\documentclass{article}

\usepackage[utf8]{inputenc}

\title{NAUTIA}

\author{Guillermo Sánchez Gutiérrez-Cabello }

\date{Abril 2020}

\begin{document}

\maketitle

\section{Plataforma UPM sobre refugiados y NAUTIA}

A pesar de que Naciones Unidas establezca en 2018, en la Agenda 2030, los Objetivos del Desarrollo Sostenible (ODSs), solo en ese mismo año hubo 70.8 millones de personas forzadas a abandonar sus lugares de residencia.

El contexto histórico actual arroja datos alarmantes, pues 37.000 personas se ven obligadas diariamente a huir. De ellas, más de 26 millones de personas son reconocidas como refugiados. Según el Alto Comisionado de la Naciones Unidas para los Refugiados (UNHCR, por sus siglas en inglés) en 2019, la mayoría de esas personas eran menores de 18 años y 28.000 eran niños desacompañados. Todas ellas se han visto obligadas a emigrar por diversas causas como la pobreza, la violación de los derechos humanos, el cambio climático, desastres naturales, por fuertes crisis políticas y sociales o por guerras. Paradójicamente, muchos de estos desplazamientos se realizan hacia poblaciones más pobres que de las que provienen y, desgraciadamente, la mayoría de las comunidades de acogida para estos refugiados también están envueltas en crisis económicas y sociales. Por todo ello, resulta imperativo dar soporte a la integración entre campos de refugiados y comunidades de acogida.

Con la integración entre ambas comunidades como idea principal, se desarrolla la Declaración de Nueva York sobre Refugiados, la cual pone en marcha un Marco Integral de Respuesta de Refugiados (CRRF, por sus siglas en inglés).

En el ámbito de esta idea integrista nace la Alianza Shire, primera alianza multiactor de acción humanitaria en España en la que participan diversas entidades del sector público, privado y universitarios. Entre ellas, se encuentra la Plataforma UPM sobre Refugiados, que surgió como un grupo interdisciplinar para, precisamente, dar apoyo a la Alianza Shire.

Los objetivos de la plataforma se encuadran entorno a la mejora de la calidad de vida de los refugiados mediante la intervención con proyectos tecnológicos y del desarrollo de una metodología interdisciplinar que permite realizar el diagnóstico de soluciones integrales.

Así nace NAUTIA (Need Assessment under a Technological Interdisciplinary Approach). NAUTIA es una metodología dedicada a identificar de manera cuantitativa las necesidades básicas de los habitantes residentes en los campos de refugiados y comunidades de acogida para poder mejorar su calidad de vida a través de propuestas basadas en la tecnología.

Esta metodología se puso en práctica en el campo de Refugiados de Shimelba, en la región de Tigray (Etiopía), en octubre de 2018 y en dos campos saharauis: Smara y Aiun.

La aplicación de esta metodología ha puesto en evidencia lo complejo que es realizar dicha aproximación integral e interdisciplinar y, por lo tanto, lo necesario de implementar una herramienta que automatice dicho proceso. En concreto, en este trabajo, se explica el desarrollo de una gran parte la segunda versión de la herramienta, la cual está centrada en la automatización de la recogida de datos y detección de necesidades de campos y comunidades con características similares a Shimelba y Tigray. Esta herramienta se denomina como NAUTIA Tool Kit y sentará las bases para que en una versión posterior evolucione y pueda ser escalable a cualquier contexto, no solo africano, sino a campos de refugiados con localización en Europa, por ejemplo.

\section{Descripción de datos de origen y definición de requisitos}

Como se ha especificado anteriormente, la herramienta ya cuenta con una primera versión. Esta versión, si bien no está para nada automatizada, sirve muy bien para entender el propósito y necesidad de automatizar la metodología NAUTIA.

La manera de trabajar con esta novel versión de \textit{NAUTIA Tool Kit} sigue los siguientes pasos:

\begin{enumerate}

\item Proceso de entrevistas, cuestionarios y trabajo de campo para levantar datos en campos de refugiados y comunidades de acogida. Todo se realiza utilizando el entorno \textit{Open Data Kit (ODK)}.

\item Proceso de pasar a mano los resultados obtenidos del primer paso a una hoja excel relativamente estructurada y con determinados cálculos ya automatizados. Es importante darse cuenta que este paso significa duplicar el trabajo, porque cada tipo de dato se replica por un lado para campos de refugiados y por otro para la comunidad de acogida.

\item Proceso de calculo de indicadores utilizando otra hoja excel de características similares a la anterior.

\item Proceso de generación de informe y gráficos de manera manual con todos los resultados obtenidos previamente.

\end{enumerate}

Como se puede observar en este breve resumen de utilización de la herramienta para aplicar la metodología, es todo bastante ineficiente y muy dependiente de la acción humana, la cual suele estar sujeta muy frecuentemente a errores, lo que conlleva de manera inevitable no poder asegurar la ausencia de resultados erróneos y alejados de la realidad.

Bajo esta problemática, Susana Muñoz Hernández, profesora de la Escuela Técnica Superior de Ingeniero Informáticos de la UPM, directora del grupo de cooperación TEDECO y colaboradora en la Plataforma UPM sobre refugiados, propone a la plataforma el desarrollo de una beca de colaboración para poder desarrollar una herramienta informática que realmente automatice todo lo que sea susceptible de ello con respecto a la primera versión de la herramienta. Dicha beca la disfruté yo y a continuación detallo todo el proceso para entender los datos de origen y los requisitos necesarios para poder implementar la herramienta.

Es importante saber que la primera versión de la herramienta no fue desarrollada por profesionales informáticos. Aunque esta primera versión estaba bien implementada y recogía todo lo necesario para poder aplicar la metodología, tenía muchas carencias a la hora de afrontar un desarrollo informático. La trazabilidad de los datos entre distintas capas de la herramienta era baja, la estructuración de los datos en la hojas Excel en ningún momento aseguraba la integridad de los datos y, por último, no existía ningún documento de requisitos.

Con todo esto sobre la mesa, se aplicó ingeniería inversa haciendo un estudio de dichas hojas Excel y manteniendo reuniones, alguna de ellas superior a las dos horas, con las dos becarias de la plataforma que habían desarrollado de manera integral la primera versión.

Los datos nacen de los siguientes cuestionarios, los cuales se anexan a este documento, realizados con la aplicación para móviles android \textit{ODK Collet} en terreno:

\begin{enumerate}

\end{enumerate}

\begin{enumerate}

\item Busines survey.

\item Local leaders survey.

\item Entities survey.

\item General form.

\item General cicizen focus group survey.

\item Women focus group survey.

\item Household survey.

\item Sanitation infrastructure survey.

\item Water infrastructure servey.

\item Waste management survey.

\item Energy infrastructure survey.

\item Shelter survey.

\item Public space survey.

\item Priorities Survey.

\item Farmyard and crops survey.

\item Comunal services survey.

\item Transport services survey.

\end{enumerate}

La aplicación permite realizar y almacenar los resultados de todas estas encuestas sin necesidad de internet. Posteriormente, una vez se tiene acceso a la red, la aplicación permite subir los resultados a un servidor de ODK (\textit{ODK aggragate}) donde éstos se pueden exportar en formato CSV o XML.

Adicionalmente a esos 17 cuestionarios, existe otra fuente de datos, la cual también se anexa al documento y ésta es una bibliografía en formato de tabla Excel. Antes de continuar analizando la gran hoja excel que es el centro de la herramienta y de donde se extraen prácticamente todos los requisitos, es necesario comentar todos los cambios de formato que se han hecho en esta bibliografía para asegurar la ausencia de errores o duplicidad en los datos (el fichero con el registro de cambios realizados de manera detallada también se anexa):

\begin{itemize}

\item Reordenación de filas correspondientes a población disgregada por sexo para que tenga un orden coherente.

\item Renombre de todos los nombres de columnas duplicados para hacerlos únicos y así evitar duplicidad de índices a la hora de trabajar con \textit{DataFrames} durante el proceso ETL.

\item Restricción de ciertas entradas de datos para limitar su rango. Ejemplos: campos de sí o no, campos con formato fecha o campos de regiones climáticas para evitar dos referencias distintas a la misma zona.

\end{itemize}

Posteriormente, toda esta información recabada se vuelca en en una hoja de cálculo Excel cuyo nombre es "20190822\\_Edición Listado de Datos.xlsx" y se estructura de la siguiente manera

\begin{enumerate}

\end{enumerate}

\begin{itemize}

\item El archivo cuenta con 12 hojas diferentes. La primera de ellas es "INSTRUCCIONES" y, la segunda, un índice que referencia al resto

\item Hoja "0. Generales": Recoge todos los datos del contexto general del país con el objetivo de contextualizar la situación en el ámbito estatal.

\item Hoja "0. Generales Campo Refugiados": Es la única con datos exclusivos del campo del refugiados y éstos son de carácter general.

\item Hoja "1. Sociales y Económicos": Recoge, por duplicado, datos de demografía, culturales, seguridad personal, economía y de género.

\item Hoja "2. Gobernanza": Recoge, por duplicado, políticas públicas (no se trata en este trabajo) y actores políticos involucrados.

\item Hoja "3. Físicos y Medioambientales": recoge, por duplicado, datos de Identificación geográfica, topografía y recursos naturales.

\item Hoja "4. Urbanísticos": Recoge, por duplicado, datos sobre información general, uso del suelo, viales y espacio público urbano.

\item Hoja "5. Infraestructuras": Recoge por duplicado, datos de agua, saneamiento y drenaje, gestión de residuos, energía (de manera triplicada, pues se divide en los sectores residencial, comercial y servicios comunitarios) y movilidad y transporte.

\item Hoja "6. Servicios": Recoge, por duplicado, datos educativos (de manera triplicada en escuela primaria, secundaria y profesional), sanitarios, socioculturales, deportivo/recreativos, religiosos, sobre cementerios y sobre tecnologías de la comunicación (TIC).

\item Hoja "7. Alojamiento": Recoge, por duplicado, datos generales, de calidad constructiva y sobre el grado de satisfacción.

\item Hoja "8. Seguridad Alimentaria": Recoge, por duplicado, datos de antecedentes, comida, fuentes de alimentación, ubicación física de cultivos y ganados y sobre la continuidad del acceso a los alimentos y tecnología.

\end{itemize}

Antes de continuar definiendo esta capa de la herramienta, me gustaría aclarar por que se hace inca pié en la recogida duplicada o triplicada de los datos. Pues en vez de haber un lugar donde se ponga un identificador único de la comunidad, independientemente de que sea de acogida o de refugiados, para una fila de datos existen dos columnas, una para el campo y otra para la comunidad de acogida. De esta manera, la herramienta queda totalmente limitada a que la relación de cardinalidad entre comunidades es 1:1, lo cual queda muy alejado de la realidad además. Por otra parte, cuando se hace referencia, por ejemplo a los datos triplicado sobre cuestiones educativas, en vez de existir un identificador único para los centros y un campo de datos sobre que tipo de educación es, se replican las filas de datos por cada tipo de escuela. Estos son algunos de los ejemplos que ponen de manifiesto la complejidad de uso de la herramienta y la necesidad de escalarla a una con fundamentos informáticos fuertes.

Continuando con la definición de esta hoja Excel, esta se estructura de la siguiente manera:

\begin{itemize}

\item La primera columna es la categoría fuente, donde se especifica de donde provienen los datos, la mayoría de las veces es un cuestionario específico. Otras especifica "trabajo de campo" el cual engloba un conjunto de formularios distintos y, dependiendo del contexto,se refiere a unos o a otros. Pero, en ningún caso, se especifica a que cuestionario se refiere según el contexto. Por último, a veces especifica que la categoría fuente es un cálculo, es decir es un dato derivado de dos o más datos ya existentes.

\item La segunda columna es el número de dato de contexto, su código.

\item La tercera columna es la descripción del dato de contexto a recopilar.

\item La cuarta columna es información del dato con respecto al campo de refugiados. A veces, de manera irregular, esta columna se divide en otros campos.

\item La quinta columna es exactamente igual que la cuarta pero en referencia a la comunidad de acogida.

\item La sexta columna son comentarios

\item De la séptima en adelante es una guía sobre la metodología para la recogida del dato concreto.

\end{itemize}

Una vez están recogidos y representados todos los datos en esta capa de la herramienta, se realiza el cálculo de indicadores, los cuales representan la situación real y las necesidades de las comunidades. Representando además las brechas que existan entre comunidad de acogida y campo de refugiados.

Estos indicadores se representan en otra hoja Excel: "(NO MODIFICAR) Edición Indicadores V1.xlsx", que se estructura de la siguiente manera:

\section{Proceso de diseño de la base de datos}

.

\section{Diseño lógico propuesto de BD}

.

\section{Creación del esquema de BD}

.

\section{Proceso ETL (Programas Python y scripts SQL)}

.

\section{Proceso de muestra y cálculo de indicadores}

.

\section{Cuadro de mando}

.

\section{Aproximación al diseño de Data Warehouse}

.

https://blogs.upm.es/refugiadosupm/quienes-somos/

http://www.itd.upm.es/alianzashire/quienes-somos/

\end{document}