

Introducción a Learning Analytics con  
ejemplos prácticos

# **UD 03. Anexo 01 - Extrayendo información de Itaca**

---

Autor: Sergi García Barea

Actualizado Septiembre 2020

## Licencia




**Reconocimiento – NoComercial - CompartirIgual (BY-NC-SA):** No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.

## Nomenclatura

A lo largo de este tema se utilizarán distintos símbolos para distinguir elementos importantes dentro del contenido. Estos símbolos son:

 **Importante**

 **Atención**

 **Interesante**

## ÍNDICE DE CONTENIDO

<b>1. Introducción</b>	<b>3</b>
<b>2. ¿Que es Itaca?</b>	<b>3</b>
<b>3. Itaca: informe de incidencias e informe de tutoría</b>	<b>3</b>
<b>4. La biblioteca de Python "Tabula-py"</b>	<b>3</b>
<b>5. Programas: "extractorINF_ALUMNOS.py" y "extractorINF_TUTORIA.py"</b>	<b>4</b>

## UD03. ANEXO 1 - EXTRAYENDO INFORMACIÓN DE ITACA

### 1. INTRODUCCIÓN

En este punto vamos a tratar la parte de extracción de información de PDFs de Itaca.

Este punto tiene una parte técnica de programación, pensada especialmente para profesores con conocimientos de informática.

Si eres un profesor de otra rama y no tienes conocimientos de programación, no te preocupes, puedes seguir el curso sin problema e ignorar esta parte.

Si aún así, eres valiente y quieres mirar este documento, te recomendamos que te centres en la parte de ejecutar los programas ya generados y que no te estresen otros detalles que no comprendas :)

### 2. ¿QUE ES ITACA?

Itaca es un sistema de gestión de la docencia (profesorado, alumnado, calificaciones, asistencia, etc.) utilizado en el sistema educativo público de la Comunidad Valenciana.

El sistema Itaca nos proporciona mucha información, pero no siempre en el mejor formato para poder realizar análisis de datos.

En este anexo comentamos tanto algunas fuentes de información presentes en Itaca , así como algunas herramientas técnicas que nos pueden ayudar a extraer información de ella.

Por último, presentamos dos programas de cosecha propia para extraer información.

### 3. ITACA: INFORME DE INCIDENCIAS E INFORME DE TUTORÍA

Itaca, (siempre según el rol de tutor) permite extraer en PDF dos informes que nos pueden ser interesantes:

- **Listado de incidencias:** genera un PDF con un nombre del estilo “**INF\_ALUMNOS.pdf**”. Contiene un listado de incidencias de un alumno en todas las materias.
- **Informe de tutoría :** genera un PDF con un nombre del estilo “**INF\_TUTORIA.pdf**”. Además de contener datos estadísticos interesantes, contiene un pequeño mapa de faltas de asistencia desglosadas por asignaturas.

### 4. LA BIBLIOTECA DE PYTHON “TABULA-PY”

Python (Version 3) es un lenguaje de programación ampliamente usado y muy potente para tareas como el análisis de datos <https://www.python.org/>

Python posee un gran número de bibliotecas para trabajar con PDFs: PDFMiner, PyPDF2, Tabula-py, etc. En este enlace [https://medium.com/@umerfarooq\\_26378/python-for-pdf-ef0fac2808b0](https://medium.com/@umerfarooq_26378/python-for-pdf-ef0fac2808b0) se pueden encontrar referencias a varias de ellas.

Para procesar los PDFs descritos anteriormente, hemos utilizado la biblioteca Tabula-py, cuya


documentación asociada es <https://tabula-py.readthedocs.io/en/latest/>. El motivo de utilizar esta biblioteca de extracción de información en PDFs, es que pasaba las tablas de PDFs a formato CSV, fácilmente utilizable en aplicaciones como KNIME <https://www.knime.com/> descrita en anteriores unidades.

## 5. PROGRAMAS: “EXTRACTORINF\_ALUMNOS.PY” Y “EXTRACTORINF\_TUTORIA.PY”

Para ejecutar estos programas, debes tener el lenguaje de programación Python instalado y lanzarlos de forma “python extractorINF\_ALUMNOS.py” o “python extractorINF\_TUTORIA.py”.

Sobre cada uno de estos ficheros:

- **“extractorINF\_ALUMNOS.py”**
  - Espera que en el directorio actual este un fichero llamado “INF\_ALUMNOS.pdf”. Generará 3 ficheros: “INF\_ALUMNOS\_bruto.csv” con los datos en bruto, “INF\_ALUMNOS\_procesado.csv” con un ligero procesado y “INF\_ALUMNOS\_anonimizado.csv” que es como el anterior, pero anonimizando alumnos.
- **“extractorINF\_TUTORIA.py”**
  - Espera que en el directorio actual esté un fichero llamado “INF\_TUTORIA.pdf”. Generará 3 ficheros: “INF\_TUTORIA\_bruto.csv” con los datos en bruto, “INF\_TUTORIA\_procesado.csv” con un ligero procesado y “INF\_TUTORIA\_anonimizado.csv” que es como el anterior, pero anonimizando alumnos.

 **Interesante:** Si quieres saber más información sobre cómo funcionan estos programas, puedes utilizar el foro y plantear tus cuestiones o contactar con el autor.