Visualizador de presencia de Universidades según rankings.

Informe Final

Proyecto Lenguajes de Programación

Integrantes:

* Fanny Valverde
* Isabela Vinces
* Pedro Mendoza

1. **Nombre del Proyecto:**

Visualizador de presencia de Universidades según rankings

1. **Justificación del Proyecto:**

A partir del desarrollo del proyecto es posible poner en práctica algunos de los temas revisados en clases, como son la extracción de información, y el procesamiento de esta, para luego llevar a cabo el desarrollo de una aplicación web que permita comparar la presencia de universidades alrededor del mundo según diferentes rankings, ya que no existe una herramienta que realice dicha actividad y facilite la búsqueda de universidades para licenciaturas, ingenierías, doctorados, etc.

1. **Objetivos:**

* Realizar web scrapping de diferentes páginas en el cual su contenido nos provea información útil para el análisis de presencia de universidades.
* Visualizar a través del análisis de rankings a nivel mundial, la presencia de las universidades.
* Realizar un procesamiento uniforme de la información y lograr una correcta organización de estos datos.
* Usar tres diferentes lenguajes de programación y hacerlos trabajar en conjunto para la realización de un visualizador de presencia de universidades.

1. **Alcance:**

El alcance del proyecto es el realizar un visualizador de presencia de universidades, a través de tres lenguajes diferentes, los cuales gracias a las facilidades que brindan cada uno. Nos ayudarán en el proceso de crear los 3 componentes que tiene el proyecto.

* Web Scrapping
* Procesamiento de la información.
* Aplicación web que permita comparar presencia entre universidades

El web scrapping se realizó de las páginas web de rankings: Webometrics y Shangai. De ahí se parte para construir el comparador esperado.

1. **Metodología**

Para la realización de proyecto usaremos los siguientes lenguajes para la creación de la aplicación web.

* Python para el web scraping en el cual se usará la librería BeautifulSoup. Los datos extraídos de las páginas seleccionadas serán almacenados en archivos persistentes, en este caso, de texto plano.
* Java para la lectura y procesamiento de la información, manejando estructuras de datos conocidas y útiles para la organización de la misma. Con esto se evaluará la selección y uniformidad de cada universidad en los diferentes rankings.
* Javascript se lo utilizará para la creación de la aplicación web que se encargará de mostrar los datos obtenidos ya analizados y organizados por región, por país, por área, antigüedad, investigación, internacionalización, alumno, etc.
* Se mostrará un gráfico comparativo con las posiciones globales entre las universidades que se estén comparando, entre distintos aspectos, además de también el ranking respectivamente en la categoría que se esté buscando. Este gráfico parte del top 100 de universidades que se van a presentar en la aplicación web.

1. **Resultados obtenidos.**

El web scrapping que se buscaba realizar en Python ocurrió exitosamente de manera que se pudo extraer tanto de Shangai como de Webometrics la información necesaria para construir nuestros rankings y las categorías que se íban a presentar en el proyecto, en archivos persistentes para luego ser leídos en otro lenguaje de programación en la siguiente etapa del proyecto.

De las páginas mencionadas se pudo obtener información de las universidades en el mundo por distintas categorías: por año, por país o región, áreas de especialización. Solo éstas pudieron ser extraídas de las páginas web y se las consideraron como las más importantes.

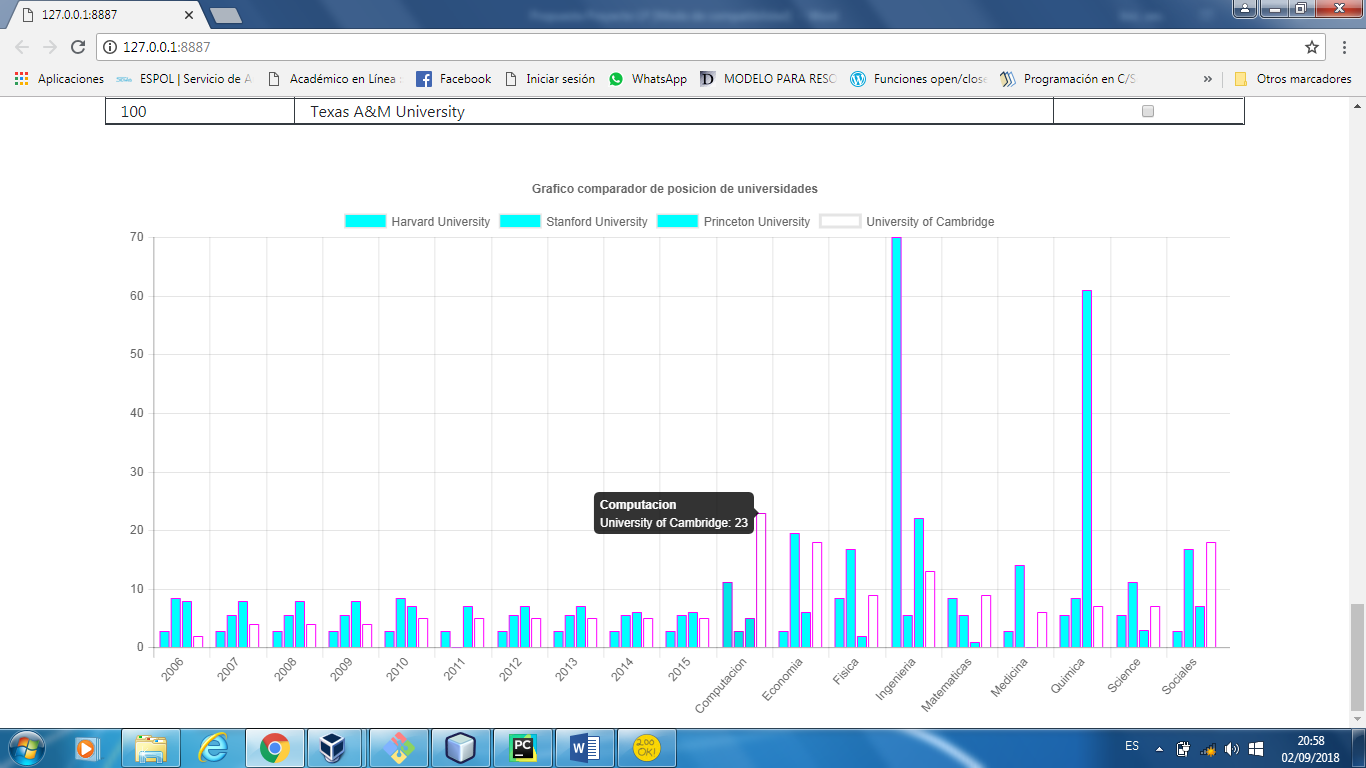
Con respecto al procesamiento de datos, se logró un correcto ordenamiento de la información, tal como se esperaba. Para esta etapa, como ya fue mencionado se utilizó Java y se utilizaron los archivos obtenidos del web scrapping para la lectura de la información. A través de una adecuada abstracción de la información, construimos dos objetos: universidades y posiciones. Seguido de esto, decidimos ordenar en una hashtable respectivamente, siendo la universidad la clave, y sus posiciones el valor. Con los datos ordenados, se encapsuló todo en varios archivos JSON para poder ser utilizado en la siguiente y última etapa.

Para la parte final, con los archivos JSON obtenidos se procedió a utilizar Javascript de manera que los datos procesados puedan ser mostrados por medio de una aplicación web básica. Esta aplicación web presenta las categorías anteriormente mencionadas y de cada categoría se pueden elegir distintas opciones según como se desee, y se mostrará a continuación una tabla con las 100 mejores universidades en la respectiva opción de la categoría escogida. Seleccionando los check box, entre una o más universidades de dicho top 100, aparece un gráfico de barras en la parte inferior con las posiciones en las distintas categorías, que es lo que se buscaba desde un principio.

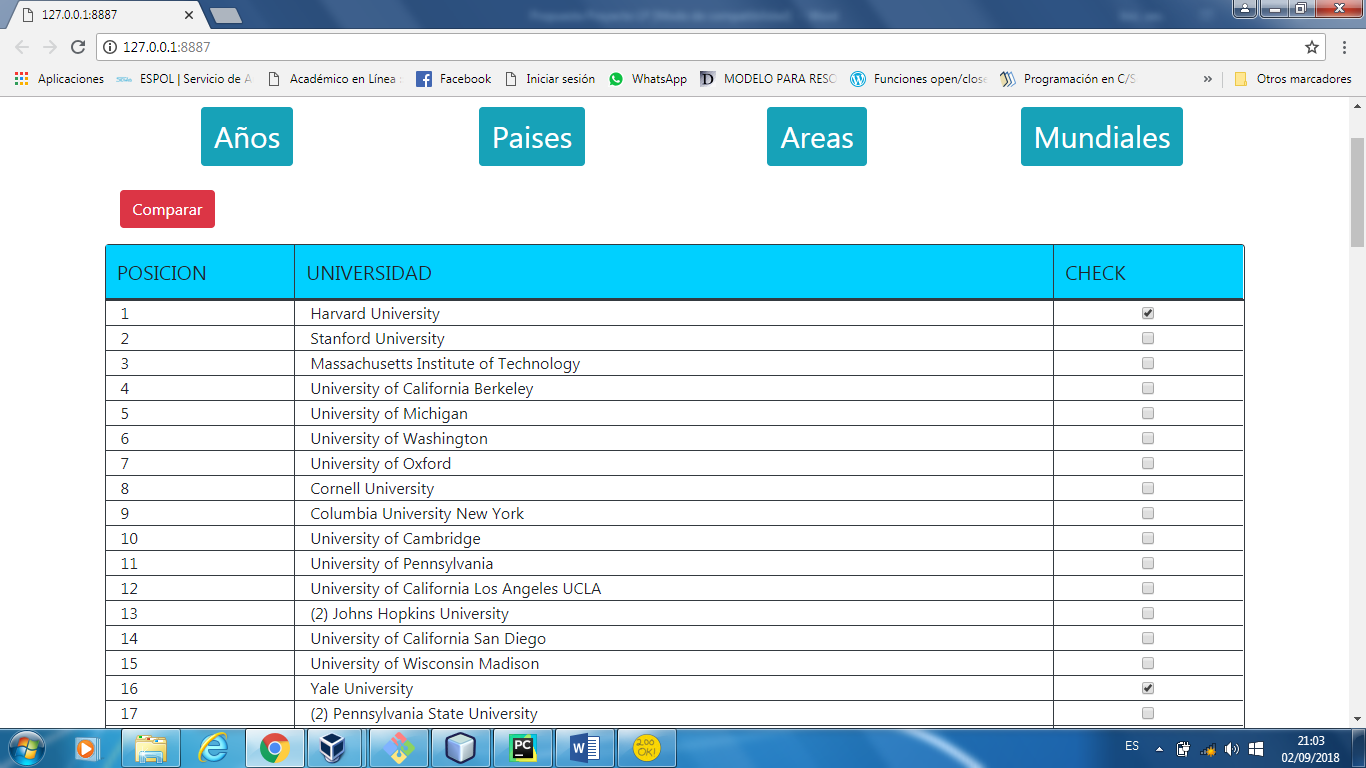
* Visualización de la aplicación.

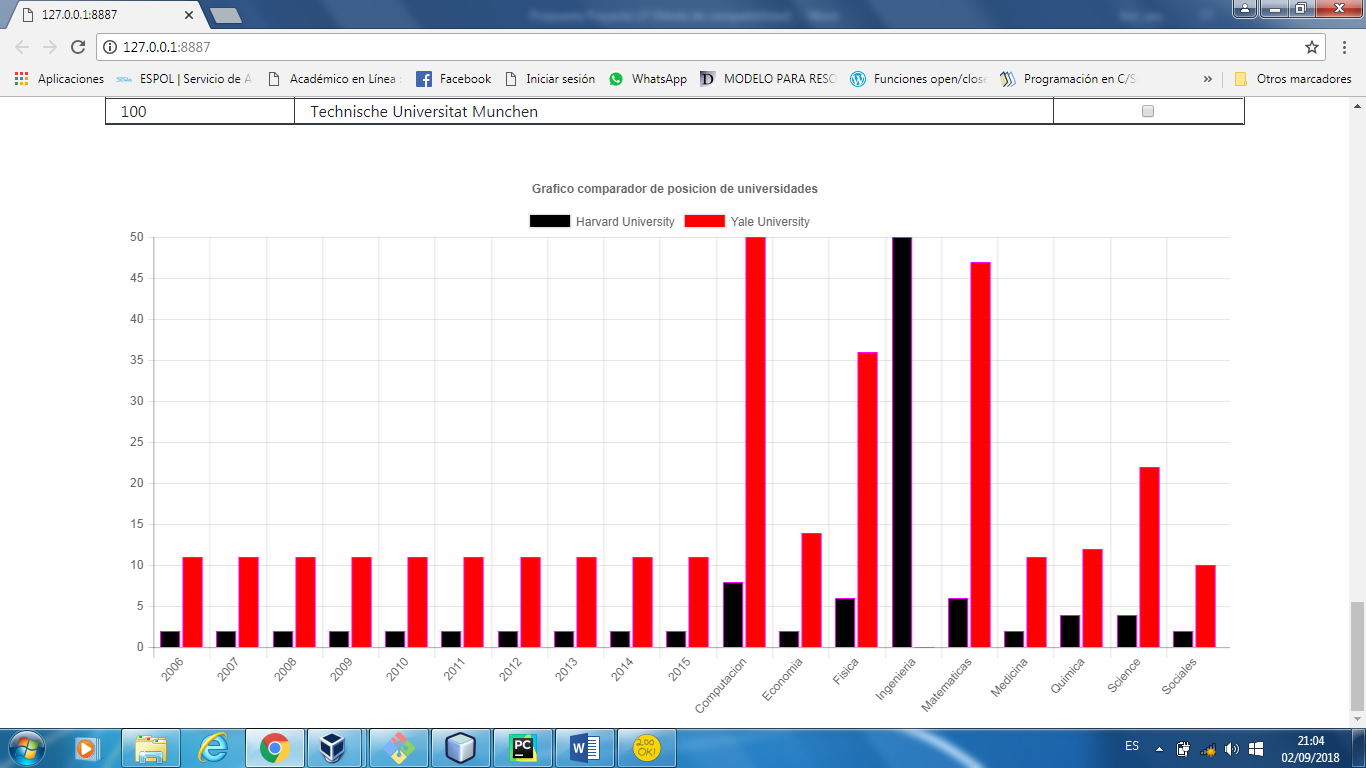
Por año.





Mundial.





1. **Conclusiones.**

* Se logró desarrollar correctamente el proyecto, tal y como se esperaba desde el inicio, cumpliendo los objetivos planteadas.
* Los rankings son una herramienta muy útil para medir calidad, específicamente de universidades, ya que la posición dice mucho sobre qué tan buena es y cuánto vale la pena estudiar en ella.
* Esta herramienta puede ser muy útil para tomar decisiones sobre qué universidad escoger en un futuro, si se piensa en universidades en el extranjero donde estudiar, hacer maestrías, doctorados, etc.
* El comparador puede servir para conocer la realidad entre distintas universidades, sobre en qué se desempeñan mejor, o cuál ha sido su evolución a través del tiempo o ver en qué área es mejor una universidad de la otra.
* Combinar varios lenguajes de programación para realizar un proyecto permite obtener los múltiples beneficios que estos ofrecen, o que se conozcan, de manera que se pueda presentar un trabajo más agradable y ordenado.

1. **Recomendaciones.**

* Es de suma importancia realizar una correcta extracción de datos, ya que esa será la base para proyectos de este estilo.
* En el momento de la extracción es importante revisar la codificación de los datos que se obtienen de la página web. Esto podría generar un error para las etapas posteriores si no se lo toma en cuenta.
* Se debe realizar un correcto análisis de cuál es la forma más óptima para organizar o procesar la información, de manera que esta pueda ser correctamente utilizada para los múltiples objetivos que se planteen.
* Es necesario tomar en cuenta los caracteres especiales que se originan
* Con un mayor conocimiento de Javascript se puede construir una aplicación web más dinámica para mejor presentación de la información.