
Estructuración de un sistema de información geoespacial para el análisis de datos de seguridad alimentaria, intervenciones nutricionales y de salud humana en Panamá

González Ortega, Kevin

Universidad Tecnológica de Panamá
Ciudad de Panamá, Panamá
ORCID: 0000-0003-0437-5136

Aguilar, Eliecer

Universidad Tecnológica de Panamá
Ciudad de Panamá, Panamá

Aizprúa, Ana Gabriela

Universidad Tecnológica de Panamá
Ciudad de Panamá, Panamá

Cedeño, Eddy

Universidad Tecnológica de Panamá
Ciudad de Panamá, Panamá

Sánchez-Galán, Javier

Universidad Tecnológica de Panamá
Ciudad de Panamá, Panamá
javier.sanchezgalan@utp.ac.pa
ORCID: 0000-0001-8806-7901

Abstract

This research addresses the scarcity of studies related to food security and its connection to poverty in Panama, highlighting the lack of geospatial information systems to identify areas where nutritional interventions have been implemented. The main objective of the study is to organize databases from previously published information and human health data collected in various communities.

The work uses web scraping techniques and a process of data extraction, transformation and loading to align them with the geometries of the country's provinces. As a result, 96 geo-referenced indicators are obtained and visualized through maps, providing data on living standards, weight and height in the period 2003-2014.

The importance of this project lies in its capacity to expand knowledge on food security and human health in Panama, issues of high relevance for the country. In addition, the information generated will be incorporated as a layer in the geographic information system PRISM (Panama Research And Integrated Sustainability Model), which will facilitate interdisciplinary analysis by allowing integration with other existing layers in the model, such as water quality, biological connectivity and biodiversity.

In summary, this research article addresses a gap in the scientific literature related to food security in Panama by developing a methodology to organize geospatial data and provide a comprehensive view of the food security and human health situation in the country. This approach has the potential to inform future policy and action in the proposed areas.

Keywords: food security, web scraping, geo-referenced indicators, geographic information system, interdisciplinary spatial data analysis

Resumen

Esta investigación aborda la escasez de estudios relacionados con la seguridad alimentaria y su conexión con la pobreza en Panamá, destacando la falta de sistemas de información geoespacial para identificar áreas donde se han implementado intervenciones nutricionales. El objetivo principal del estudio es organizar bases de datos a partir de información previamente publicada y datos de salud humana recopilados en diversas comunidades.

El trabajo utiliza técnicas de web scraping y un proceso de extracción, transformación y carga de datos para alinearlos con las geometrías de las provincias del país. Como resultado, se obtienen 96 indicadores georreferenciados que se visualizan a través de mapas, proporcionando datos sobre niveles de vida, peso y talla en el período 2003-2014.

La importancia de este proyecto radica en su capacidad para ampliar el conocimiento

sobre seguridad alimentaria y salud humana en Panamá, temas de alta relevancia para el país. Además, la información generada se incorporará como una capa en el sistema de información geográfica PRISM (Panama Research And Integrated Sustainability Model), lo que facilitará el análisis interdisciplinario al permitir la integración con otras capas existentes en el modelo, como calidad del agua, conectividad biológica y biodiversidad.

En resumen, este artículo de investigación aborda una brecha en la literatura científica relacionada con la seguridad alimentaria en Panamá al desarrollar una metodología para organizar datos geospaciales y proporcionar una visión integral de la situación de la seguridad alimentaria y la salud humana en el país. Este enfoque tiene el potencial de informar políticas y acciones futuras en las áreas propuestas.

Palabras claves: seguridad alimentaria, web scraping, indicadores georreferenciados, sistema de información geográfica, análisis interdisciplinario de datos espaciales

1. INTRODUCCIÓN

El presente artículo aborda un proceso de recopilación y procesamiento de datos relacionados con intervenciones nutricionales en el ámbito de la salud pública. Para llevar a cabo este proceso, se emplearon técnicas de web scraping para obtener archivos PDF del Programa de Nutrición (PdN) del Ministerio de Salud de Panamá (MINSa). Estos archivos, inicialmente no estructurados, contenían información valiosa sobre indicadores nutricionales y datos geospaciales. A lo largo de este trabajo, se describirá detalladamente el procedimiento seguido, que incluye la extracción de tablas desde los PDF, la consolidación de datos con el uso de Pandas, y la obtención de archivos geospaciales para la creación de visualizaciones interactivas. Finalmente, se presenta un dashboard interactivo desarrollado con Plotly que facilita la exploración de los indicadores nutricionales y su representación geográfica.

2. METODOLOGÍA

A. Web Scraping de archivos PDF

Se emplearon técnicas de webscraping para recopilar los reportes de intervenciones nutricionales disponibles en el sitio web del Programa de Nutrición (PdN) del MINSa. Esta acción permitió obtener un conjunto inicial de datos en formato PDF para su posterior análisis.

B. Extracción de tablas con indicadores nutricionales.

Para obtener información específica contenida en las tablas dentro de los reportes, se utilizó

el software Tabula. Tabula utiliza técnicas de OCR (Reconocimiento Óptico de Caracteres) para extraer datos de tablas y los exporta en formato CSV. Este paso facilitó la obtención de datos estructurados a partir de documentos no estructurados.

C. Procesamiento de archivos que contienen los indicadores nutricionales.

Se empleó la biblioteca Pandas para combinar todos los archivos CSV generados en la etapa de extracción en un solo dataframe. Esta consolidación simplifica la gestión y análisis de los datos al presentarlos en un formato estructurado y uniforme.

D. Obtención y Procesamiento de Archivos Geoespaciales

Para obtener las geometrías utilizamos un repositorio abierto de información geográfica [1]. Una vez obtenido el GeoDataFrame procedimos a agregar los códigos ISO-3166-2 que sirvan como llave común al momento de crear las visualizaciones de tipo mapa. Adicionalmente, exportamos el GeoDataFrame a un archivo GeoJSON. Este archivo GeoJSON se utiliza como fuente de datos para la creación y visualización de mapas.

E. Unificación y Visualización de los indicadores y archivos geoespaciales.

Este archivo GeoJSON se utiliza como fuente de datos para la creación y visualización de mapas a través de un *dashboard* interactivo desarrollado con Plotly, un conjunto de bibliotecas de Python especializadas en la generación de gráficos interactivos. El *dashboard* proporciona una interfaz de usuario amigable que permite a los usuarios explorar los diferentes mapas y utilizando controles como el selector de los indicadores de nutrición, para una experiencia de usuario más cómoda.

3. RESULTADOS

Al completar la investigación se lograron obtener 96 indicadores de 3 documentos correspondiente a datos censales. En la Tabla 1 se muestran los datos tabulares extraídos del VI Censo de Talla 2007, de forma similar ocurre con las Tablas 2 y 3 con los datos tabulares extraídos de VII Censo Nacional de Talla 2014 y el Informe de Situación Nutricional en Grupos de Población en Panamá 2016 respectivamente. En la Figura 1 se muestra el *dashboard* interactivo desarrollado con Plotly que muestra los 96 indicadores obtenidos.

SEVERA EN ESTADO NUTRICIONAL DE ACUERDO A TALLA/EDAD DE ESCOLARES

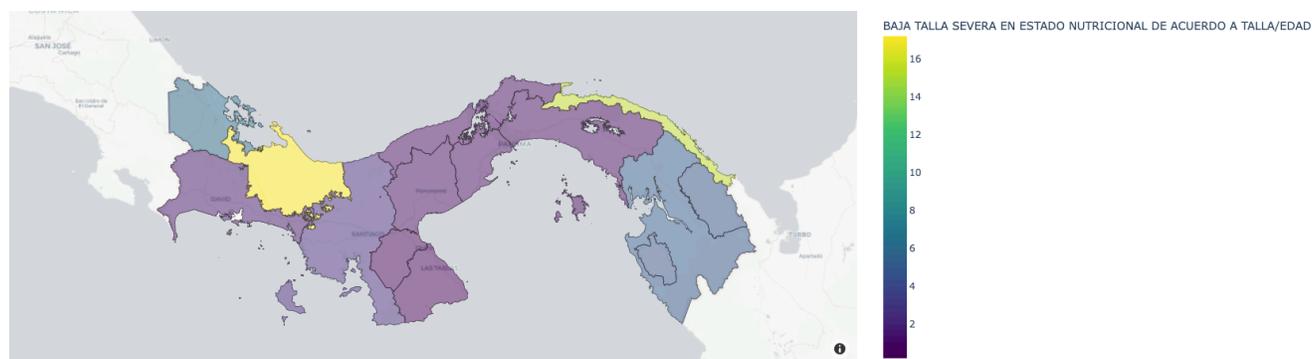


Fig. 1: Sistema de Información Geoespacial de Datos sobre Nutrición

Tabla 1. Tablas Extraídas del VI CENSO DE TALLA 2007.

Datos Tabulares Extraídos	Cantidad de Indicadores
Población total de escolares en primer grado de las escuelas oficiales y particulares que participaron en el VI censo nacional de talla de escolares de primer grado, Panamá, 2007.	7
Cobertura nacional y por provincia del VI Censo Nacional de Talla de Escolares de Primer Grado de 6 a 9 años, Panamá, 2007.	3
Prevalencia de retardo en talla por área, según provincia de los escolares de 6 a 9 años. VI Censo Nacional de Talla de Escolares de Primer Grado, Panamá, 2007.	8
Prevalencia de retardo en talla en escolares por provincia en las escuelas particulares VI Censo Nacional de Talla de Escolares de Primer Grado, Panamá, 2007.	6
Situación nutricional de escolares de 6 a 9 años según provincia. VI Censo Nacional de Talla de Escolares de Primer Grado, Panamá, 2007.	5
Prevalencia de retardo en talla por sexo, según provincia. VI censo de talla. Panamá, 2007.	6
Prevalencia de retardo en talla por edad, según provincia.	10

Tabla 2. Tablas Extraídas del VII CENSO NACIONAL DE TALLA 2014.

Datos Tabulares Extraídos	Cantidad de Indicadores
Cobertura nacional y por provincias DEL VII censo nacional de talla de escolares de primer grado de 6 a 9 años, panamá, 2013.	3
Talla promedio y mediana a los 7 años con 0 meses, por provincias y comarcas. VII censo nacional de talla, panamá 2013.	4
Estado nutricional de acuerdo con talla/edad de escolares según provincias y comarcas. VII censo nacional de talla, panamá 2013.	5

Prevalencia de baja talla por área geográfica según provincias y comarcas. VII censo nacional de talla, panamá 2013	9
Prevalencia de baja talla por edad según provincias y comarcas. VII censo nacional de talla, panamá 2013	12
Clasificación del nivel de prevalencia de baja talla según provincias y comarcas. VII censo nacional de talla, panamá, 2013	1
Diferencia y cambio porcentual de las prevalencias de baja talla entre los censos 2007 y 2013, provincias y comarcas. Panamá 2013	4
Prevalencia de baja talla por área, censos 2007 y 2013, según provincias y comarcas. Panamá	8

Tabla 3. Tablas Extraídas del INFORME DE SITUACIÓN NUTRICIONAL EN GRUPOS DE POBLACIÓN EN PANAMÁ 2016.

Datos Tabulares Extraídos	Cantidad de Indicadores
Estado nutricional en adultos según región de salud. Panamá 2014.	5

4. CONCLUSIONES

En este estudio, se muestra la efectividad de combinar técnicas de web scraping, procesamiento de datos y visualización interactiva para obtener una visión integral de los indicadores nutricionales en Panamá. La recopilación de datos a partir de documentos no estructurados, como los PDF, se facilitó mediante el uso de Tabula, el software de OCR. La consolidación de datos con Pandas permitió crear un formato uniforme para el análisis de los indicadores, y la obtención de datos geoespaciales enriqueció la presentación de información con visualizaciones de mapas.

El dashboard interactivo desarrollado con Plotly ofrece una herramienta accesible y amigable para que los usuarios exploren estos datos de manera intuitiva. En este contexto, se lograron obtener 96 indicadores de relevancia nutricional de diversos documentos censales. Este enfoque interdisciplinario y tecnológico contribuye a una comprensión más profunda y completa de los desafíos nutricionales en Panamá, facilitando así la toma de decisiones informadas en el ámbito de la salud pública y la nutrición.

REFERENCIAS

- [1] M. Solano, "Panama Province Boundaries - 2016." [Online]. Available: https://stridata-si.opendata.arcgis.com/datasets/74bc448a57914572b71fd3c949bce51c_0/explore

AUTORIZACIÓN Y LICENCIA CC

Los autores autorizan a APANAC XIX a publicar el artículo en las actas de la conferencia en Acceso Abierto (Open Access) en diversos formatos digitales (PDF, HTML, EPUB) e integrarlos en diversas plataformas online como repositorios y bases de datos bajo la licencia CC:

Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

Ni APANAC XIX ni los editores son responsables ni del contenido ni de las implicaciones de lo expresado en el artículo.