

**Modelo genérico de web scraping que facilite la integración, visualización y normalización de datos inmobiliarios en EE. UU**



**Daniel Felipe Bravo Calvache  
Luis Fernando Gómez Gómez**

Director: PhD. Carlos Alberto Cobos Lozada

**Anexo A - Fuentes de información**

**Universidad del Cauca  
Facultad de Ingeniería Electrónica y Telecomunicaciones  
Departamento de Sistemas  
Grupo de I+D en Tecnologías de la Información (GTI)  
Área de interés en Gestión de la Información  
Popayán, octubre de 2021**

## LISTA DE TABLAS

Tabla 1: Listado de páginas analizadas .....	4
Tabla 2: Atributos de búsqueda de información general .....	6
Tabla 3: Atributos de búsqueda de impuestos .....	6
Tabla 4: Atributos de búsqueda de documentos .....	8
Tabla 5: Campos de información general encontrados .....	12
Tabla 6: Campos de información de impuestos .....	14
Tabla 7: Campos de información de documentos .....	16
Tabla 8: Atributos equivalentes de información general .....	20
Tabla 9: Atributos equivalentes de impuestos .....	21
Tabla 10: Atributos equivalentes de Documentos .....	23

## Fuentes de información

Se presentarán las fuentes de información para las ciudades de Estados Unidos escogidas para analizar, estas páginas contienen toda la información inmobiliaria que se abarcó en el modelo de información. En las tablas mostradas a continuación se evidencian las diferentes URLs que resultan ser las fuentes de información, también se reflejan los datos que cada una de las páginas proveen y la selección de los atributos comunes de todos los mostrados y que serán contenidos en el modelo de información. En primer lugar, se presenta la tabla de URLs para cada uno de los modelos (Información general, impuestos y escrituras), luego, para estos mismos anteriores se muestran los diferentes campos de búsqueda que cada una de las páginas fuentes de información proveen para realizar la búsqueda de una parcela. Después se nombran los diferentes campos de información que cada una de las páginas proveen. Finalmente se muestra la agrupación de los campos comunes de todas las páginas.

### Lista de páginas analizadas

Para lograr un proceso de scraping exitoso se hizo necesario analizar y familiarizarse con diferentes sitios web que proveen información inmobiliaria en de EE. UU. Esto se realizó para diferentes estados del país como se muestra en la Tabla 1.

CIUDAD	INFORMACIÓN GENERAL	IMPUESTOS	ESCRITURAS
Miami Dade, FL	<a href="https://www.miamidade.gov/Apps/PA/propertysearch/#/">https://www.miamidade.gov/Apps/PA/propertysearch/#/</a>	<a href="https://miamidade.com/act_webdev/dallas/searchbyproperty.jsp">https://miamidade.com/act_webdev/dallas/searchbyproperty.jsp</a>	<a href="https://onlineservices.miamidadeclerk.com/officialrecords/StandardSearch.aspx">https://onlineservices.miamidadeclerk.com/officialrecords/StandardSearch.aspx</a>
Dallas, TX	<a href="https://www.dallasact.com/">https://www.dallasact.com/</a>	<a href="https://www.dallasact.com/act_webdev/dallas/searchbyproperty.jsp">https://www.dallasact.com/act_webdev/dallas/searchbyproperty.jsp</a>	<a href="https://dallas.tx.publicsearch.us/">https://dallas.tx.publicsearch.us/</a>
King County, WA	<a href="https://blue.kingcounty.com/Assessor/eRealProperty">https://blue.kingcounty.com/Assessor/eRealProperty</a>	<a href="https://payment.kingcounty.gov/Home/Index?app=PropertyTaxes">https://payment.kingcounty.gov/Home/Index?app=PropertyTaxes</a>	<a href="https://recordsearch.kingcounty.gov/LandmarkWeb/">https://recordsearch.kingcounty.gov/LandmarkWeb/</a>
San Francisco, CA	<a href="https://sfplanninggis.org/pim/">https://sfplanninggis.org/pim/</a>	<a href="https://sanfrancisco-ca.county-taxes.com/public">https://sanfrancisco-ca.county-taxes.com/public</a>	<a href="https://recorder.sfgov.org/#!/simple">https://recorder.sfgov.org/#!/simple</a>
Houston, TX	<a href="https://hcad.org/property-search/real-property/">https://hcad.org/property-search/real-property/</a>	<a href="https://public.hcad.org/records/">https://public.hcad.org/records/</a>	<a href="https://public.hcad.org/records/">https://public.hcad.org/records/</a>

Denver, CO	<a href="https://www.denvergov.org/apps/realpropertyapplication/RealProperty.asp">https://www.denvergov.org/apps/realpropertyapplication/RealProperty.asp</a>	<a href="https://www.denvergov.org/property/">https://www.denvergov.org/property/</a>	<a href="https://countyfusion3.kofiletech.us/countyweb/disclaimer.do">https://countyfusion3.kofiletech.us/countyweb/disclaimer.do</a>
Boston, MA	<a href="https://www.cityofboston.gov/assessing/search/">https://www.cityofboston.gov/assessing/search/</a>	<a href="https://www.invoicecloud.com/portal/">https://www.invoicecloud.com/portal/</a>	<a href="https://pubrec6.hillsclerk.com/ORIPublicAccess/customSearch.html">https://pubrec6.hillsclerk.com/ORIPublicAccess/customSearch.html</a>
Tampa, FL	<a href="https://gis.hcpafl.org/propertysearch/#/nav/Basic%20Search">https://gis.hcpafl.org/propertysearch/#/nav/Basic%20Search</a>	<a href="https://hillsborough.county-taxes.com/public">https://hillsborough.county-taxes.com/public</a>	<a href="https://pubrec6.hillsclerk.com/ORIPublicAccess">https://pubrec6.hillsclerk.com/ORIPublicAccess</a>
Fort Lauderdale, FL	<a href="https://web.bcpa.net/BcpaClient/#/Record-Search">https://web.bcpa.net/BcpaClient/#/Record-Search</a>	<a href="https://broward.county-taxes.com/public">https://broward.county-taxes.com/public</a>	<a href="https://officialrecords.broward.org/AclaimWeb">https://officialrecords.broward.org/AclaimWeb</a>
Jacksonville, FL	<a href="https://paopropertysearch.coj.net/Basic/Search.aspx">https://paopropertysearch.coj.net/Basic/Search.aspx</a>	<a href="https://fl-duval-taxcollector.publicaccessnow.com/PropertyTaxSearch.aspx">https://fl-duval-taxcollector.publicaccessnow.com/PropertyTaxSearch.aspx</a>	<a href="https://or.duvalclerk.com/">https://or.duvalclerk.com/</a>
Palm Beach, FL	<a href="https://www.pbcgov.com/papa/PAPAGIS/papasalesearch2.html">https://www.pbcgov.com/papa/PAPAGIS/papasalesearch2.html</a>	<a href="https://pbctax.manatron.com/Tabs/PropertyTax.aspx">https://pbctax.manatron.com/Tabs/PropertyTax.aspx</a>	<a href="https://www.pbcgov.com/papa/PAPAGIS/papasalesearch2.html">https://www.pbcgov.com/papa/PAPAGIS/papasalesearch2.html</a>

Tabla 1: Listado de páginas analizadas

## Campos de búsqueda

Para lograr el éxito del scraper se necesitó apoyarse en los campos de búsqueda de cada una de las páginas analizadas, por lo cual, el scraper entrará a las páginas y buscará el campo de búsqueda correspondiente para buscar la información de cada parcela. En la Tabla 2, Tabla 3 y Tabla 4 se muestran los atributos de búsqueda que cada página permite ingresar a la hora de realizar la búsqueda de una parcela, las tablas muestran los atributos de búsqueda de las páginas de información general, impuestos y documentos respectivamente.

ATRIBUTOS DE BÚSQUEDA DE INFORMACIÓN GENERAL	
CIUDAD	CAMPOS DE BÚSQUEDA
MIAMI DADE	Owner name Folio Subdivision name Address
DALLAS	Owner name Account number By address Bussines name
TAMPA	Folio/parcel number Owner name Address
FORT LAUDERDALE	Name Address Folio
JACKSONVILLE	Real State number Property Owner Information Property Address / Steet # Street Name
PALM BEACH	Name Search Book Search Document Search Property Search Recorded Date Search Unindexed Property Search Court Case Number Search Certificate Search
DENVER	Search by address Search by parcel number
KING COUNTY	Search by Adreess Search by Parcel Number Search by Property Name
SAN FRANCISCO	Address Parcel

	Planning application number
BOSTON	Search by address Search by parcel

Tabla 2: Atributos de búsqueda de información general

<b>ATRIBUTOS DE BÚSQUEDA DE IMPUESTOS</b>	
<b>CIUDAD</b>	<b>CAMPOS DE BÚSQUEDA</b>
MIAMI DADE	Owner name Address Account number
DALLAS	Owner Address Account Fiduciary
TAMPA	Owner name Address Account number
FORT LAUDERDALE	Owner name Address Account number
JACKSONVILLE	Owner Name Business Name Account Number Location Address
PALM BEACH	Owner name Property control number Mailing Address Property Address
DENVER	Address number Schedule number
KING COUNTY	Tax account Parcel number
SAN FRANCISCO	Bill # Block-lot Property location
BOSTON	Property Address 10 digit Ward + Parcel

Tabla 3: Atributos de búsqueda de impuestos

<b>ATRIBUTOS DE BÚSQUEDA DE DOCUMENTOS</b>	
<b>CIUDAD</b>	<b>CAMPOS DE BÚSQUEDA</b>
MIAMI DADE	Name/document search Cfn number Recording book / page

	Legal description
DALLAS	Grantor/Grantee subdivision doc type doc # Book/volume/page
KING COUNTY	Grantor/Grantee Document Type Book/Page Consideration Recording Number Parcel ID Record Date Legal Advanced Legal
SAN FRANCISCO	Grantor/Grantee Name Document Date Document Number Old Doc Number Block Lot
DENVER	Names Reception Number Book - Page Marriage License / Civil Union Ordinances / Resolutions Grantor Indexes Lastname by range of year Groom Indexes Lastname by range of year Bride Indexes Lastname by range of year Miscellaneous Lastname Indexes
BOSTON	Name Search Book Search Document Search Property Search Recorded Date Search Unindexed Property Search Court Case Number Search Certificate Search
TAMPA	Book/Page Case # Date Document Type DP Name and DOB

	DP Registration # Instrument # Legal Search Name Sales Price Simple Searches
FORT LAUDERDALE	Name Book/Page Document Type Instrument # Case # Consideration Amount Date Recorded Simple Search Parcel ID
JACKSONVILLE	Name Instrument # Doc Type Record Date Consideration Book/Page Case #
PALM BEACH	Party Name Document Type Property Type CFN Book/Page

*Tabla 4: Atributos de búsqueda de documentos*



## Datos inmobiliarios de cada página web

Para conocer los datos que conformarían la información común entre todas las páginas analizadas se hizo necesario listar los diferentes campos informativos de cada sitio web, así lograr agrupar los datos con distintas denominaciones que hacen referencia al mismo concepto, separar los campos poco comunes y a partir de los campos más comunes obtener el modelo de información.

CAMPOS DE INFORMACIÓN GENERAL	
CIUDAD	CAMPOS DE INFORMACIÓN
MIAMI DADE	Folio Sub-divison Property address Owner Mailing address Pa primary zone (building type) Dormitorios Beds / baths / half Floors Living units Actual area Living area Adjust area Lot size Year built
DALLAS	Residential account # Living area Year built Total area Area Living area Depreciation # baths (full/half) # bedrooms State code Address Owner name Mailing address Roof type Construction type Fireplaces Air condition Heating Basement # wet bars

	Building type
KING COUNTY	Parcel Name (owner) Site address Year built Number of baths Condition Lot size Condition Property type Land sqft Road access Zoning Living units Bedrooms Present use Total finished area
SAN FRANCISCO	Parcel Parcel address Year built Building area Parcel area Units Rooms Bathrooms Bedrooms Basement Zoning districts Use type Construction type
DENVER	Property type Owner address Owner name Address Parcel id Year built Building square foot Bedrooms Full baths Half baths Basement Lot size Zoning
BOSTON	Parcel id Address Property type

	Lot size Living area Year built Owner on (date) Owner mailing address
TAMPA	Owner name Mailing address Address Folio Property use Year build Type of build Bedroom Bathrooms Roof structure Roof cover Gross area Heated area Subdivision Zone Total land units
FORT LAUDERDALE	Property id Property owner/s Mailing address Physical address Neighborhood Property use Bldg under air s.f.: Effective year Year built Unit/beds/baths Units Zoning
JACKSONVILLE	Owner name Address Primary site address Re # # buildings Subdivision Total area Year built Roof struct Roofing cover Bedrooms Baths Rooms/units

	Heating type Air condition Gross area Heating area Property use Zoning assessment Effective area
PALM BEACH	Location address Parcel control number Subdivision Owners(sin label) Owner mailing address Total square feet Acres Number of units Year built Bedrooms Full baths Half baths Roof structure Roof cover Property use Zoning

Tabla 5: Campos de información general encontrados

<b>CAMPOS DE INFORMACIÓN DE IMPUESTOS</b>	
<b>CIUDAD</b>	<b>CAMPOS DE INFORMACIÓN</b>
MIAMI DADE	Amount due Owner Assessed value Ad valorem Non-ad valorem Total discountable No discount nava Total tax
DALLAS	Current tax levy Total amount due  Tabla resumen Payment amount Tax year paid Payer Receipt date
KING COUNTY	Paid Tax payer name

	Improvement value Balance Interest Penalty Tabla resumen Date Amount Penalty/interest paid
SAN FRANCISCO	Total Tabla resumen Paid? Total amount Semester Assessee
DENVER	Total Due Paid Interest Liens/fees Original tax levy Exemption Total assessed value Assessed improvements Assessed land
BOSTON	Total due Personal exemptions Fees Interest Assessed owner
TAMPA	Amount due Owner Assessed value Ad valorem Non-ad valorem Total discountable No discount nava Total tax
FORT LAUDERDALE	Amount due Owner Assessed value Ad valorem Non-ad valorem Total discountable No discount nava Total tax
JACKSONVILLE	Year

	Total Taxes Fees Interest Discount Paid Amount due Paid by
PALM BEACH	Paid by Year Bill number Amount paid Total due Tax Discount Penalty/fee Interest

Tabla 6: Campos de información de impuestos

<b>CAMPOS DE INFORMACIÓN DE DOCUMENTOS</b>	
<b>CIUDAD</b>	<b>CAMPOS DE INFORMACIÓN</b>
MIAMI DADE	First party (code) Second party Rec book Rec page Rec date Orig. Cfn no Orig. Book Orig. Page Misc reference Document type Document date Entry date Subdivision name Pages Legal description Block no Folio no Plat book Plat page Section Township Range
DALLAS	Document number Number of pages

	Recorded date Book Volume Page Instrument date Consideration Parties names Legal description Lot Block
KING COUNTY	Instrument # Book Page Book type Doc type Number of pages Doc link Grantor Grantee Legal description
SAN FRANCISCO	Grantor Grantee Description (doc type) Document number Document date Pages
DENVER	Document type Document status Reception number Book Page Recorded date Number of pages Grantor Grantee Associated documents from Documents associated to
BOSTON	Doc. # Rec. Date Rec time Type desc. # of pgs. Book Page Consideration Doc. Status

	Grantor Grantee Grantor/grantee number
TAMPA	Document type Recording date Grantor Grantee Book Page Legal Direct link
FORT LAUDERDALE	Record date Book type Book Page Instrument number Number of pages Doc type
JACKSONVILLE	Record date Book type Book Page Instrument number Number of pages Doc type Grantor Grantee Consideration # of aux pages Doclegals / parcel#
PALM BEACH	Type Date/time Cfn Book type Book Page Pages Consideration Party 1 Party 2 Legal

Tabla 7: Campos de información de documentos



## Atributos equivalentes

Luego de obtener la lista de toda la información disponible en los diferentes sitios web, se agrupa cada campo de información junto con sus sinónimos.

<b>ATRIBUTOS EQUIVALENTES DE INFORMACIÓN GENERAL</b>		
<b>DOMINIO</b>	<b>ATRIBUTO</b>	<b>SINÓNIMOS OBTENIDOS DE TODOS LOS SITIOS WEB</b>
Información general	Subdivision	Sub-division
	Owner	Property owner/s Owners Owner name Name
	Parcel number	Folio Residential account # Property id Parcel control number Re # Parcel id Parcel number Hcad account (hcad variable) Parcel
	Address	Address Location address Physical address Site address Primary site address "" (no está etiquetada, direccion debajo del nombre) Parcel address Property address
	Bedrooms	# bedrooms Unit/beds/baths Room: bedroom Beds / baths / half Bedrooms
	Year built	Aparece y se escribe igual en todas
	Lot size	Parcel area Acres Land area Total land units Total area Units (square foot) Area Land sqft

Building type	Pa primary zone Use type Property type Property use (podria pasarse aqui) State class code State code Type of build
Building area	Total square feet Total area under air Bldg under air s.f. Effective area Total living area Building square foot Total square footage Total finished area Actual area Living area
Heating	Heating Heated area Heating / ac Heating type
Air conditioned area	Air condition Total area under air Heating / ac Area under air
Zoning	Zone Zoning assessment Zoning districts
Floors	
Mailing address	Address Owner mailing Mailing address Owner contact Owner address Owner mailing address
Roof structure	Roof struct
# buildings	Number of units Living units Units
Neighborhood	
Effective area	
Grade	
Route access	
Condition	
Gross area	
Adjust área	

	Depreciation	
	Roof cover	Roofing cover
	Roof type	
	Use type	Land use code
	Construction type	
	Fireplaces	
	Basement	
	Wet bar	
	Rooms	Unit/beds/baths Room: total Rooms/units
	Baths	Beds / baths / half # baths (full/half) Bathrooms Baths Full bath Room: half bath Room: full bath Full baths Number of baths Half baths Unit/beds/baths
Extra features	Feature name	Feature description Accessory type Improvement type
	Description	
	Year built	Year Eff year Year roll
	Floor	
	Exterior wall	
	Construction	
	Quality	
	Total units	Units
	Value	
Area	Sqft	
Información adicional	Comunity development district	
	Existing land use	
	Empowerment zone	
	Urban development	
	School	

	School grade	Schools
	Historic landmark district	
	Foodplain designation	
	Road access	
	Parking	
	Services nearby	
	Port facilities	
	Enterprise zone	

*Tabla 8: Atributos equivalentes de información general*

<b>ATRIBUTOS EQUIVALENTES DE IMPUESTOS</b>		
<b>DOMINIO</b>	<b>ATRIBUTO</b>	<b>SINÓNIMOS OBTENIDOS DE TODOS LOS SITIOS WEB</b>
Taxes	Due	Total due Balance Amount due Total current taxes due Total amount due
	Payer	Paid by Tax payer name Assessee Assessed owner Owner
	Assessed value	Appraised Total assessed value
	Discountable value:	
	No discount nava:	
	Total tax:	Total Paid Amount paid Current tax levy
	Year	Tax year paid
	Taxes before	Original tax levy Amount Tax
	Fees	Liens/fees Penalty Penalty/fee
	Interest	

	Discount	
	Bill number	
	Exemption	Personal exemptions
	Assessed improvements	Improvement value
	Assessed land	
	Paid?	
	Semester	
Ad-valorem	Taxing authority	Distribution information Taxing jurisdiction Ad valorem
	Millage	Millage rate
	Assessed	Assessed value
	Exemption	Exemption amount
	Taxable	Taxable value
	Tax	Gross tax Taxes
	Tax total	Dollars Total tax Total
	Taxing code	
	Credit	
	Savings	
	Percent	Rate Tax rate
Non ad-valorem	Levying authority	Ad valorem
	Rate	
	Amount	
	Total	Total tax
	Levying code	
	Gross tax	
	Credit	
	Net tax	
Savings		

Tabla 9: Atributos equivalentes de impuestos

ATRIBUTOS EQUIVALENTES DE DOCUMENTOS		
DOMINIO	ATRIBUTO	SINÓNIMOS OBTENIDOS DE TODOS LOS SITIOS WEB
Documentos	Grantor	Parties names First Party (Code) Party 1
	Grantee	Parties names

	Party 2 Second Party
Book	Rec Book
Page	Rec Page
Rec Date	Recorded Date Date/Time
Orig. CFN No	
Orig. Book	
Orig. Page	
Misc Reference	
Doc Type	Document Type Type Type Desc. Description
Document Date	Instrument Date
Entry Date	
Subdivision Name	
Legal Description	Lot Legal
Block No	Block
Folio No	
Plat book	
Plat Page	DocLegals / Parcel#
Section	
Township	
Range	
Document Number	Instrument # CFN Doc. # Reception Number Instrument Number
Pages	Number of Pages # of Pgs.
Volume	
Consideration	
Book Type	
Doc Link	Direct link
Document Status	Doc. Status
Associated Documents From	
Documents Associated To	
Rec Time	

	Grantor/Grantee Number	
	# of AUX Pages	

*Tabla 10: Atributos equivalentes de Documentos*