Development of a web application for data analysis of behavioral patterns of selected environmental and epidemiological variables in two localities of Bogotá.

Jairo Augusto Cortes Mendez[1], Mario Fernando Castro Fernández[2], Fredys Alberto Simanca Herrera[3], Jaime Alberto Páez Páez[4], Oscar Barragán[5], Diana Lancheros[6]

[1] Universidad Cooperativa de Colombia, Jairo.cortes@campusucc.edu.co

[2] Universidad Cooperativa de Colombia, Mario.castrof@campusucc.edu.co

[3] Universidad Cooperativa de Colombia, fredys.simanca@campusucc.edu.co

[4] Universidad Cooperativa de Colombia, jaime.paez@campusucc.edu.co

[5] Universidad Cooperativa de Colombia, oscar.barrgan@campusucc.edu.co

[6] Universidad Cooperativa de Colombia, diana.lancheros@campusucc.edu.co

Abstract

The project for the development of a web application that allows measuring air conditions and the risk of atmospheric pollution, based on the concentration ranges of six polluting factors (PM10, PM2.5, O3, CO, SO2, NO2), initially under the geographical delimitation of two localities of the city of Bogota (Puente Aranda and Ferias). The above is possible, taking as a reference the information obtained from environmental sources (SISAIRE, IDEAM and Ministry of Environment). The project is derived from the research called Determining the correlation of environmental variables that affect health conditions in two localities of Bogota, applying data analytics, presented in the internal call of the Universidad Cooperativa de Colombia.

The application provides significant data for integration with statistical analysis tools, in order to allow a correlation between variable data, achieving a better reaction to the risk in each of the levels of contamination.

DOI Number: 10.14704/nq.2022.20.13.NQ88169 **Neuro Quantology 2022; 20(13):1353-1359**

Introduction

According to Lara I. and Chacón N. (2021), Bogotá´s Air Quality and Health Risk Index -IBOCA- has been created to inform and structure in an articulated way how the air quality is and the level of risk generated for human health by variables and patterns associated with air pollution, and as a result the generation of early notifications and emergencies due to air pollution in the city is proposed, allowing the articulation of citizen and institutional measures aimed at protecting the environment of the city and the health and quality of life of its inhabitants.

With the above, there is a need to develop a web tool that allows automated calculations, both for concentrations of particulate matter (PM 10 and PM2.5) and pollutant gases (SO2, NO2, CO and O3), in order to perform analysis of the calculated information in statistical tools and data mining, for preventive decision-making.

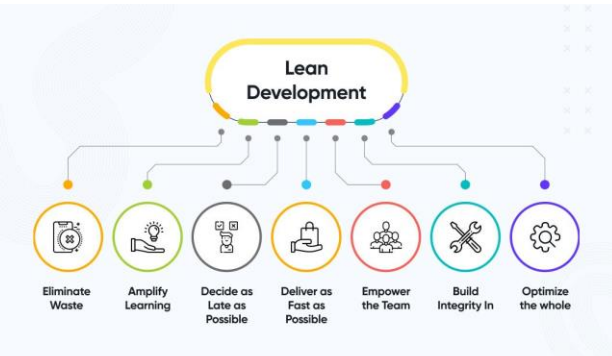
The objectives of the research are to develop a web application that allows the analysis of data on the behavior patterns of the selected variables, in order to perform an automated calculation of air quality indexes from particulate matter and pollutant gases at environmental and epidemiological level in two locations of Bogota. Some specific objectives were also oriented to diagnose the functional and non-functional requirements, design the Web application according to the established requirements, create the Web application for data analysis of the selected variables, implement the Web application in the user and administration dimensions, and verify the operation of the Web application. The question that guides this research is: How through the development of a web application can be generated analysis on behavioral patterns for the calculation of the air quality index in two localities of Bogota?

Methodology

In this development process, we worked under the Lean Software Development paradigm (Garcia, 2017), which is an iterative methodology and considered as an agile method that originates from the processes of Toyota factories in Japan where lean manufacturing was used as a method to avoid and take advantage of the waste that was generated in order to optimize the entire production chain. Likewise, in the software development process these optimized processes are required through a method called Lean Software Development based on three fundamental purposes: To optimize and decrease the times for delivery of products, to decrease their value and finally to contribute to the removal of errors that can be generated in the process, thus improving each of its processes with quality. (Lasa et al, 2017).

This work philosophy is made up of seven principles and 22 tools as shown in Figure 1, which, when effectively implemented in the development of the web application, will develop an efficient and effective system. The seven principles described by LSD technology are: elimination of flaws and waste, improve learning processes, make decisions as late as possible, deliver as fast as possible, empower the team, create integrity and visualize the whole.

Figure 1 – The 7 principles of Lean software development



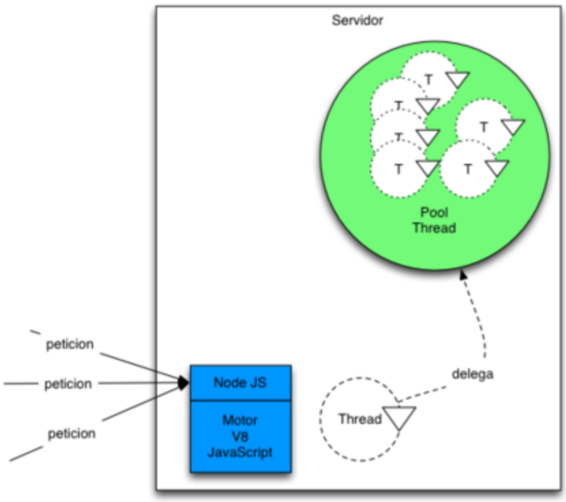
Source: taken and adapted from https://www.techmonsterr.com

Results

With the application of the Lean Development methodology (Garcia, 2017) the expected results were obtained in the development of the web application for the analysis of variables and behavioral patterns for the calculation of the air quality index in two locations in the city of Bogotá. Initially the following activities were performed in the initial stage, as was the documentation on the air quality indexes in the city of Bogotá and the regulations that establish the parameters that define the levels of prevention and alert or emergencies for air pollution. After these activities, the needs were identified to determine the requirements and correlate them with the requirements of the web application and the regulations on air pollution, as well as the protocols for monitoring and follow-up of air quality, including the guidelines to take into account when evaluating the most harmful polluting factors.

As a result of the mentioned process, technologies were selected to generate the software architecture developed in Java Script Node. To develop the application, we implemented the Node.js technology, based on Google's JavaScript V8engine, which was designed to be executed in a code environment in a browser, running in its architecture very easily and quickly. As a result, Node.js optimization processes were also generated, which are done through a set of threads, also the set of threads is formed by the libuv library. This contains its own structure and environment called asynchronous multithread, this Node.js technology transfers the activities to the set of threads as shown in Figure 2.

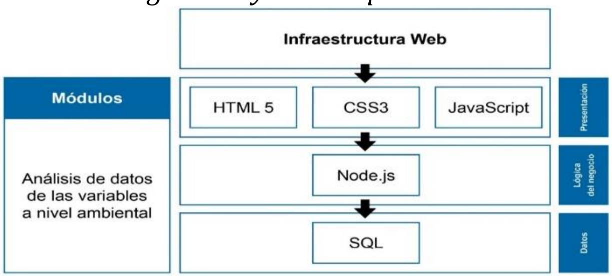
Figure 2 – Node.JS Asynchronous Tasks Scheme



Source: taken and adapted from https://www.genbeta.com/desarrollo/como-funciona-node-js

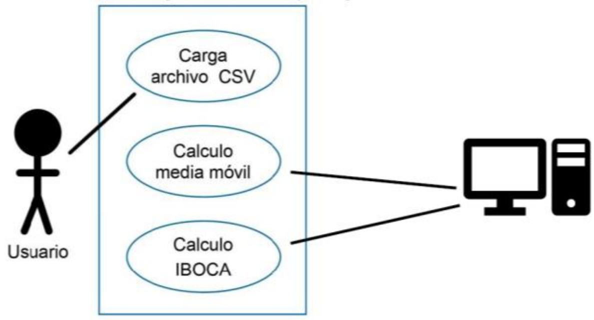
The Web application oriented to the development of IBOCA, in its software design is done by structuring a process consisting of data, its architecture, interfaces and algorithms, thus translating the software requirements into a designed solution, as can be analyzed in its system architecture in Figure 3.

Figure 3. System Arquitecture



The process of describing the architecture begins with the loading of CSV files, which contain the records by hours in periods of 30 days, also containing the concentration ranges of the pollutants so that later the environmental variables are loaded and calculated to obtain the air quality index for Bogota - IBOCA. The moving average is determined in an established period of time for each of the pollutants taken from the CSB file with records of the concentration ranges of the pollutant factors. In general terms, the actions can be visualized by means of a use case diagram, as shown in Figure 4.

Figure 4. Use cases software process



The technological infrastructure used for the IBOCA calculation was tested on a multiplatform operating system, with browsers such as Chrome, Firefox, Microsoft Edge, Safari, and the technologies implemented were HTML5, CSS3, Node.js SQL with the general use of JavaScript.

In its elaboration, a work path was traced in which is defined; parameterization of variables and constants, methods to implement the development, technologies that best adapt to the need of the development, and execution times in programming. This work path is described below (Ballesteros et al, 2020):

Parameterization of variables: In tables 1 and 2, it can be seen how each of the variables to be used has been defined:

Table 1. Particulate Matter Codes

Particulate Matter	PM 10
	PM 2.5

Table 2. Polluting gases code

Polluting gases	CO - Carbon Monoxide
	O3 – Ozone
	SO2 - Sulfur Dioxide
	NO2 - Nitrogen Dioxide

Parameterization of Constants: Table 3 shows how the constants to be used in the model were coded and defined (Hernández et al, 2012).

Table 3. Parameterization of Constants

				µg/m3		µg/m3		ppm		ppb		ppb		ppb			
ICA		Color	Categoria	PM-10 (24h)		PM-2,5 (24h)		CO (8h)		SO2 (1h)		NO2 (1h)		O3 (8h)		O3 (1h)1	
I-*bajo*	I-*Alto*			PC-*bajo*	PC-*Alto*	PC-*bajo*	PC-*Alto*	PC-*bajo*	PC-*Alto*	PC-*bajo*	PC-*Alto*	PC-*bajo*	PC-*Alto*	PC-*bajo*	PC-*Alto*	PC-*bajo*	PC-*Alto*
0	10		Favorable	0	54	0	12	0	4.4	0	35	0	53	0	59	------	------
10.1	20		Moderada	55	154	12.1	35.4	4.5	9.4	36	75	54	100	60	75	------	------
20.1	30		Regular	155	254	35.5	55.4	9.5	12.4	76	185	101	360	76	95	245	323
30.1	40		Mala	255	354	55.5	150.4	12.5	15.4	186	304	361	649	96	115	324	401
40.1	60		Muy Mala	355	424	150.5	250.4	15.5	30.4	305	604	650	1249	116	374	402	794
60.1	1000000		Peligrosa	425	604	250.5	500.4	30.5	50.4	605	1004	1250	2049	374	938	795	1185

- Design and programming of the web application logic and database -Backend-.

The logic of the application functions is designed taking as reference the following initial requirements:

R1: Load, storage and reading of CSV file.

R2: Calculation of Moving Average of each pollutant factor.

R3: Calculation of IBOCA index for each pollutant.

R4: Data printing method.

The application does not have an authentication system.

- Design and programming of the client view -Frontend-.

Interface of the application is developed in HTML5 and CSS language, which is integrated in Node.JS with EJS view format.

- Development of unit and integration tests -Testing-.

Connectivity between database and application is validated.

Verification of flat file loading in the defined structure.

Hierarchical deployment of the implemented logic is monitored.

- Definition of the conclusions of the development cycle.

The tests performed to the software allowed to guarantee the fulfillment of the requirements in their entirety, raised in the development of the research for the analysis of data and behavior patterns of the environmental and epidemiological variables in the two selected localities in Bogota. Several test cases were generated, such as the connection to the server at port 5400, connection to the SQL Express database - IBOCA database, validation of the loading of CSV files to the database, the validation process of the operation of the moving average and IBOCA (Bogota, 2010).

In its execution in general the architecture of the application and the modules developed at Backend and database level, can be analyzed below. For the development, Node.Js and SQL Express technologies were implemented and the connection to the application server was configured through Localhost, port 5400 as shown in the following figure.

Figure 5. Port configuration code

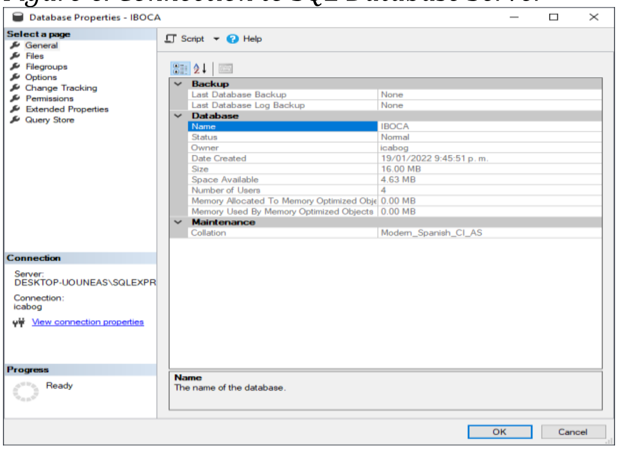
```
src > config > JS server.js > ...
const express = require('express');
const path = require('path');
const bodyParser = require('body-parser');
const app = express();

// configuracion
app.set('puerto',process.env.PORT || 5400);
app.set('view engine','ejs');
app.use('/assets', express.static('../assets'));
app.set('views', path.join(__dirname, '../app/views'));


// middleware
app.use(bodyParser.urlencoded({extended: false}));
module.exports =app;
```

The SQL Express Database server connection configuration in Figure 6.

Figure 6. Connection to SQL Database Server



Neuro Quantology | October 2022 | Volume 20 | Issue 13 | Page 1353-1359 | doi: 10.14704/nq.2022.20.13.NQ88169
Jairo Augusto Cortes Mendez, Mario Fernando Castro Fernández, Fredys Alberto Simanca Herrera, Jaime Alberto Páez Páez[4], Oscar Barragán, Diana Lancheros/ Development of a web application for data analysis of behavioral patterns of selected environmental and epidemiological variables in two localities of Bogotá.

The connection code is shown below:

Figure 7. SQL database connection code.

```
src > config > JS dbConnection.js > ...
const sql = require('mssql');

var config = {
user: 'icabog',
password: '1234',
server: 'DESKTOP-UOUNEAS\\SQLEXPRESS',
database: 'IBOCA',
options: {
  trustServerCertificate: true,
  }
};
module.exports = config;

sql.connect(config, err => {
  if(err){
      throw err ;
  }
  console.log("Conexión Exitosa a BD !");

});
```

For the definition of the DDL database, the following tables were analyzed: Measurement tables (Receives base records sent by the application), ICA points (Contains the constant values corresponding to the concentration ranges for each pollutant factor, according to the regulatory framework). Localities table (Parameterization of localities), 1Hr. (Calculations for exposure times of 1 hour), Table 8Hrs. Calculations (Calculations for exposure times of 8 hours), Table 24Hrs. Calculations (Calculations for exposure times of 24 hours), SQL scalar functions were used for the calculation of IBOCA by pollutant factor. Some triggers used in the measurement database in Figure 8.

Figure 8. TRIGGER that are inserted into the measurement database

```
SQLQuery3.sql - DE...BOCA (icabog (58))
USE [IBOCA]
GO
/****** Object:  Trigger [dbo].[inserta_calculos_var]    Script Date: 26/01/2022 7:06:52 p. m. ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
ALTER TRIGGER [dbo].[inserta_calculos_var]
   ON [dbo].[mediciones]
   AFTER INSERT
AS
BEGIN
    -- SET NOCOUNT ON added to prevent extra result sets from
    -- interfering with SELECT statements.
    SET NOCOUNT ON;

    declare @fecha datetime, @id_localidad int

    select @fecha = try_convert(datetime, fecha, 103), @id_localidad = id_localidad from inserted

    insert into calculos_8h(
        id_localidad, fecha, ppm_co, promedio_co, ppb_oz, promedio_oz
    )
    select top 1 A.id_localidad, A.fecha,
        A.ppm_co, avg(sum(A.ppm_co)) over (order by A.fecha asc)  as promedio_co,
        A.ppb_oz, avg(sum(A.ppb_oz)) over (order by A.fecha asc)  as promedio_oz
    from (
            select a.id_localidad, a.fecha, a.ppm_co, a.ppb_oz
            from calculos_8h as a
            where a.id_localidad = @id_localidad and a.fecha > DATEADD(HH,-8,@fecha) and a.fecha <= @fecha
            union all
            select id_localidad, try_convert(datetime, fecha, 103), co, ozono
            from inserted
        ) as A
    group by A.ppm_co, A.ppb_oz, A.fecha, id_localidad
    order by A.fecha desc
```

Discussion

As part of the analysis of the results and the existing information, other forms of analysis of the variables and patterns associated with the air quality index can be generated, allowing these results to make better decisions. Some elements that describe the resulting application:

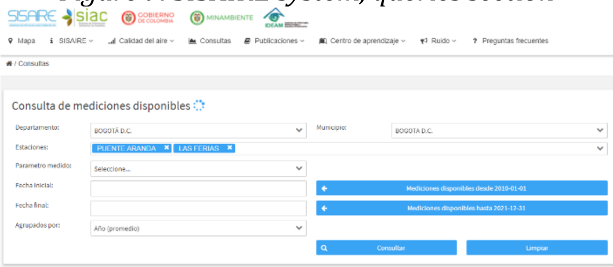
The user must access the Air Quality information system -SISAIRE- (http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml), to consult and download the published information, selecting the required fields and in a format that matches the following structure, table 4:

Table 4. Identification of required fields for downloading information.

id_localidad	fecha	co	ozono	pm10	pm2	so2	no2

Figure 9 shows the SISAIRE system interface where queries are made.

Figure 9. SISAIRE system, queries section



After downloading the information, the file must be converted to a CSV format delimited by commas, in order to perform the massive upload to the web application.

It is important to take into account that for now the application has created two locations with the following IDs as shown in table 5:

Table 5. Selected localities of Bogota

ID_LOCALIDAD	NAME
01	**Puente Aranda**
02	**Ferias**

The web application has an interface, where the first interaction that is performed is on the "Select file" button, which allows us to choose the CSV file to be uploaded, as shown in Figure 10.

Figure 10. Bogota Air Quality Index System. File selection interface

To load the file, the CSV document must have the name, carga_plano_pa.csv". After selecting the document, the "Send file" button must be clicked, which sends it to the application's server storage, for subsequent reading and insertion of records in the database, as shown in Figure 11.

Figure 11. Bogota Air Quality Index System. Send file interface

Finally, click on the "View Data" button, which displays on the screen the records and calculations made by the web application, as shown in Figure 12.

Figure 12. Bogota Air Quality Index System. Interface to visualize the data

Conclusions

As a result of the project, the calculations of air quality indexes are integrated into a Web tool, in accordance with the environmental regulatory framework.

This integrated application with statistical analysis tools is available to the user.

The knowledge of environmental indicators and processes is strengthened.

This development is created as an initial product, which is open to improvements and complements according to later requirements in other projects.

The contribution of the research to the understanding of the workflow in the field of software engineering applied to environmental processes.

The necessary documentation was gathered on current public health issues, on pollution policies, protocols and data management on air quality monitoring and tracking, thus achieving a comprehensive knowledge base for the analysis of variables and patterns for decision-making.

References

Ballesteros, E. A., Montealegre, E. L., Zambrano, D. L., Ortíz, E. Y., & Arango, C. D. (2020). Informe Anual de Calidad del aire de Bogota 2019 Red de Monitoreo de Calidad del Aire de Bogotá RMCAB.

Bogotá, A. m. (2010). PLAN DECENAL DE DESCONTAMINACIÓN DEL AIRE PARA BOGOTÁ 2010- 2020.

Garcia, A. A. (2017). *Metodos agiles: Scrum, Kanban, Lean.* Madrid: Larouse.

Hernández L; Aristizábal G; Salgado Y; Cantor L; Medina K. y Reyes J. (2012). Asociación entre la contaminación del aire y la morbilidad por enfermedad respiratoria aguda en menores de cinco años en tres localidades de Bogotá. Pediatría, Volume 45, Issue 2, June 2012, Pages 124-138. DOI: https://doi.org/10.1016/S0120-4912(15)30011-2

Lara I, Chacón N. (2021). Ficha técnica del indicador Índice Bogotano de Calidad del Aire (IBOCA). Obtenido de: https://saludata.saludcapital.gov.co/osb/fichas_tecnicas_osb/salud-ambiental/OSB_SaludAmb-IBOCA.pdf

Lasa C., Alvarez, A., Otros (2017). Métodos Agiles. Scrum, Kanban, Lean. Anaya Multimedia.

IDEAM. (2022). *Contaminación y calidad ambiental* . Obtenido de Instituto de Hidrología, Meteorología y Estudios Ambientales: http://www.ideam.gov.co/web/contaminacion-y-calidad-ambiental/

MINAMBIENTE. (2022). *Normativa*. Obtenido de Ministerio de Ambiente y Desarrollo Sostenible: https://www.minambiente.gov.co/normativa/

MINISTERIO DE AMBIENTE, VIVIENDA Y DESARROLLO TERRITORIAL. (Febrero de 2008). Protocolo para el monitoreo y seguimiento de la calidad del aire. *Manual de operación de sistemas de vigilancia de la calidad del aire*. Bogotá, Colombia.

SECRETARÍA DISTRITAL DE AMBIENTE. (11 de Diciembre de 2015). Resolución Conjunta Número 2410. *Resolución Conjunta Número 2410*. Bogotá, Colombia.

SISAIRE - IDEAM. (2018). *IDEAM*. Obtenido de Instituto de Hidrología, Meteorología y Estudios Ambientales: http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml