Documentación proyecto de obtención de información de figuras

Coordinación de Investigación, Desarrollo y Tecnología

# Pre-requisitos

* Curso por inspeccionar
* Token de un usuario de Canvas

# Requisitos

* Las páginas de semana deben estar ordenadas correctamente en la sección de Módulos.
* Si una página tiene una información adicional en otra página, debe existir una etiqueta <a> con el atributo “data-api-returntype” y su valor debe ser “True”, ignorando los enlaces que contengan la palabra “semana”. Esto permite almacenar las imágenes de las páginas externas en el orden correspondiente.
* Las imágenes que se mencionan en el contenido deben ubicarse dentro de la carpeta “Imágenes” para ser consideradas.
* Las imágenes deben estar siempre en una etiqueta <p> para poder localizar el título y la nota.
* El título de la imagen debe estar en una etiqueta <p> antes de la etiqueta contenedora de la imagen.
* La nota de la imagen debe estar en una etiqueta <pre> y a continuación de la etiqueta contenedora de la imagen.

# Características

* Se almacena información solo de las páginas de “Semanas” y páginas que estén llamadas internamente. Se excluye información de:
  + Foros
  + Tareas
  + Cuestionarios
* Las páginas que se omiten en el algoritmo son:
  + Inicio
  + Información general de la asignatura
  + Perfil del profesor
  + Horario de tutorías
  + Créditos
  + Versión offline
  + Glosario
  + Referencias bibliográficas
* La información de las semanas se la toma de acuerdo con la organización de los módulos del curso.
* Actualmente solo funciona para cursos en español y en inglés.

# Recomendaciones

* Realizar un control de calidad de la extracción, ya que existen tablas o recursos que constan como imágenes, mismas que el algoritmo no va a poder identificar ni diferenciar.
* Validar que la extracción de la información adicional (título, nota y texto alternativo) haya sido generado correctamente, ya que la forma de virtualizar de estos elementos varía en los cursos.
* Si existe información que no se ha considerado en los documentos resultantes, informar a la coordinación para revisar y corregir.

# Resultado

* Se guarda un documento Word con el código SIS del curso. Por ejemplo: PERI\_1104\_META.docx

# Ejemplo

* **Curso:**Teoría de la Imagen
* **Tiempo de generación:** 2 minutos con 45 segundos
* **Número de imágenes:** 67 imágenes
* **Número de páginas:** 12 páginas