

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA



FACULTAD DE ECONOMIA Y PLANIFICACION DEPARTAMENTO DE ESTADISTICA E INFORMATICA

Curso: Lenguaje de Programación II

Título: Trabajo Final

Profesor: Ana Cecilia Vargas Paredes

Integrantes:

- | | |
|-----------------------------------|----------|
| • Huazo Negron, Gianella Angelina | 20211817 |
| • Castillo Morin, André Benjamin | 20210829 |
| • Romero Cuenca, Mary Carmen | 20191312 |

Grupo: A

Fecha de entrega del informe: 08/01/2022

Lima - Perú

2022

DESCRIPCIÓN DEL CASO

Se quiere obtener información estructurada de las tesis de pregrado y postgrado de Estadística, Informática o de carreras similares desarrolladas a nivel nacional, a través de los repositorios institucionales de las universidades.

REPOSITORIOS ESTUDIADOS

- Universidad San Ignacio de Loyola

URL: <https://repositorio.usil.edu.pe/home>




- Universidad Nacional Mayor de San Marcos

URL: <https://cybertesis.unmsm.edu.pe/>



- Pontificia Universidad Católica del Perú

URL: <https://tesis.pucp.edu.pe>


PUCP
Login

Principal

Repositorio Digital de Tesis y Trabajos de Investigación PUCP

Elija un grado académico para listar sus colecciones

1. Doctorado
2. Maestría
3. Licenciatura
4. Bachillerato
5. Segunda Especialidad

Acerca del Repositorio

- Navegación
- Formato de tesis y trabajos
- Formulario de Autorización
- Preguntas Frecuentes

Búsquedas

DESARROLLO

Extracción de datos Universidad San Ignacio de Loyola

Agentes móviles para la recuperación personalizada de información en la municipalidad de Los Olivos - 2015

Página de artículo simple

dc.contributor.advisor	Fernández-Iparraguirre, Jaddy	
dc.contributor.author	Saucedo Meneses, Alessandra Seyla	
dc.contributor.author	Angulo Corzo, Rolando Anthony	
dc.date.accessioned	2017-10-12T18:39:05Z	
dc.date.available	2017-10-12T18:39:05Z	
dc.date.issued	2017	
dc.description.abstract	La presente investigación tuvo como propósito describir la utilización de los agentes móviles para la recuperación de la información personalizada en la Municipalidad de Los Olivos. Estudio descriptivo simple, con una muestra de 56 trabajadores del área de ofimática que laboran en la Municipalidad de Los Olivos. Se utilizó una encuesta que mide la utilización de agentes móviles en las dimensiones: bibliotecas digitales, agentes de software y agentes de usuario. Los resultados mostraron que el nivel de utilización de agentes móviles en sus dimensiones es bajo.	es_ES
dc.description.uri	Tesis	es_ES
dc.format	application/pdf	
dc.identifier.uri	https://repositorio.usil.edu.pe/handle/usil/2774	
dc.language.iso	spa	
dc.publisher	Universidad San Ignacio de Loyola	es_ES
dc.publisher.country	PE	es_ES
dc.rights	info:eu-repo/semantics/openAccess	

Cada una de las tesis del repositorio cuenta con un apartado especial conocido como los metadatos. En esta tabla, podemos ver cómo nuestros datos requeridos aparecen identificados con un nombre diferente, por lo tanto creamos los siguientes objetos:

```
links_usil = ['https://repositorio.usil.edu.pe/browse/typeuri?scope=5306afac-1d51-4564-ba61-d6185853d80&value=Tesis&bbm.rpp=80']
datos_requeridos = ["dc.publisher", "dc.title", "dc.contributor.author", "thesis.degree.name", "dc.contributor.advisor", "dc.description.abstract", "dc.date"]
datos_requeridos_2 = ["Asesor", "Nombre del tesista", "Año", "Resumen", "Institución", "Resumen", "Grado"]
```

Parte 1

Creamos una función que, dada una URL ingresada en ella, encuentre todas las URL de los metadatos de todas las tesis presentes.

Partimos inspeccionando el html de la página con una tesis.

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Escriba las primeras letras

Examinar

a.lead.item-list-title.ng-star-inserted 690 x 54.44

Agentes móviles para la recuperación personalizada de información en la municipalidad de Los Olivos - 2015

(Universidad San Ignacio de Loyola, 2017) Saucedo Meneses, Alessandra Seyla; Angulo Corzo, Rolando Anthony; Fernández-Iparraguirre, Jaddy

La presente investigación tuvo como propósito describir la utilización de los agentes móviles para la recuperación de la información personalizada en la Municipalidad de Los Olivos. Estudio descriptivo simple, con una muestra de 56 trabajadores del área de informática que laboran en la Municipalidad de Los Olivos. Se utilizó una encuesta que mide la utilización de agentes móviles en las dimensiones: bibliotecas digitales, agentes de software y agentes de usuario. Los resultados mostraron que el nivel de utilización de agentes móviles en sus dimensiones es bajo.

Artículo

Aplicación de IOT en el proceso de fertilización de mandarinas mediante un sistema de red de sensores

(Universidad San Ignacio de Loyola, 2021) Cruz Gavilan, Luis David; Maguina

La presente investigación tiene como objetivo proponer un sistema basado en Internet de las Cosas (IOT) que brinde apoyo a la gestión de fertilizantes como consecuencia de una problemática en la agricultura en la región del valle de

```
<ul _ngcontent-dspace-angular-c291 class="list-unstyled ng-tns-c291-2">
  <li _ngcontent-dspace-angular-c291 class="mt-4 mb-4 ng-tns-c291-2 ng-star-inserted">
    <ds-listable-object-component-loader _ngcontent-dspace-angular-c291 class="ng-tns-c291-2" _ngghost-dspace-angular-c128">
      <div _ngcontent-dspace-angular-c128">
        <ds-item-list-element _ngghost-dspace-angular-c206 class="ng-star-inserted">
          <ds-item-search-result-list-element _ngcontent-dspace-angular-c206 _ngghost-dspace-angular-c205">
            <ds-type-badge _ngcontent-dspace-angular-c205 class="ng-star-inserted">
              </ds-type-badge>
              <!-->
            <ds-truncatable _ngcontent-dspace-angular-c205 _ngghost-dspace-angular-c203 class="ng-star-inserted">
              <div _ngcontent-dspace-angular-c203 dsdragclick">
                <a _ngcontent-dspace-angular-c205 rel="noopener noreferrer" class="lead item-list-title ng-star-inserted" href="/items/e431c6dd-f82f-4712-8958-57ed3c52e955" target="_self"></a>
                <!-->
                <!-->
                <span _ngcontent-dspace-angular-c205 class="text-muted">
                  </span>
                <div _ngcontent-dspace-angular-c205 class="item-list-abstract ng-star-inserted">
                  <!-->
                </div>
              </ds-truncatable>
            </div>
          </ds-item-search-result-list-element>
        </ds-item-list-element>
      </li>
    </ul>
```

Para obtener el url de cada una de las tesis, primero buscamos todos los 'a' con el 'class' particular usando la función 'find_all()'. A cada uno de los resultados se le aplica 'get('href')' para conseguir parte del URL. Procedemos a completar cada link y almacenarlo a un objeto lista de la siguiente manera.

```
import requests
from bs4 import BeautifulSoup
def getTesisUsil(links):
    links_tesis = [] #creamos una lista vacia
    for link in links:
        html = requests.get(link).text #lo convierte a texto
        soup = BeautifulSoup(html, 'html.parser') #nos ayuda para la extraccion de datos
        links_tesis.extend(list(map(lambda x: 'https://repositorio.usil.edu.pe/'+x.get("href")+'/full', soup.find_all('a', class_="lead item-list-title"))))
        # .extend() agrega cada uno de los elemtos a la lista en el orden en el que aparecen
        # list(map(lambda x:.....,....)) es un filtro que aplica una condición a cada elemento de una lista
    return links_tesis

getTesisUsil(links_usil)
```

Parte 2

Creamos una función que extraiga de una URL de metadatos ingresada, todos los datos de la tabla que coincidan con 'datos_requeridos' definido al inicio de nuestro desarrollo.

Inspeccionamos el html de la tabla de metadatos.

Agentes móviles para la recuperación personalizada de información en la municipalidad de Los Olivos - 2015

td.ng-tns-c500-2	210.3 × 49.11	Página de artículo simple
dc.contributor.advisor	Fernández-Iparraquirre, Jaddy	
dc.contributor.author	Saucedo Meneses, Alessandra Seyla	
dc.contributor.author	Angulo Corzo, Rolando Anthony	
dc.date.accessioned	2017-10-12T18:39:05Z	
dc.date.available	2017-10-12T18:39:05Z	
dc.date.issued	2017	
dc.description.abstract	La presente investigación tuvo como propósito describir la utilización de los agentes móviles para la recuperación de la información personalizada en la Municipalidad de Los Olivos. Estudio descriptivo simple, con una muestra de es_ES	

```
rigger-fadeInOut ng-star-inserted">
▼<div _ngcontent-dspace-angular-c500 class="ng-tns-c500-2 ng-star-inserted" style>
  ▶<ds-item-alerts _ngcontent-dspace-angular-c500 class="ng-tns-c500-2" _ngghost-
    dspace-angular-c362>...</ds-item-alerts>
  ▶<ds-item-versions-notice _ngcontent-dspace-angular-c500 class="ng-tns-c500-2">...
    </ds-item-versions-notice>
  <ds-view-tracker _ngcontent-dspace-angular-c500 class="ng-tns-c500-2" _ngghost-
    dspace-angular-c412>...</ds-view-tracker>
  ▶<div _ngcontent-dspace-angular-c500 class="d-flex flex-row ng-tns-c500-2">...
    </div> <!--
  ▶<div _ngcontent-dspace-angular-c500 class="simple-view-link my-3 ng-tns-c500-2
    ng-star-inserted">...</div>
  <!-->
  ▼<table _ngcontent-dspace-angular-c500 class="table table-responsive table-strip
    ed ng-tns-c500-2">
    ▼<tbody _ngcontent-dspace-angular-c500 class="ng-tns-c500-2">
      ▼<tr _ngcontent-dspace-angular-c500 class="ng-tns-c500-2 ng-star-inserted">
        <td _ngcontent-dspace-angular-c500 class="ng-tns-c500-2">
          dc.contributor.advisor</td> == $0
        <td _ngcontent-dspace-angular-c500 class="ng-tns-c500-2">Fernández-
          Iparraquirre, Jaddy</td>
        <td _ngcontent-dspace-angular-c500 class="ng-tns-c500-2"></td>
        </tr>
        <!-->
        <!-->
        ▶<tr _ngcontent-dspace-angular-c500 class="ng-tns-c500-2 ng-star-inserted">...
          </tr>
        ▶<tr _ngcontent-dspace-angular-c500 class="ng-tns-c500-2 ng-star-inserted">...
          </tr>
        <!-->
        <!-->
        ▶<tr _ngcontent-dspace-angular-c500 class="ng-tns-c500-2 ng-star-inserted">...
          </tr>
        <!-->
        <!-->
```

De lo cual sabemos que cada 'tr' con su 'class' particular representa a cada fila de la tabla y, de estas, cada 'td' corresponde a un elemento de la fila. Buscamos entonces quedarnos con la parte 'string' de cada 'td' en una fila 'tr'. Si el 'string' coincide con 'datos_requeridos' entonces almacenaremos sus datos en un diccionario de la siguiente manera:

```
def getDatosDeTesis(linktesis): #Comenzamos la función
    datos = []
    html = requests.get(linktesis).text
    soup = BeautifulSoup(html, 'html.parser')
    #Creamos un objeto 'tabla' donde buscamos los 'tr' de class_="ng-tns-c505-8 ng-star-inserted"
    tabla = soup.find_all("tr", class_="ng-tns-c542-2 ng-star-inserted")

    #De cada fila en tabla buscamos todos los 'td', los volvemos string y enlistamos en 'columnas'
    for fila in tabla:
        columnas = list(map(lambda x: x.string, fila.find_all("td")))[0:2]
        #La primera columna de 'columnas' en datos
        dato = columnas[0]

        #Si 'dato' coincide con 'datos_requeridos' almacenar en 'datos'
        if(dato in datos_requeridos):
            datos.append(columnas)

        #Nos quedamos con la primera y segunda columna de 'datos', almacenando en varios diccionarios
    datos = [{item[0]:item[1]} for item in datos]
    respuesta = {} #nuevo diccionario

    #Insertamos los diccionarios como elementos del nuevo diccionario
    for dato in datos:
        respuesta.update(dato)

    #Enlistamos los valores del diccionario 'respuesta' y los emparejamos con 'datos_requeridos_2' con zip()
    #Se almacena todo a un nuevo diccionario 'respuesta_datos'
    respuesta_datos = dict(zip(datos_requeridos_2,list(respuesta.values())))

    return respuesta_datos #nos retorna respuesta_datos
```

Parte 3

Finalmente, en esta parte del desarrollo crearemos nuestro archivo .csv del dataframe creado con Pandas para ordenar todos los datos obtenidos cuando se aplican a todas las tesis.

```
import pandas as pd
#importamos la biblioteca pandas que nos permite estructurar los datos extraídos
todas_las_tesis = []
#creamos una lista vacía
links_tesis_usil = getTesisUsil(links_usil)
#asignamos a 'links_tesis_usil' los links extraídos por la función 'getTesisPucp'
for link in links_tesis_usil:
    todas_las_tesis.append(getDatosDeTesis(link))
#Aplicamos la función creada anteriormente "getDatosDeTesis(linktesis)" a cada uno de los links
#los resultados se almacenan a 'todas_las_tesis'
df = pd.DataFrame(todas_las_tesis)
#creamos una data frame
df.to_csv('usil_todas_las_tesis.csv', index=False, header=True, encoding="utf-8")
#guardamos el data frame en un archivo de extensión .csv
```

En el código, guardamos en un objeto todos los URL de metadatos que obtengamos con nuestra primera función 'getTesisUsil()' para que después a cada link en esta lista se le aplique la segunda función 'getDatosDeTesis()'. Todo lo obtenido se guarda en un objeto lista 'todas_las_tesis' para que, finalmente, este se convierta en dataframe con pandas.

	Asesor	Nombre del tesista	Año	Resumen	Institución	Grado
0	Fernández-Iparraquirre, Jaddy	Angulo Corzo, Rolando Anthony	2017	Agentes móviles para la recuperación personal...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
1	Fajardo Castro, Jimmy Robert	Maguiña Vilchez, Ernesto Andres	2021	Aplicación de IOT en el proceso de fertilizaci...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
2	Tena de Sebastián, Cecilia Marín	Pillaca Huamani, Víctor Alfredo	2018	Aplicación de ITIL v3.0 para mejorar la gestió...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
3	Escajadillo Muñoz, José Isaac	Pita Marchena, José Eduardo	2014	Aplicación de metodologías ágiles como soporte...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
4	Chavez Espinoza, Jonathan Jorge	Grandez Márquez, Miguel Ángel	2017	Aplicación de minería de datos para determinar...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
5	Tena de Sebastián, Cecilia Marín	Ríos Piñan, Darlyn Brigitte	2018	Aplicación del programa BCPVTX01 para elevar e...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
6	Díaz Arenas, Daniel Jesús	Pérez Valencia, Fernando David	2022	Controles del centro de seguridad de internet ...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
7	Zerpa Zerpa, Luis Jonathan	López Raez, Ruby Lucette	2018	Desarrollo de un aplicativo móvil para el anál...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
8	Prada Baca, Katia	Novoa Chuquiguel, Frank Dux	2022	Desarrollo de un método de recolección de imág...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
9	Soto Soto, Luis	Agustín Oscanoa, David José	2016	Desarrollo de una aplicación Android para incr...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
10	Jiménez Motte, Fernando	Ayuyque Arenas, Kevin José Marino	2016	Diseño de un sistema de clasificación de señal...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas

Si deseamos ver las principales estadísticas de lo hallado:

```
df.describe()
#Muestra la estadística del Dataframe df
```

	Asesor	Nombre del tesista	Año	Resumen	Institución	Grado
count	56	56	56	56	56	56
unique	31	56	13	56	1	1
top	Guevara Julca, José Zulu	Angulo Corzo, Rolando Anthony	2018	Agentes móviles para la recuperación personal...	Universidad San Ignacio de Loyola	Ingeniero Informático y de Sistemas
freq	4	1	12	1	56	56

Las siguientes extracciones cuentan con un proceso similar.

Extracción de datos Universidad Nacional Mayor de San Marcos

PARTE 1

Ubicar los url requeridos.



Códigos

```
links_unmsm = ["https://cybertesis.unmsm.edu.pe/handle/20.500.12672/129", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18699", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18698", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18037", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/17304", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/17060", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16679", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16552", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16433", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16064", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/15478", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/15474", "https://cybertesis.unmsm.edu.pe/handle/20.500.12672/10776"]
datos_requeridos = [{"dc.contributor.advisor", "dc.contributor.author", "dc.date.issued", "dc.description.abstract", "dc.publisher", "dc.subject", "dc.title", "dc.type"}, {"dc.subject", "dc.title", "dc.type"}]
datos_requeridos_2 = ["Asesor", "Nombre de tesista", "Año de publicación", "Resumen", "Insitución", "Titulo de la tesis", "Grado"]
```

```
import requests
from bs4 import BeautifulSoup
def getTesisUnmsm(links):
    links_Tesis = []
    for link in links:
        html = requests.get(link).text
        soup = BeautifulSoup(html, 'html.parser')
        links_Tesis.extend(list(map(lambda x: 'https://cybertesis.unmsm.edu.pe/' + x.find("a").get("href") + '?show=full', soup.find_all("a"))))
    return links_Tesis
getTesisUnmsm(links_unmsm)
```

```
['https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18699?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18698?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18037?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/17304?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/17060?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16679?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16552?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16433?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16064?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/15478?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/15474?show=full',
'https://cybertesis.unmsm.edu.pe/handle/20.500.12672/10776?show=full',
```

Activar Windows
Ve a Configuración para activar Windows.

Parte de web scraping, ubicar las direcciones requeridas en el archivo html.

[illegible]

Códigos

```
import requests
#Librerías
from bs4 import BeautifulSoup
def getDatosDeTesis(linktesis):
    #Comenzamos la función
    datos = []
    html = requests.get(linktesis).text
    soup = BeautifulSoup(html, 'html.parser')
    #Creamos una tabla donde buscamos los 'tr' de class="ds-table-row"
    tabla = soup.find_all("tr", class_="ds-table-row")
    for fila in tabla:
        columnas = list(map(lambda x: x.string, fila.find_all("td")))[0:2]
        #De cada fila en tabla buscamos todos los 'td', los volvemos string y enlistamos en 'columnas'
        #La primera columna de 'columnas' en datos
        dato = columnas[0]
        #Si 'dato' coincide con 'datos_requeridos' almacenar en 'datos'
        if(dato in datos_requeridos):
            datos.append(columnas)
            #Nos quedamos con la primera y segunda columna de 'datos', almacenando en varios diccionarios
    datos = [{item[0]:item[1]} for item in datos]
    response = {} #nuevo diccionario
    #Insertamos los diccionarios como elementos del nuevo diccionario
    for dato in datos:
        response.update(dato)
        #Enlistamos los valores del diccionario 'respuesta' y los emparejamos con 'datos_requeridos_2' con zip()
        #Se almacena todo a un nuevo diccionario 'respuesta_datos'
    response_data = dict(zip(datos_requeridos_2, list(response.values())))
    return response_data #fin de la función
getDatosDeTesis('https://cybertesis.unmsm.edu.pe//handle/20.500.12672/5144?show=full') #Probamos con un link de tesis
```

PARTE 3

Creamos un data frame para ubicar los datos extraídos en una tabla, para finalmente se guarde en un archivo csv.

```
import pandas as pd
#importamos la biblioteca pandas que nos permite estructurar los datos extraídos
todas_las_tesis = []
#creamos una lista vacía
links_tesis_unmsm = getTesisUnmsm(links_unmsm)
#asignamos la función
for link in links_tesis_unmsm:
    todas_las_tesis.append(getDatosDeTesis(link))
#Aplicamos la función creada anteriormente "getDatosDeTesis(linktesis)" a cada uno de los links
df = pd.DataFrame(todas_las_tesis)
#creamos una data frame
df2=df.dropna()
###eliminamos las filas que contienen NaN
df2.to_csv('unmsm_tesis.csv', index=False, header=True, encoding="utf-8")
#guardamos el data frame en un archivo de extensión .csv
```

VISUALIZACIÓN DE DATOS

df2

	Aesor	Nombre de testista	Año de publicación	Resumen	Institución	Titulo de la tesis	Grado
0	Mauricio Sánchez, David Santos	Bravo Mullo, Silvia Jeaneth	2019	Se analizaron seis aspectos sobre la detección...	Universidad Nacional Mayor de San Marcos	Contribuciones para la Detección de Ataques DI...	Doctor en Ingeniería de Sistemas e Informática
1	Mauricio Sánchez, David Santos	Oñate Andino, Mayra Alejandra	2022	Determina la influencia de los 13 factores cri...	Universidad Nacional Mayor de San Marcos	Contribuciones al gobierno de las tecnologías ...	Doctor en Ingeniería de Sistemas e Informática
2	La Serna Palomino, Nora Bertha	Ayovi Ramirez, Marco Wellington	2022	Pocas veces el éxito de una empresa surge de l...	Universidad Nacional Mayor de San Marcos	Modelo corporativo basado en gestión de la cal...	Doctor en Ingeniería de Sistemas e Informática
3	Pró Concepción, Luzmila Elisa	Garcia Quilachamin, Washington Xavier	2021	La revolución tecnológica del siglo XXI ha con...	Universidad Nacional Mayor de San Marcos	Algoritmo en detección positiva de la imagen d...	Doctor en Ingeniería de Sistemas e Informática
4	Rodríguez Rafael, Glen Darío	Rosero Miranda, Raúl Hernán	2021	Actualmente, en diferentes contextos, en mayor...	Universidad Nacional Mayor de San Marcos	Método para reducir pruebas de regresión de so...	Doctor en Ingeniería de Sistemas e Informática
5	Mauricio Sánchez, David Santos	Santisteban Pazos, José Luis	2021	El emprendurismo es un responsable directo de ...	Universidad Nacional Mayor de San Marcos	Contribuciones al desarrollo de una Startup de...	Doctor en Ingeniería de Sistemas e Informática
6	La Serna Palomino, Nora Bertha	Cadena Moreano, José Augusto	2021	La presente investigación se desarrolla en el ...	Universidad Nacional Mayor de San Marcos	Técnica eficiente para reconocimiento facial g...	Doctor en Ingeniería de Sistemas e Informática
7	Aguilar Alonso, Igor Jovino	Mera Macias, Angel Cristian	2021	Para desarrollar una buena gestión de servicio...	Universidad Nacional Mayor de San Marcos	Metodología para la construcción del catálogo ...	Doctor en Ingeniería de Sistemas e Informática
8	Gamboa Cruzado, Javier Arturo	Chancusig Chisag, Juan Carlos	2021	La investigación se encuentra enmarcada en la ...	Universidad Nacional Mayor de San Marcos	Implementación de un modelo de adopción de la ...	Doctor en Ingeniería de Sistemas e Informática
9	Bayona Oré, Luz Sussy	Morales Lozada, José Vicente	2020	Desarrolla un modelo conceptual integrado de l...	Universidad Nacional Mayor de San Marcos	Modelo conceptual integrado de e-gobierno para...	Doctor en Ingeniería de Sistemas e Informática

Extracción de datos Pontificia Universidad Católica del Perú

(Tabla de metadatos de una tesis PUCP)

Login

[Principal](#) /
 [3. Licenciatura](#) /
 [Facultad de Ciencias e Ingeniería](#) /
 [Ingeniería Informática](#) /
 [Ver ítem](#)

Implementación de un sistema de información para la gestión de unidades de manipulación en almacenes de grandes dimensiones



PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA

INSTITUTO TECNOLÓGICO

Implementación de un sistema de información para la gestión de unidades de manipulación en almacenes de grandes dimensiones

Tesis para optar por el título profesional de Ingeniero Informático que presenta
Paul Rodrigo Cárdenas Muela

Instituto:
Dr. Manuel Pinedora Tapia Andújar
Rg. María García Rivas

Lima, marzo del 2022

URI
<http://hdl.handle.net/20.500.12404/23555>

Un almacén cuenta con tres funciones importantes. La primera está relacionada a la recepción, se debe realizar una verificación registrando la cantidad y las características de los productos recibidos para después ser almacenados. La segunda función abarca el control y gestión de los productos dentro de la infraestructura física. La tercera se refiere a la selección de productos a distribuir y el monitoreo de la entrega de estos al cliente. Una de las formas más comunes de administrar productos en un almacén de gran envergadura es mediante el uso de unidades de manipulación. Estas estructuras han sido utilizadas ampliamente como un importante equipo de logística unitaria, puesto que mejoran el proceso de carga y transporte de bienes. Además, su uso puede reducir la tasa de daños y errores, la necesidad de mano de obra y otros costos de manipulación. Existen varias consideraciones a tomar en cuenta al momento de recepcionar y despachar algún producto. Los almacenes pueden tener diferentes tipos de infraestructura y cada área dentro del almacén puede ser expuesta a distintos factores ambientales. Además, existen productos que deben estar distanciados de otros por su naturaleza, por ejemplo, no es posible mantener productos de consumo humano cerca de productos de limpieza. Otro punto a considerar son las políticas de rotación de productos, debido a que también influyen en la ubicación que deben tener en el almacén. Partiendo de esta problemática, el objetivo de este trabajo de fin de carrera es implementar un sistema de información para el soporte a la gestión de operaciones en la recepción y relación de productos para el despacho en almacenes de grandes dimensiones.

Acerca del Repositorio

- Navegación
- Formato de tesis y trabajos
- Formulario de Autorización
- Preguntas Frecuentes

- ☒ Buscar en tesis PUCP
- ☐ Esta colección

LISTAR

Todo Tesis PUCP

- Grados académicos y colecciones
- Por fecha de publicación
- Autores

Ver/
 Texto completo (5.430Mb)

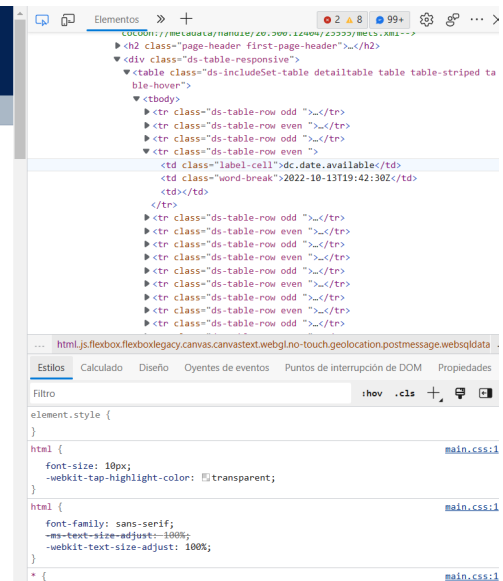
Fecha
 2022-10-13

dc.contributor.advisor	Manrique Villavicencio, Lileya
dc.contributor.author	Colan Sanchez, Carla Andrea
dc.date.accessioned	2022-12-19T12:53:46Z
dc.date.available	2022-12-19T12:53:46Z
dc.date.created	2022
dc.date.issued	2022-12-19
dc.identifier.uri	http://hdl.handle.net/20.500.12404/23932
dc.description.abstract	La retroalimentación del aprendizaje es un proceso complejo del enfoque formativo de la evaluación. Dicho proceso consiste en detectar los aprendizajes logrados y los aspectos de mejora de los estudiantes en relación con los criterios establecidos para proporcionarles información suficiente a fin de provocar mejoras en su aprendizaje. El objetivo de la presente investigación es describir las percepciones docentes sobre el proceso de retroalimentación del

Mostrar el registro sencillo del ítem

Implementación de un sistema de información para la manipulación en almacenes

dc.label.cell	174.73 x 37
Color	#333333
Font	14px "Helvetica Neue", Helvetica, Arial, s...
Padding	8px
ACCESSIBILITY	
Contrast	Aa 12.63 
Name	dc.date.available
Role	gridcell
Keyboard-focusable	
dc.date.available	2022-10-13T19:42:30Z
dc.date.created	2022
dc.date.issued	2022-10-13
dc.identifier.uri	http://hdl.handle.net/20.500.12404/23555
dc.description.abstract	Un almacén cuenta con tres funciones importantes. La primera está relacionada a la recepción, se debe realizar una verificación registrando la cantidad y las características de los productos recibidos para después ser almacenados. La segunda función abarca el control y gestión de los productos dentro de la infraestructura física. La tercera se refiere a la selección de productos a distribuir y el monitoreo de la entrega de estos al cliente. Una



Código que será usado para sacar la información

Parte '0'

Empezamos nuestro código definiendo 3 objetos lista almacenados según el repositorio de la universidad escogida y de la manera en que hayan guardado los ficheros de tesis.

```
In [1]: links_pucp = ["https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/14", "https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/14/recent-subm",
datos_requeridos = ["dc.publisher", "dc.title", "dc.contributor.author", "thesis.degree.name", "dc.contributor.advisor", "dc.description.abstract", "dc.date",
datos_requeridos_2 = ["Asesor", "Nombre del tesista", "Año", "Resumen", "Institución", "Resumen", "Grado"]
```

Parte 1

Con el siguiente código, creamos una función que nos extraiga los links en donde se encuentran los ficheros de cada una de las tesis, almacenandose en una lista

```
In [2]: import requests
from bs4 import BeautifulSoup
def getTesisPucp(links):
    links_Tesis = [] #creamos una lista vacia
    for link in links:
        html = requests.get(link).text #lo convierte a texto
        soup = BeautifulSoup(html, 'html.parser') #nos ayuda para la extraccion de datos
        links_Tesis.extend(list(map(lambda x: 'https://tesis.pucp.edu.pe'+x.find("a").get("href")+'?show=full' , soup.find_all("h4", class_="artifact-title")
# .extend() agrega cada uno de los elementos a la lista en el orden en el que aparecen
# list(map(lambda x:.....,....)) es un filtro que aplica una condición a cada elemento de una lista
# por lo tanto nos pide que hallemos los 'a' con 'href' encontrados en los 'h4' de la class="artifact-title"
# y que a su vez se le agregue en el inicio un 'https://tesis.pucp.edu.pe' y al final un '?show=full'
        return links_Tesis
getTesisPucp(links_pucp)
```

Parte 2

Aquí creamos la función que extraerá los datos requeridos de la tesis cuyo link es ingresado.

```
1: import requests #Librerías
from bs4 import BeautifulSoup

def getDatosDeTesis(linktesis): #Comenzamos la función
    datos = []

    html = requests.get(linktesis).text
    soup = BeautifulSoup(html, 'html.parser')
    #Creamos una tabla donde buscamos los 'tr' de class_='ds-table-row'
    tabla = soup.find_all("tr", class_="ds-table-row")

    #De cada fila en tabla buscamos todos los 'td', los volvemos string y enlistamos en 'columnas'
    for fila in tabla:
        columnas = list(map(lambda x: x.string, fila.find_all("td")))[0:2]

        #La primera columna de 'columnas' en datos
        dato = columnas[0]

        #Si 'dato' coincide con 'datos_requeridos' almacenar en 'datos'
        if(dato in datos_requeridos):
            datos.append(columnas)

    #Nos quedamos con la primera y segunda columna de 'datos', almacenando en varios diccionarios
    datos = [{item[0]:item[1]} for item in datos]
    respuesta = {} #nuevo diccionario

    #Insertamos los diccionarios como elementos del nuevo diccionario
    for dato in datos:
        respuesta.update(dato)

    #Enlistamos los valores del diccionario 'respuesta' y los emparejamos con 'datos_requeridos_2' con zip()
    #Se almacena todo a un nuevo diccionario 'respuesta_datos'
    respuesta_datos = dict(zip(datos_requeridos_2, list(respuesta.values())))

    return respuesta_datos #fin de la función

#Probamos con un link de tesis
getDatosDeTesis('https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/22146?show=full')
```

Parte 3

Procedemos a crear un objeto con los datos de todas las tesis y convertirlo en un data frame en la extensión deseada '.csv'

```
import pandas as pd
#importamos la biblioteca pandas que nos permite estructurar los datos extraídos
todas_las_tesis = []
#creamos una lista vacía
links_tesis_pucp = getTesisPucp(links_pucp)
#asignamos a 'links_tesis_pucp' los links extraídos por la función 'getTesisPucp'
for link in links_tesis_pucp:
    todas_las_tesis.append(getDatosDeTesis(link))
#Aplicamos la función creada anteriormente "getDatosDeTesis(linktesis)" a cada uno de los links
#Los resultados se almacenan a 'todas_las_tesis'
df = pd.DataFrame(todas_las_tesis)
#creamos un data frame
df.to_csv('pucp_todas_las_tesis.csv', index=False, header=True, encoding="utf-8")
#guardamos el data frame en un archivo de extensión .csv
```

Probamos 'df'

```
df
#observamos el data frame
```

Podemos ver el df en la siguiente imagen

	Asesor	Nombre del tesista	Año	Resumen	Institución	Grado
0	Bruzza Moncayo, Mariuxi Alexandra	Llaque Agramonte, Josue Manuel	2022-12-06	Implementación de un sistema integral de gesti...	Pontificia Universidad Católica del Perú	Ingeniero Informático
1	Tupia Anticona, Manuel Francisco	Canasa Mayta, Paul Rodrigo	2022-10-13	Implementación de un sistema de información pa...	Pontificia Universidad Católica del Perú	Ingeniero Informático
2	Flores García, Luis Alberto	Cabello Roca, Percy Ernesto	2022-10-13	Implementación de sistemas de gestión de segur...	Pontificia Universidad Católica del Perú	Ingeniero Informático
3	Huamán Monzón, Fernando Miguel	Recalde Manrique, Diego Andre Walter	2022-10-13	Implementación de un sistema de gestión de act...	Pontificia Universidad Católica del Perú	Ingeniero Informático
4	Murillo Veliz, Braulio Oscar	Alpiste Sarmiento, Daniel Mijail	2022-09-05	Automatización del proceso de gestión de órden...	Pontificia Universidad Católica del Perú	Ingeniero Informático
5	Baldeón Medrano, Johan Paul	Salazar Suasnabar, Hubert Eliot	2022-09-05	Desarrollo de un videojuego educativo configur...	Pontificia Universidad Católica del Perú	Ingeniero Informático
6	Lena Valega, Ángel Gabriel	Arrué Pajares, Sebastián Daniel	2022-08-17	Implementación de un Sistema de Información Ho...	Pontificia Universidad Católica del Perú	Ingeniero Informático
7	Tupia Anticona, Manuel Francisco	Castillo Lopez, Anderson Jesús	2022-08-11	Diseño de un modelo para identificar amenazas ...	Pontificia Universidad Católica del Perú	Ingeniero Informático
8	Rojas Polo, Jonatan Edward	Goyzueta Meneses, Cesar Eduardo	2022-07-22	Propuesta de mejora de la cadena de suministro...	Pontificia Universidad Católica del Perú	Ingeniero Informático
9	Huamán Monzón, Fernando Miguel	Girbau Roque, Christian Diego	2022-07-21	Diseño de un sistema de gestión de protección ...	Pontificia Universidad Católica del Perú	Ingeniero Informático
10	Beltrán Castañón, César Armando	Sosa Pezo, Carlos Alberto	2022-07-11	Determinación y segmentación del nivel socio-e...	Pontificia Universidad Católica del Perú	Ingeniero Informático

Por último le aplicamos un describe para ver las estadísticas

```
df.describe()
#Le aplicamos al df la funcion describe() para poder visualizar la estadística que contiene
```

	Asesor	Nombre del tesista	Año	Resumen	Institución	Grado
count	40	40	40	40	40	40
unique	20	40	36	40	1	2
top	Beltrán Castañón, César Armando	Llaque Agramonte, Josue Manuel	2022-10-13	Implementación de un sistema integral de gesti...	Pontificia Universidad Católica del Perú	Ingeniero Informático
freq	5	1	3	1	40	38