15-1-2016

Caso Práctico

Integración y monitorización de fuentes de información



Javier García Pérez - 100290698 José Manuel Fernández Ruiz - 100290892

Máster en Ingeniería Informática

Curso 1

integración de sistemas informáticos

UC3M

Índice

[Introducción 1](#_Toc438497208)

[Resumen ejecutivo 2](#_Toc438497209)

[Objetivo 2](#_Toc438497210)

[Alcance 2](#_Toc438497211)

[Solución diseñada 3](#_Toc438497212)

[Aspectos claves 3](#_Toc438497213)

[Conclusiones 3](#_Toc438497214)

[SEGUIR 5](#_Toc438497215)

[ANEXO I: Cantidad de estaciones por provincia 6](#_Toc438497216)

[ANEXO II: Elementos químicos y variables medidas 7](#_Toc438497217)

[ANEXO III: Temas, hashtags y cuentas Twitter 8](#_Toc438497218)

Índice de Tablas

[Tabla 1 Núm. Estaciones por provincia 6](#_Toc438497219)

[Tabla 2 Elementos químicos y variables medidas 7](#_Toc438497220)

[Tabla 3 Temas, hashtags y cuentas de tweets 8](#_Toc438497221)

Índice de ilustraciones

[Ilustración 1 PM10 y PM25 7](#_Toc438497222)

# Introducción

El siguiente documento recoge la memoria de la realización del caso práctico de la asignatura Integración de Sistemas Informáticos del Máster en Ingeniería Informática de la Universidad Carlos III de Madrid. Este caso práctico tratará sobre el análisis y estudio de la contaminación atmosférica en toda España entre los años 2014 y 2015 en base a los datos recogidos por las estaciones de calidad del aire del país y a la opinión social de la gente a través de Twitter.

En este documento se mostrará, entre otras cosas, un resumen ejecutivo con el objetivo, alcance, principales aspectos claves del trabajo, etc., un análisis y diseño del problema de integración propuesto, unas visualizaciones y guía sobre la prueba de concepto, y finalmente. unas conclusiones sobre los objetivos alcanzados y problemas que se han tenido a lo largo de la elaboración del caso práctico.

Este trabajo ha sido realizado por el grupo 2 compuesto por los alumnos José Manuel Fernández Ruiz y Javier García Pérez.

A lo largo de la memoria sean añadido una serie de anexos con información complementaria sobre la realización del caso práctico.

# Resumen ejecutivo

Este apartado recoge una panorámica del trabajo realizado. Los puntos a destacar de este apartado son:

* **Objetivo:** recoge las metas a conseguir.
* **Alcance:** la extensión que tendrá la solución propuesta.
* **Solución diseñada:** resumen sobre la solución diseñada.
* **Aspectos claves:** principales características que proporciona la solución diseñada.
* **Conclusiones:** breve resumen sobre lo principal de los apartados anteriores.

## Objetivo

El objetivo del caso práctico es abordar el tema de la contaminación atmosférica en España desde diferentes puntos de vista para su posterior análisis. Para esta meta se han usado datos validados de las estaciones de calidad del aire de las diferentes provincias del país y opiniones sobre la contaminación realizadas por las personas en Twitter.

De esta manera, se pretende hacer una comparación de los niveles de contaminación reales extraídos de las estaciones con lo que realmente percibe y expresa la gente mediante sus mensajes en Internet. Se mostrarán niveles de todas las provincias de España según los datos oficiales de las estaciones de aire y de la misma manera, para estos lugares, comprobaremos si las personas también perciben y expresan este hecho mediante su opinión online. Ej. En Madrid el nivel medido de NO2 es muy alto y se observan quejas en twitter de como el gran volumen de tráfico afecta al aire de la ciudad.

Toda esta información se mostrará en unos dashboards intuitivos y completos, que permitirán analizar una provincia de forma individual o comparándola con otra.

## Alcance

Para lograr los objetivos anteriores se partirá de los datos recogidos por las estaciones de aire de España durante el año 2015, aunque se usarán datos de 2014 para todas esas estaciones que todavía no hayan validado sus datos de 2015. A la hora de mostrar la información en la solución se usarán medidas diarias.

Se empezará con un total de 331 estaciones de calidad del aire con sus respectivas medidas para las fechas mencionadas. Estas estaciones abarcan todo el territorio español salvo las provincias de las CCAA de Andalucía y Extremadura, y las ciudades autónomas de Ceuta y Melilla, debido a la falta de documentación compatible con la que se pudiera trabajar y extraer datos. Se pueden consultar el número de estaciones de aire recopiladas por provincia en la tabla del *ANEXO I: Cantidad de estaciones por provincia*. Por otro lado, los contaminantes medidos por las estaciones y que se analizarán en la solución se pueden ver en el *ANEXO II: Elementos químicos y variables medidas*. Matizar que cada estación de calidad de aire no mide todos los contaminantes listados, si no que la cantidad que mida dependerá de su ubicación y función.

Por otro lado, en cuanto la opinión social, se recopilarán tweets entre 2014-2015 en base a mensajes publicados con los temas, hashtags y cuentas listados en el *ANEXO III: Temas, hashtags y cuentas Twitter*. Para la prueba conceptual únicamente se usarán tweets entre el 1 de noviembre y 31 de diciembre debido a que la API de Twitter únicamente permite recuperar mensajes con antigüedad de una semana máximo.

## Solución diseñada

Hablar por encima de la aplicación web, de mongodb, meaningcloud,etc

## Aspectos claves

Vender la aplicación, decir lo que tiene q la hace importante, comparador, datos preprocesados y limpios para descargar etc

## Conclusiones

Breve resumen de los puntos anteriores

MAS EN DISEÑO ESTA PARTE QUE AQUÍ: En cuanto a la información recopilada de las personas, la mayoría, está recogida en una colección en MongoDB de tweets, estos mensajes abarcan en cuanto a fecha desde principios de noviembre hasta finales del año 2015. Para recopilar estos tweets se han usado palabras clave tales como las recogidas en PONER ANEXO. Aparte de esta fuente de datos también se ha usado el portal de quejas ciudadanas del ayuntamiento de Madrid y … PONER MÁS SI USAMOS MÁS

El trabajo final que se mostrará de visualización de los datos no abarcará la totalidad de estos aspectos mencionados, sino que se verá simplificado para mostrar únicamente como prueba conceptual como sería el resultado. Esto no implica que en el diseño de esta memoria no se explique cómo sería el proceso para su desarrollo completo posterior.

# SEGUIR

DECIR HERRAMIENTAS USADAS. FORMATO DE COLECCIONES EN MONGODB, DISEÑO ETC, ver enunciado para ver que más poner

# ANEXO I: Cantidad de estaciones por provincia

La siguiente tabla muestra el número de estaciones de calidad de aire por provincia cuyos datos se han recopilado para la solución del caso práctico.

|  |  |  |  |
| --- | --- | --- | --- |
| Provincia | Cantidad | Provincia | Cantidad |
| A Coruña | 6 | **La Rioja** | 5 |
| Álava | 8 | **Las Palmas** | 20 |
| Albacete | 1 | **León** | 5 |
| Alicante | 12 | **Lérida** | 1 |
| Almería | 0 | **Lugo** | 4 |
| Asturias | 5 | **Madrid** | 26 |
| Ávila | 1 | **Málaga** | 0 |
| Badajoz | 0 | **Melilla** | 0 |
| Barcelona | 24 | **Murcia** | 8 |
| Burgos | 6 | **Navarra** | 9 |
| Cáceres | 0 | **Ourense** | 2 |
| Cádiz | 0 | **Palencia** | 2 |
| Cantabria | 11 | **Pontevedra** | 4 |
| Castellón | 23 | **Salamanca** | 3 |
| Ceuta | 0 | **Santa Cruz de Tenerife** | 30 |
| Ciudad Real | 5 | **Segovia** | 1 |
| Córdoba | 0 | **Sevilla** | 0 |
| Cuenca | 1 | **Soria** | 2 |
| Gerona | 1 | **Tarragona** | 6 |
| Granada | 0 | **Teruel** | 1 |
| Guadalajara | 2 | **Toledo** | 3 |
| Guipúzcoa | 14 | **Valencia** | 25 |
| Huelva | 0 | **Valladolid** | 6 |
| Huesca | 3 | **Vizcaya** | 25 |
| Islas Baleares | 17 | **Zamora** | 1 |
| Jaén | 0 | **Zaragoza** | 2 |
| Total | 331 | | |

Tabla Núm. Estaciones por provincia

# ANEXO II: Elementos químicos y variables medidas

La siguiente tabla muestra el número de elementos químicos y variables medidas por las estaciones de calidad del aire guardadas.

|  |  |  |
| --- | --- | --- |
| Abreviatura | Nombre | Unidades |
| BEN | Benceno | µg/m3 |
| CO | Monóxido de carbono | µg/m3 |
| NO | Monóxido de nitrógeno | µg/m3 |
| NO2 | Dióxido de nitrógeno | µg/m3 |
| NOx | Óxidos de nitrógemo | µg/m3 |
| O3 | Ozono | µg/m3 |
| OXL | Ortoxileno | µg/m3 |
| PM10 | Partículas en suspensión < 10 um **(\*)** | µg/m3 |
| PM25 | Particulas en suspensión < 2,5 um | µg/m3 |
| SH2 | Sulfuro de Hidrógeno | µg/m3 |
| SO2 | Dióxido de azufre | µg/m3 |
| TOL | Tolueno | µg/m3 |
| XIL | Xileno | µg/m3 |

Tabla Elementos químicos y variables medidas

**(\*)-** *PM10* y *PM25* recogen información sobre partículas compuestas por polvo, cenizas, polen, cemento, hollín y metales pesados que hay en la atmósfera. El diámetro de cada partícula es el siguiente:

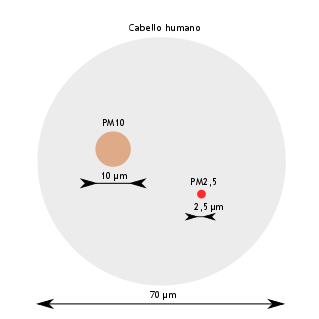


Ilustración PM10 y PM25

# ANEXO III: Temas, hashtags y cuentas Twitter

La siguiente tabla muestra en base a qué temas, hashtags y cuentas se han recopilado los mensajes usados en la solución final.

|  |  |  |
| --- | --- | --- |
| Temas | Hashtags | Cuentas |
| alergia contaminación | #airemadrid | ASURCAI |
| basura Barcelona | #airebarcelona | ecogestos |
| basura calle | #basura | EPmedioambiente |
| basura calles | #basuramadrid | FilterQueenES |
| basura ciudad | #calidadaire |  |
| basura Madrid | #contaminaciobarcelona |
| basura valencia | #contaminacion |
| basureros | #contaminacionatmosferica |
| calles limpias | #contaminacionmadrid |
| CO2 | #desperdicios |
| contaminacion aire | #madridsincontaminacion |
| contaminacion atmosférica | #medioambientemadrid |
| huelga recogida basura | #malolorbcn |
| limpieza Madrid | #pestebcn |
| NO2 | #residuos |
| PM10 |  |
| papeleras Barcelona |
| puntos limpios |
| papeleras Madrid |
| reciclaje |
| recogida de basura |
| SO2 |
| suciedad |
| vertederos |

Tabla Temas, hashtags y cuentas de tweets