mix de estabilidad política etc. (da idea de fiabilidad en la continuidad del suministro según cambios políticos, estabilidad legislativa … Cuanto más alto menos fiable sería el suministro por parte de ese país)

**t - (de cada PAÍS) – índice de ponderación del análisis CRM para futuros estudios- -qué es**

**MATERIAL**

**Críticos (valoración 2020)**

**no críticos**

**CADENA DE SUMINISTRO**

* + - **Presentación**
    - **Resumen etapa 1 y etapa 2**

ETAPA 1: Extracción (extracción, minería, etc) – obtención de material “bruto”

PAISES PRODUCTORES ETAPA 1

Cuota

WGI

t

ETAPA 2: Procesado (o refinamiento) - obtención de material procesado o “refinado”

PAISES PRODUCTORES ETAPA 2

Cuota

WGI

t

**USO DEL MATERIAL**

**MATERIAL**

**Clasificación – ore, LEE, etc**

APLICACIONES

Share o cuota

Aplicación oficial

Grupo de aplicación resumido

Grupo NACE2 de uso o aplicación de un material

Total de valor del grupo NACE2

Aportación proporcional del material al total del grupo NACE2

PREPROCESADO DE DATOS

* Carga de datos
  + Revisión de opción automática
  + Tratamiento y revisión semi-manual
  + Carga y formateo en Excel
    - Uso de macros y fórmulas
    - Campos numéricos de texto a numérico
    - Campos de texto, eliminando espacios finales
    - Eliminar datos sobrantes
      * Cuotas de países por material que superan 100%
      * Datos “<1%” adicionales a 100%
      * Ajuste de datos contrastando con informes origen
        + CRM
        + NO CRM
        + Complementar con fuentes externas cuando es necesario
* Formato de datos
* Asegurar proporciones de suministro corresponde al 100%

TRATAMIENTO DE LOS DATOS – PROCESADO CON GRAFOS

* POR QUÉ GRAFOS
* ANTECEDENTES
* CAPACIDADES DE LOS GRAFOS
* MODELO USANDO GRAFOS
  + Justificación
  + Explicación del modelo
  + Nodos
    - propiedades
  + Relaciones
    - propiedades
  + Interrelaciones y algoritmos
    - Grado de un nodo
    - Camino
    - Dependencias

CASO DE USO PARA MATERIALES BÁSICOS EN BATERÍAS ELÉCTRICAS

Intro y justificación

funcionamiento básico de una batería

Materiales frecuentes en las baterías eléctricas

PROCESADO CON GRAFOS NEO4J

* INTRO A NEO4J
* INSTALACIÓN
* VENTAJAS E INCONVENIENTES
  + No SQL
  + VISUALIZACIÓN
    - Exportar SVG
    - Uso potencial D3 SVG
* LIBRERÍA DE USO DE NEO4J CON PYTHON
  + Justificación
    - Librería
    - python
  + Instalación
    - Anaconda
    - Jupyter notebook
    - librerías
  + utilización
    - el sw
    - repositorio
* CARGA DE LOS DATOS EN NEO4J
  + Ficheros CSV
  + Datos y propiedades
* LA BASE DE DATOS NEO4J CREADA
  + Características técnicas y nombre – tamaño , localización …
  + docker
  + acceso
* EL MODELO EN NEO4J
  + Nodos
    - Propiedades
  + Relaciones
    - propiedades
* PROCESANDO EL MODELO: DATOS GENERALES
  + Consultas básicas
  + Consultas ¿avanzadas?
* PROCESANDO EL MODELO: BATERÍAS ELÉCTRICAS
  + Nodos específicos
    - Propiedades
  + Relaciones específicas
    - propiedades
  + Consultas básicas específicas
  + Consultas ¿avanzadas? Específicas
* PROCESANDO EL MODELO: EXPORTAR MODELO VISUAL: SVG
  + Opciones de exportación
  + Por qué SVG
  + Ejemplo sencillo
  + Usos potenciales SVG y D3

RESULTADOS

* + De dominio – cadena de suministro global
  + De análisis de datos – fuentes de datos, procesado y limpieza
  + Tecnológicos y procesado
    - Preprocesado
      * Librería usar tablas desde PDF en otros casos
      * Office : EXCEL, WORD y MACROS con acciones manuales
    - Procesado
      * DOCKER
      * NEO4J
      * PYTHON
        + LIBRERÍAS NEO4J
      * VISUALIZACIÓN
        + NEO4J
        + SVG
        + D3

APORTACIONES FUTURAS

* Aplicar al modelo europeo sobre datos recopilados
* Identificar ubicación de minas y factorías en países y potenciales medios de transporte , circuitos y tiempos para medir la huella de carbono de cada suministro -- Librería de distancias entre países para valorar suministros
  + Tipos de transporte combinados y asignables según material, origen, destino, distancias, frecuencia de envío, seguridad de la ruta, alternativas de ruta y otros parámetros analizables:
    - tierra (camión, tren/ pequeño-mediano-grande)
    - mar(barco / pequeño-mediano-grande)
    - aire(avión)
      * (librería de distancias – referencias-)
      * Valorar huella de carbono
        + Documentación cálculo de huella por tipo de transporte – referencias-
* Uso potencial D3 Grafos sobre los datos orige
* Aplicar al modelo europeo sobre datos recopilados
* Profundizar o automatizar:
  + Localizar personas de contacto y fuentes de datos originales para automatizar y evolucionar el modelo o concepto con más perímetro de alcance de datos incluso para trabajos de doctorado, automatizando desde el origen.
* Uso de otras herramientas de base de datos sobre grafos como evaluación comparativa sobre el mismo alcance o similares
* Aplicación de D3 SVG para visualización completa con alternativas de selección sobre la visualización y la base de datos