SIW - P8

Primer Ejercicio

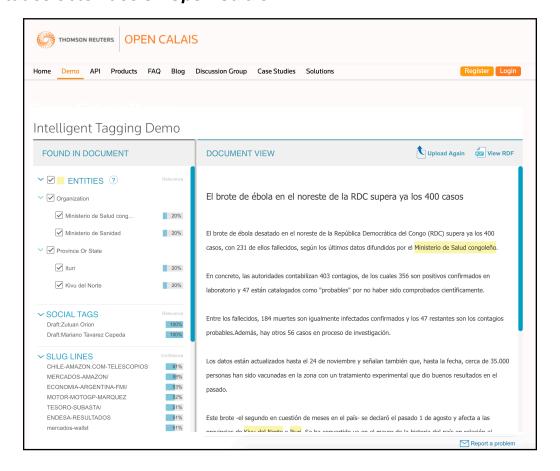
Herramienta Open Calais

Enlace:

https://www.eldiario.es/sociedad/brote-ebola-noreste-RDC-supera_0_839566057.html

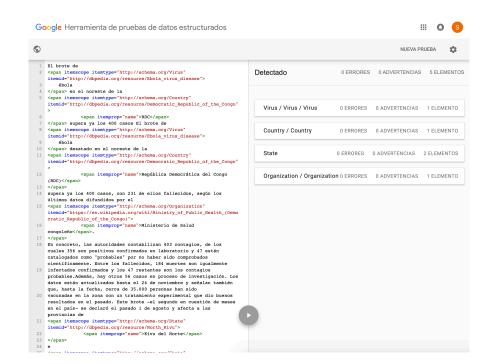
Texto seleccionado: "texto-P8.txt"

Resultados obtenidos en Open Calais:





Texto plano con etiquetado HTML usando microdatos + GSDTT:

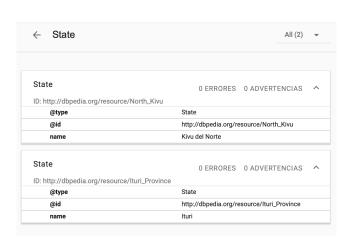


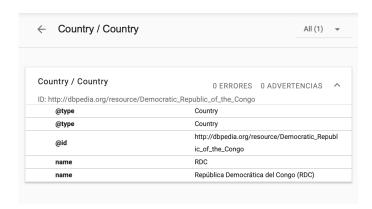
La herramienta Google Structured Data Testing Tool valida el HTML etiquetado que he realizado para el artículo ("UO251370-P8-Ejercicio1.html").

La herramienta Open Calais no identificó ni el país RDC y el virus/ enfermedad ébola, ambos los añade en este etiquetado que valida la herramienta de Google.









Cuestiones planteadas:

¿Deberían tener valor para itemid todos los ítems?

Sí, representa un ente único dentro de tipo que se le ha asignado, por ejemplo un persona o un lugar, lo que lo transforma en algo necesario.

¿Qué valor debería asignarse?

Un valor que le represente de forma única y que este ligado al tipo que se le ha asignado.

¿Si hubiera varios candidatos cuál escogerías? ¿Por qué?

Normalmente, si se trata de una utensilio, país, y cosas por el estilo, la más actual, que suelen ser los procedentes de Wikidata, porque así garantizamos que la información es lo más complementa posible.

(En mi caso he usado dbpedia ya que la url es más comprensible)

Si se trata de una persona y posee web personal, ésta sería nuestra primera opción, si es así con la página de wikipedia o similares sería suficiente.

¿Cuál crees que es la solución de compromiso que podría resultar menos polémica en la mayoría de casos?

Si existe en Wikidata, poner esta, si no mirar en dbpedia, y si no está en ninguno de los sitio recurrir a la página de wikipedia o a la web propiedad de la entidad a identificar.

¿Qué inconvenientes concretos te ha supuesto la obligación de incrustar los metadatos allí donde te "forzaba" la estructura del texto?

Un problema fueron la fechas, probé a poner el tipo Date de <u>schema.org</u> en las dos fechas que aparecen pero como éstas no están escritas en el formato: "ISO 8601 date format", la herramienta de Google no los detectaba. Por esta razón quite dichas etiquetas.

Segundo Ejercicio

El archivo .json modificado es: "UO251370-P8-Ejercicio2.json" El archivo .html es: "UO251370-P8-Ejercicio2.html"

La herramienta de Google lo valida con dos advertencias para los artículos que no poseen imagen:

