

Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos



Grado en Ingeniería Informática

Trabajo Fin de Grado

DATOS ENLAZADOS sobre BICICLETAS en CALLES de MADRID

Autor: Rubén Rodríguez Álvarez Tutor(a): Oscar Corcho García Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado Grado en Ingeniería Informática

 $\it T\'{i}tulo:$ DATOS ENLAZADOS sobre BICICLETAS en CALLES de MADRID Abril 2020

Autor: Rubén Rodríguez Álvarez Tutor: Oscar Corcho García

Inteligencia Artificial ETSI Informáticos

Universidad Politécnica de Madrid

La publicación de datos abiertos por parte de ayuntamientos e instituciones públicas está cada vez más extendido y en un futuro se irá incrementando. Estos datos son completamente accesibles y reutilizables para cualquier fin que un usuario quiera darles, lo cual da mucha libertad a la hora de crear aplicaciones y dar valor a esa información. Esta información, en su mayoría, está publicada en formato RDF, CSV u hojas de calculo. La publicación por parte de los ayuntamientos suele contener errores, incoherencias u otros problemas que impiden su correcta utilización. Es por ello que para su uso debe hacerse un análisis y determinar una estrategia y modelización de los mismos.

En trabajo consistirá en crear una aplicación que, a partir de los datos proporcionados por el ayuntamiento de Madrid, pueda hacer una valoración de las distintas calles acorde con la seguridad de las mismas para circular en bicicleta. Todo esto siguiendo los principios de Web Semantica y Linked Data que permita realizar un buen modelado de los datos, para su correcto funcionamiento y para una posible reutilización posterior.

Primero será necesario seleccionar los vocabularios con los que se va a trabajar. Para ello se reutilizarán los ya creados en la plataforma vocab.ciudadesabiertas.es y se crearán nuevos a partir del portal de datos del ayuntamiento de Madrid (datos.madrid.es). Se reutilizará el vocabulario de Callejero, definido en [1]. Y se han diseñado tres nuevos vocabularios a partir de los datasets proporcionados por el ayuntamiento de Madrid. El correspondiente a Accidentes con implicación de Bicicletas, accesible en formato CSV y XLSX en la web [3]. El correspondiente a Ciclocarriles, accesible en formato XLS y KML en la dirección [4]. Y el correspondiente a Calles Tranquilas, accesible en formato XLS y KML en la dirección [6].

En este proyecto se va a desarrollar una aplicación que, a partir de estos datos y conociendo la ruta entre 2 puntos dentro de Madrid, se pueda determinar la seguridad de una ruta para ir en bicicleta. Para ello se hará uso de una API externa, de la cual se obtendrá la ruta entre las posiciones dadas por el ususario. Una vez se tengan las calles por las cuales el navegador GPS guiará al ciclista hasta su destino, se comprobará una a una su seguridad. Para esta comprobación se hará uso de los datos mencionados antes. Para ello, se le asignará un identificador único a cada calle, el cual es proporcionado por el Callejero del ayuntamiento [7] y definido como se ha mencionado anteriormente en el portal ciudades Abiertas. Este identificador ha sido asignado a los otros tres datasets, cuyos vocabularios han sido definidos en el contexto de este trabajo, y de esta forma es posible hacer una busqueda rápida y concisa de cada calle por la que transitará la ruta, para así conocer mejor las características de ellas y determinar su seguridad.

Más adelante se detallarán los elementos de cada vocabulario utilizado en la aplicación. En términos generales, en la siguiente figura se muestran los vocabularios definidos para este proyecto y sus enlaces con el Callejero y entre ellos.

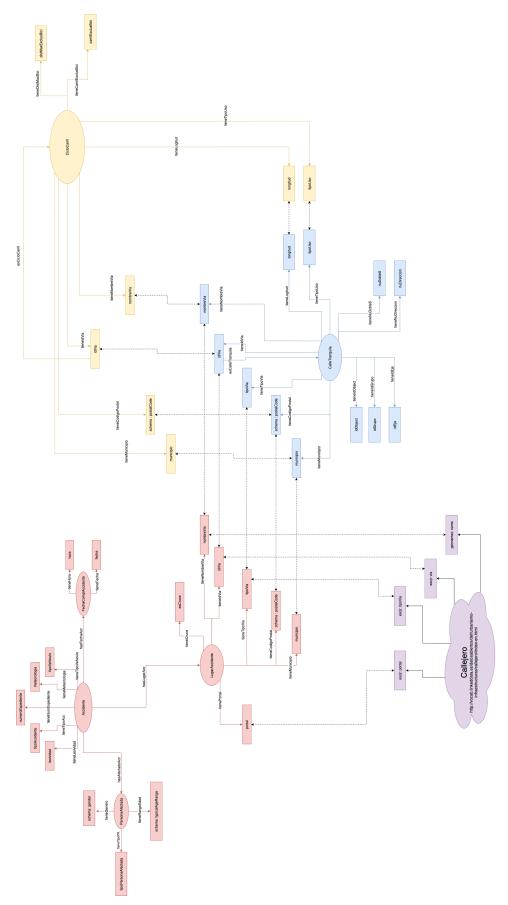


Figura 1: Diagrama de los vocabularios usados en la aplicacion

Tabla de contenidos

1.	Intr	oducci	on	1
2.	Des	arrollo		3
	2.1.	ulario de Accidentes de Bicicletas	4	
		2.1.1.	Clases	6
		2.1.2.	Propiedades de datos	10
		2.1.3.	Propiedades de objeto	12
2.2. Vocabulario de CicloCarriles		Vocab	ulario de CicloCarriles	20
		2.2.1.	Clases	22
		2.2.2.	Propiedades de datos	24
		2.2.3.	Propiedades de objeto	25
2.3.		Vocab	ulario de Calles Tranquilas	27
		2.3.1.	Clases	29
		2.3.2.	Propiedades de datos	32
		2.3.3.	Propiedades de objeto	33
2.4. Transformaciones en los vocabularios: Fase 1				37
2.5. Transformaciones en los vocabularios: Fase 2		Formaciones en los vocabularios: Fase 2	50	
		2.5.1.	Cambios relativos a CicloCarriles	50
		2.5.2.	Cambios relativos a Accidentes de Bicicletas	52
		2.5.3.	Cambios relativos a Calles Tranquilas	54
3.	Res	ultado	s y conclusiones	57
Bi	bliog	rafía		60
Aı	iexo			61

Capítulo 1

Introducción

Capítulo 2

Desarrollo

2.1. Vocabulario de Accidentes de Bicicletas

Para el vocabulario asociado con los accidentes de bicicletas en el ayuntamiento de Madrid se ha elegido la siguiente fuente Datos. Madrid [3]. En ella se muestran los accidentes de tráfico con implicación de bicicletas dentro de la jurisdicción del ayuntamiento.

La organización del conjunto de datos se hará siguiendo el diagrama 2.1

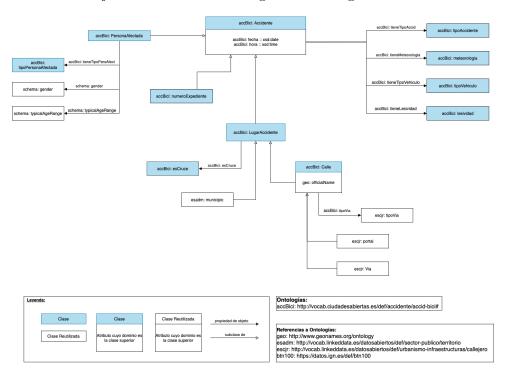


Figura 2.1: Diagrama de Ontología de Accidentes con implicación de bicicletas en Madrid.

Para la representación de los datos de accidentes de trafico con bicicletas se han definido varias clases y propiedades. Se han reutilizado elementos ya definidos en el vocabulario de Callejero [1] y de Schema [2].

En la siguiente tabla se muestran los Namespaces usados.

esadm	http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio#
escjr	http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
schema	http://schema.org#
www-geonames-org	http://www.geonames.org/
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#

Figura 2.2: Namespaces usados para Accidentes de Bicicletas

Se ha optado por mantener la separación de elementos como fecha y hora, calle y numero debido a que la fuente de origen los tiene así dispuestos y en la posterior aplicación final que se va a construir será más conveniente tener esa información por separado, para poder disponer de datos a horas con menos luminosidad o calles completas(sin conocer la posición exacta), por ejemplo.

Para este conjunto de datos se ha optado por añadir, además de los ya proporcionados por la fuente de origen del ayuntamiento, nueva información como la propiedad .esCruce", el municipio, el tipo de via o el identificador de via. Son propiedades inferidas de la información proporcionada que permiten que sea más sencillo su tratamiento y uso, para esta u otras aplicaciones que puedan tener estos datos. EsCruce se obtendrá del nombre de la calle, del cual atendiendo a varios patrones se puede determinar si el accidente ha ocurrido en una intersección de dos o más vías. El Municipio se ha añadido para su posible reutilización posterior utilizando otros datasets de otras localidades, en este caso será siempre "Madrid". El Tipo de Via de nuevo se obtendrá, a partir de los nombres de las calles, siguiendo unos patrones y palabras clave que determinen si es Calle, Avenida, Plaza... lo cual facilitará además su posterior cruce con otros datasets que también utilicen esta propiedad. El Identificador de Via se obtendrá comparando el nombre de la via y su tipo con el Callejero de Madrid, el cual proporcionará este valor único que represente la via.

En este conjunto de datos se ha hecho un cambio importante con respecto al original y que será detallada en la sección Transformaciones en los vocabularios Fase1. Los accidentes que han ocurrido entre un cruce de vias se han separado en tantos registros como vias interfieran. De este modo será mucho más simple la búsqueda de accidentes ocurridos en una via y se podrá hacer una búsqueda más sencilla de ellas. Para no perder ninguna información el nombre del lugar donde ha ocurrido el accidente se conservará igual (es decir, con la descripción del cruce de vias) y se podrá identificar si dos o más registros pertenecen al mismo accidente por el número de expediente, el cual se conserva igual en ambos.

2.1.1. Clases

Accidente

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#Accidente

Siniestro ocurrido con implicación de bicicletas.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

En dominio de:

PersonaAfectada, LugarAccidente, meteorología,

lesividad, tipoVehiculo, tipoAccidente,

numeroExpediente

Tiene Superclases:

fecha, time

PersonaAfectada

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#PersonaAfectada

La persona perjudicada por el accidente de tráfico.

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Subclase de:

Accidente

En dominio de:

tipoPersonaAfectada, gender, typicalAgeRange

LugarAccidente

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#LugarAccidente

El momento en el que ocurrió el siniestro.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Subclase de:

Accidente

En dominio de:

esCruce, municipio, Calle

Calle

<u>IRI:</u> http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#Calle

Representación de una via de una ciudad.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Tiene subclase:

tipoVia, Via, portal

Tiene Superclases:

LugarAccidente nombre oficial

Portal

IRI: http://vocab.linkeddata.es/datosabiertos/def/
urbanismo-infraestructuras/callejero#Portal

Ha sido definido por la plataforma ciudadesabiertas [18]. Subacceso independiente exterior (al aire libre) a una misma construcción. Para una misma construcción, con un mismo número de vía, pueden existir varias entradas que pueden estar numeradas con números o letras. [fuente: Modelo de Direcciones de la Administración General del Estado v.2]

<u>Definida por:</u> http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero

Tiene Superclases: LugarAccidente

Municipio

IRI: http://vocab.linkeddata.es/datosabiertos/def/sector-publico/ territorio#Municipio

Se ha reutilizado la definición de Municipio proporcionada por vocab.linkeddata.es [13] Un Municipio es el ente local definido en el artículo 140 de la Constitución española y la entidad básica de la organización territorial del Estado según el artículo 1 de la Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local. Tiene personalidad jurídica y plena capacidad para el cumplimiento de sus fines. La delimitación territorial de Municipio está recogida del REgistro Central de Cartografía del IGN. Su nombre, que se especifica con la propiedad dct:title, es el proporcionado por el Registro de Entidades Locales del Ministerio de Política Territorial, en

http://www.ine.es/nomen2/index.do

Definida por: http://purl.org/derecho/vocabulario

http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio http://www.ign.es/ign/resources/acercaDe/tablon/ModeloDireccionesAGE

Tiene Superclases:

LugarAccidente

Via

IRI: http://vocab.linkeddata.es/datosabiertos/def/
urbanismo-infraestructuras/callejero#Via

Se ha reutilizado la definición de Municipio proporcionada por vocab.linkeddata.es [14] Vía de comunicación construida para la circulación. En su definición según el modelo de direcciones de la Administración General del Estado, Incluye calles, carreteras de todo tipo, caminos, vías de agua, pantalanes, etc. Asimismo, incluye la pseudovía., es decir todo aquello que complementa o sustituye a la vía. En nuestro caso, este término se utiliza para hacer referencia a las vías urbanas. Representación numérica de la misma.

<u>Definida por:</u> http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero

Tiene Superclases:

Calle

${\bf Desarrollo}$

${\bf numero Expediente}$

IRI: http:

//vocab.ciudadesabiertas.es/def/accidente/accid-bici#numeroExpediente

El identificador del siniestro que proporciona el ayuntamiento.

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Tiene Superclases:

Accidente

2.1.2. Propiedades de datos

fecha

IRI: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#fecha

Fecha en la que ocurre el siniestro.

 $Definida\ por: \verb|http://vocab.ciudadesabiertas.es/def/accidente/accid-bici|$

Dominio:

Accidente

Rango:

date

hora

IRI: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#hora

Hora en la que ocurre el siniestro.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

time

nombre oficial

 $\underline{\mathbf{IRI:}}\;\mathtt{http://www.geonames.org/ontology\#officialName}$

Definido en el callejero de DatosAbiertos [1]. Un nombre en el idioma oficial local.

Definida por: http://www.geonames.org/ontology

Dominio: Calle

typical Age Range

IRI: https://schema.org/typicalAgeRange

Rango de edad en el que se encuentra la persona afectada. Seguirá el siguiente formato definido por Schema.org: 10-12 [17]

Definida por: https://schema.org/typicalAgeRange

Dominio:

PersonaAfectada

Rango:

xsd:string

2.1.3. Propiedades de objeto

tiene Persona A fectada

IRI: http:

//vocab.ciudadesabiertas.es/def/accidente/accid-bici#tienePersonaAfectada

Persona que se asocia a un accidente.

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

PersonaAfectada

tieneLugarAccidente

IRI: http:

//vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneLugarAccidente

Lugar de los acontecimientos asociados a un accidente.

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

LugarAccidente

tieneCalle

 $\underline{\mathbf{IRI:}}\ \mathtt{http://vocab.ciudadesabiertas.es/def/accidente/accid-bici\#tieneCalle}$

Calle asociada al lugar donde ocurrió el accidente.

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

 ${\bf Lugar Accidente}$

Rango:

Calle

$tiene Tipo Vehicu \\ lo$

IRI: http:

//vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneTipoVehiculo

Tipo de vehiculo afectado, p.ej. Bicicleta, Bicicleta EPAC (pedaleo asistido). Se han definido los siguientes elementos:

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoVehiculo/BICICLETA

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoVehiculo/BICICLETA-EPAC

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

concept

tiene Meteorologia

IRI: http:

//vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneMeteorologia

Condiciones ambientales que se dan en el momento del siniestro. Se han definido varios tipos posibles:

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/meteorologia/DEPEJADO

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/meteorologia/LLUVIA-DEBIL

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/meteorologia/LLUVIA-INTENSA

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/meteorologia/NUBLADO

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/meteorologia/GRANIZANDO

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/meteorologia/DESCONOCIDO

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

${\bf tiene Num Expediente}$

IRI: http:

 $// {\tt vocab.ciudadesabiertas.es/def/accidente/accid-bici\#tieneNumExpediente}$

El identificador del siniestro que proporciona el ayuntamiento.

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

numeroExpediente

tieneTipoAcc

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneTipoAcc

Tipo de accidente asociado. Se han definido para ello varios tipos posibles:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/COLISION http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/COLISION-DOBLE

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/COLISION-MULTIPLE

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/ALCANCE http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/CHOQUE-NO-VEHICULO

 $\verb|http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/ATROPELLO-PEATON|$

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/VUELCO http:

 $// vocab. \verb|ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/CAIDA| | http: \\$

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/tipoAcc/OTROS

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

tieneLesividad

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneLesividad

Codigo que indica la gravedad del siniestro para la persona afectada.

Para su uso se han definido los siguientes elementos:

01 Atencion en urgencias sin posterior ingreso. - LEVE:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/01 02 Ingreso inferior o igual a 24 horas - LEVE:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/02 03 Ingreso superior a 24 horas. - GRAVE:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/03 04 Fallecido 24 horas - FALLECIDO:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/04 05 Asistencia sanitaria ambulatoria con posterioridad - LEVE:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/05 06 Asistencia sanitaria inmediata en centro de salud o mutua - LEVE:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/06 07 Asistencia sanitaria solo en el lugar del accidente - LEVE:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/07 14 Sin asistencia sanitaria:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/14 77 Se desconoce:

http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/lesividad/77

Definida por:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

Accidente

Rango:

gender

IRI: https://schema.org/gender

Género de la persona afectada.

Seguirá el formato definido por Schema.org [15] Se utilizarán las siguientes definidas en la clase:

http://schema.org/Male http://schema.org/Female http://schema.org/Mixed

Definida por: https://schema.org/gender

Dominio:

PersonaAfectada

Rango:

genderType [16]

${\bf tiene Tipo Pers A fect}$

IRI: http:

//vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneTipoPersAfect

Persona a la que afecta el accidente. Puede ser Conductor, peaton, testigo o viajero. Se han definido los siguientes elementos:

http://vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/persAfec/CONDUCTOR

http:

 $// vocab. \verb|ciudadesabiertas.es/recurso/accidente/accid-bici/persAfec/PEATON| | http: \\$

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/persAfec/TESTIGO http:

//vocab.ciudadesabiertas.es/recurso/accidente/accid-bici/persAfec/VIAJERO

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio:

PersonaAfectada

Rango:

tipoVia

IRI: http://vocab.linkeddata.es/datosabiertos/def/
urbanismo-infraestructuras/callejero#tipoVia

Se ha reutilizado la definición de tipoVia proporcionada por vocab.linkeddata.es [12]. Tipo de vía, que será representado mediante la clasificación en SKOS de URI http://vocab.linkeddata.es/datosabiertos/kos/urbanismo-infraestructuras/tipo-via. Por ejemplo, estas serán las URIs correspondientes a calles y plazas http://vocab.linkeddata.es/datosabiertos/kos/urbanismo-infraestructuras/tipo-via/CL http://vocab.linkeddata.es/datosabiertos/kos/urbanismo-infraestructuras/tipo-via/PL

<u>Definida por:</u> http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero

Dominio: Calle

Rango: concept

tieneMunicipio

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneMunicipio

Municipio en el que ocurrió un accidente.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio: LugarAccidente

Rango: municipio

tieneIdVia

IRI: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#tieneIdVia

Identificador de calle asociado.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio: Calle

Rango: Via

tienePortal

IRI:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#tienePortal

Numero de la calle donde ha ocurrido el accidente, si procede.

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

Dominio: Calle

Rango: Portal

esCruce

<u>IRI:</u> http://vocab.ciudadesabiertas.es/def/accidente/accid-bici#esCruce

Si el accidente ocurrió en un cruce entre 2 o más vías. Se han definido los siguientes elementos:

http://vocab.ciudadesabiertas.es/def/accidente/accid-bici/esCruce/TRUE http://vocab.ciudadesabiertas.es/def/accidente/accid-bici/esCruce/FALSE

Definida por: http://vocab.ciudadesabiertas.es/def/accidente/accid-bici

<u>Dominio:</u> LugarAccidente

Rango: concept

2.2. Vocabulario de CicloCarriles

Para el vocabulario asociado con los ciclocarriles para ciclistas se han obtenido los datos del portal de datos abiertos del ayuntamiento de Madrid [4], en el cual se muestran las calles que disponen de ciclocarriles y alguna de sus características.

La organización del conjunto de datos se hará siguiendo el diagrama 2.3

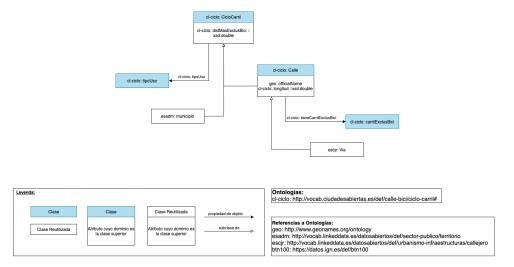


Figura 2.3: Diagrama de Ontología de ciclocarriles en Madrid.

Para la representación de los datos de ciclocarriles para ciclistas se han definido varias clases y propiedades. Se han reutilizado elementos ya definidos en el vocabulario de Callejero de ciudadesabiertas.es [1]. Se han añadido elementos como el identificador de via y el municipio (que siempre será Madrid). El identificador de via se añadirá para cada caso a partir del nombre. Para ello se hará una reducción del nombre de la via a palabras clave, proceso detallado en la sección de Transformaciones de Vocabularios, y se cruzará con el dataset del callejero de Madrid [7]. Se ha optado por omitir la propiedad "MinSimpTol" debido a que no aporta valor al conjunto al tener solo 2 valores, 0 para calles sin carril bici y 0.20 para calles que si disponen de él, información que puede inferirse del campo "MaxSimpTol" (renombrada "distMaxExclusBici"), con valor 0 para el primer caso y valor distinto de 0 para el segundo. Para representar esto se ha añadido la propiedad "carrilExclusBici" con valor booleano indicando si dispone de ese carril exclusivo o no. La fecha proporcionada por el ayuntamiento se ha omitido debido a que no se sabe con exactitud su significado. En caso de que fuese fecha de creación del ciclocarril se debería añadir en futuras actualizaciones del vocabulario, sin embargo al ser en todos los registros la misma cabe la posibilidad de que sea la fecha de inserción en el dataset, lo cual no aporta información relevante y podría inducir a errores. Para este caso el valor de "tipoUso" será siempre CICLOCARRIL. Se han transformado los valores del campo longitud a metros (expresados en kilómetros en los datos de origen) para poder reutilizarlos con más facilidad.

En la siguiente tabla se muestran los Namespaces usados.

	http://vocab.ciudadesabiertas.es/def/ciclo-carril#
esadm	http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio#
escjr	http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
schema	http://schema.org#
www-geonames-org	http://www.geonames.org/
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#

Figura 2.4: Namespaces usados para Ciclocarriles

Cabe destacar en este conjunto de datos la ausencia de la propiedad TipoVia, presente en los demás utilizados. En este caso los nombres de las calles contenían únicamente palabras clave y privaban de la capacidad de obtener este atributo. En cambio si dispone de tipoUso, propiedad que indica si es una calle peatonal o ciclocarril.

Debido a la falta de disponibilidad de una leyenda o información proporcionada por el ayuntameinto de Madrid, para este conjunto de datos no se han podido conocer con exactitud el significado de sus datos. Se ha obtenido de forma aproximada esa información pero podria no ser correcta.

2.2.1. Clases

CicloCarril

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#CicloCarril

Via con uno o más carriles destinados al tránsito de ciclistas. No necesariamente exclusivos para el tránsito de bicicletas, pero si con señalización y limitaciones adaptadas para ello.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Tiene subclase:

Calle, municipio, tipoUso

Tiene Superclase:

distMaxExclusBici

Calle

IRI: http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#Calle

Representación de una via de una ciudad.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Tiene subclase:

carrilExclusBici, Via

Tiene Superclases:

CicloCarril, longitud, nombre oficial

Via

IRI: http://vocab.linkeddata.es/datosabiertos/def/
urbanismo-infraestructuras/callejero#Via

Se ha reutilizado la definición de Municipio proporcionada por vocab.linkeddata.es [14] Vía de comunicación construida para la circulación. En su definición según el modelo de direcciones de la Administración General del Estado, Incluye calles, carreteras de todo tipo, caminos, vías de agua, pantalanes, etc. Asimismo, incluye la pseudovía., es decir todo aquello que complementa o sustituye a la vía. En nuestro caso, este término se utiliza para hacer referencia a las vías urbanas. Representación numérica de la misma.

<u>Definida por:</u> http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero

Tiene Superclases:

Calle

Municipio

IRI: http://vocab.linkeddata.es/datosabiertos/def/sector-publico/ territorio#Municipio

Se ha reutilizado la definición de Municipio proporcionada por vocab.linkeddata.es [13] Un Municipio es el ente local definido en el artículo 140 de la Constitución española y la entidad básica de la organización territorial del Estado según el artículo 1 de la Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local. Tiene personalidad jurídica y plena capacidad para el cumplimiento de sus fines. La delimitación territorial de Municipio está recogida del REgistro Central de Cartografía del IGN. Su nombre, que se especifica con la propiedad det:title, es el proporcionado por el Registro de Entidades Locales del Ministerio de Política Territorial, en

http://www.ine.es/nomen2/index.do

Definida por:

http://purl.org/derecho/vocabulario

http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio http://www.ign.es/ign/resources/acercaDe/tablon/ModeloDireccionesAGE

Tiene Superclases:

CicloCarril

2.2.2. Propiedades de datos

longitud

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#longitud

Longitud de la calle descrita. Esta propiedad está referida a la calle que contiene un ciclocarril (calle completa). La longitud del ciclocarril estará representada por la propiedad distMaxExlusBici.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Dominio: Calle

Rango: double

distMaxExclusBici

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#distMaxExclusBici

Longitud del carril exclusivo de bicicletas dentro de la calle. En caso de que no haya ciclocarril, el valor será 0.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Dominio: CicloCarril

Rango: double

nombre oficial

IRI: http://www.geonames.org/ontology#officialName

Definido en el callejero de DatosAbiertos [1]. Un nombre en el idioma oficial local.

Definida por: http://www.geonames.org/ontology

Dominio: Calle

2.2.3. Propiedades de objeto

tiene Calle

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#tieneCalle

Calle asociada a un CicloCarril.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

<u>Dominio:</u> CicloCarril

Rango: Calle

tipoUso

IRI: http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#tipoUso

Identificador del tipo de uso que puede tener la calle. Se han definido 2 clases para ello:

- http://vocab.ciudadesabiertas.es/kos/urbanismo-infraestructuras/calle/tipo-uso/CICLOCALLE
- $\verb|http://vocab.ciudadesabiertas.es/kos/urbanismo-infraestructuras/calle/tipo-uso/PEATONAL|$

Dominio:

CicloCarril

Rango: concept

tieneIdVia

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#tieneIdVia

Identificador de calle asociado.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Dominio: Calle

Rango: Via

tiene Municipio

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril#tieneMunicipio

Municipio que contiene una calle.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Dominio: CicloCarril

Rango: municipio

tieneCarrilExclusBici

 $\underline{\mathbf{IRI:}} \; \texttt{http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril\#} \\ \mathsf{tieneCarrilExclusBici}$

Indicador de si las bicicletas disponen o no de un carril propio para su circulación. Se han definido 2 clases para ello:

- http://vocab.ciudadesabiertas.es/recurso/calle-bici/ciclo-carril/carrilExlus/EXCLUSIVO-BICI
- http://vocab.ciudadesabiertas.es/recurso/calle-bici/ciclo-carril/carrilExlus/NO-EXLUSIVO-BICI

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/ciclo-carril

Dominio: CicloCarril

Rango: concept

2.3. Vocabulario de Calles Tranquilas

Para el vocabulario asociado con las calles tranquilas para ciclistas en el ayuntamiento de Madrid se ha elegido la fuente de Datos. Madrid [6]. Proporcionada por el ayuntamiento de Madrid, se muestran las calles más apropiadas para el tránsito de ciclistas. No se dan especificaciones de los criterios utlizados que han llevado a estas calles a formar parte de la lista. Sin embargo, si se puede observar en algnas de ellas ciertos patrones, como que no forman parte de las vias principales de la capital y que son poco transitadas. La misma web ofrece un archivo KML y permite que se puedan mostrar sobre un mapa en Open Street Map [8], lo cual proporciona una idea general de su disposición y posibles criterios utilizados.

La organización del conjunto de datos se hará siguiendo el diagrama 2.5.

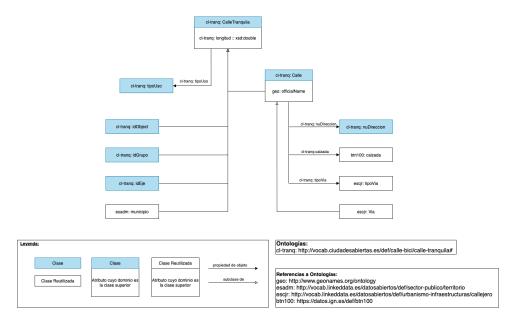


Figura 2.5: Diagrama de Ontología de Calles Tranquilas en Madrid.

Para la representación de los datos de calles tranquilas para ciclistas se han definido varias clases y propiedades. Se han tomado como ejemplo el vocabulario definido en ciudades abiertas de Callejero [1] y el definido en datos.ign sobre calzadas [9].

Se han realizados ciertas modificaciones con respecto al dataset original para que puedan utilizarse sus datos más eficazmente. Se ha optado por la separación del tipo de via del nombre, conservandola en éste y creando una nueva propiedad que permita saber su clase. Algunos ejemplos serían Calle, Avenida, Plaza... Se ha añadido el identificador de la via, obtenido a partir del nombre y cruzado con el dataset del callejero de madrid [7]. El identificador permitirá hacer búsquedas mucho más rápidas sobre los datos en caso de querer hacerla filtrando por la calle, que es el caso de la aplicación final que se desea realizar para este proyecto. En este caso el municipio será siempre madrid, pero en caso de que se quisiera reulizar en otros proyectos a mayor escala sería necesario conocer la zona geográfica donde se encuentra, por tanto también se ha añadido, aunque con el valor fijo de Madrid, que corresponde al código 28079, proporcionado por el Instituto Nacional de Estadistica[10].

Estas transformaciones se detallarán más adelante en la seccion: Transformaciones en los vocabularios.

Se ha optado por omitir la propiedad ID_TIPO, ya que representa lo mismo que TX_CAPA

(el uso que tiene la via) y se ha optado por la segunda por ser representado con texto, más visual y representativo a la hora de su utilización.

El tipo de uso se ha definido en http://vocab.ciudadesabiertas.es/def/calle-bici#tipoUso debido a que es común para todas las calle-bici (tanto calles tranquilas como ciclocarriles), por lo tanto se ha optado porque los valores sean definidos en una categoria que englobe a las dos. La dirección se ha definido también en esta categoría común ya que podría ser añadida a ciclocarriles por el ayuntamiento o ser común con otros dataset de una categoría similar(http://vocab.ciudadesabiertas.es/def/calle-bici#nuDireccion).

En la siguiente tabla se muestran los Namespaces usados.

	http://vocab.ciudadesabiertas.es/def/calles-tranquilas#
esadm	http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio#
escjr	http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
schema	http://schema.org#
www-geonames-org	http://www.geonames.org/
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#

Figura 2.6: Namespaces usados para Calles Tranquilas

Debido a la falta de disponibilidad de una leyenda o información proporcionada por el ayuntameinto de Madrid, para este conjunto de datos no se han podido conocer con exactitud el significado de sus datos. Se ha obtenido de forma aproximada esa información pero podria no ser correcta.

2.3.1. Clases

CalleTranquila

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#CalleTranquila

Calle apropiada para el tránsito de ciclistas siguiendo los cirterios del ayuntamiento de Madrid.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Tiene subclase:

 $id Object, \qquad id Grupo, \qquad id Eje, \qquad tipo Uso,$

municipio, Calle **Tiene Superclase:**

longitud

Calle

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#Calle

Representación de una via de una ciudad.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Tiene subclase:

longitud, nuDireccion, calzada, tipoVia

Tiene Superclases:

CalleTranquila

nombre oficial

Via

IRI: http://vocab.linkeddata.es/datosabiertos/def/
urbanismo-infraestructuras/callejero#Via

Se ha reutilizado la definición de Municipio proporcionada por vocab.linkeddata.es [14] Vía de comunicación construida para la circulación. En su definición según el modelo de direcciones de la Administración General del Estado, Incluye calles, carreteras de todo tipo, caminos, vías de agua, pantalanes, etc. Asimismo, incluye la pseudovía., es decir todo aquello que complementa o sustituye a la vía. En nuestro caso, este término se utiliza para hacer referencia a las vías urbanas. Representación numérica de la misma.

<u>Definida por:</u> http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero

Tiene Superclases:

Calle

Municipio

IRI: http://vocab.linkeddata.es/datosabiertos/def/sector-publico/ territorio#Municipio

Se ha reutilizado la definición de Municipio proporcionada por vocab.linkeddata.es [13] Un Municipio es el ente local definido en el artículo 140 de la Constitución española y la entidad básica de la organización territorial del Estado según el artículo 1 de la Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local. Tiene personalidad jurídica y plena capacidad para el cumplimiento de sus fines. La delimitación territorial de Municipio está recogida del REgistro Central de Cartografía del IGN. Su nombre, que se especifica con la propiedad det:title, es el proporcionado por el Registro de Entidades Locales del Ministerio de Política Territorial, en

http://www.ine.es/nomen2/index.do

Definida por: http://purl.org/derecho/vocabulario

http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio http://www.ign.es/ign/resources/acercaDe/tablon/ModeloDireccionesAGE

Tiene Superclases:

CalleTranquila

idObject

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#idObject

Identificador propio del ayuntamiento.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Tiene Superclases:

CalleTranquila

idGrupo

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#idGrupo

Identificador propio del ayuntamiento.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Tiene Superclases:

CalleTranquila

idEje

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#idEje

Identificador propio del ayuntamiento.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Tiene Superclases:

CalleTranquila

2.3.2. Propiedades de datos

nombre oficial

IRI: http://www.geonames.org/ontology#officialName

Definido en el callejero de DatosAbiertos [1]. Un nombre en el idioma oficial local.

Definida por: http://www.geonames.org/ontology

Dominio: Calle

longitud

IRI:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#longitud

Longitud de la calle o tramo de la calle descrito. Su medida es Shape_leng. Es propiedad de CalleTranquila y no de calle porque cabe la posibilidad de que no coincidan ambas medidas debido a tramos de vias más inseguros. A falta de documentación sobre el dataset se ha supuesto esta hipótesis.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: CalleTranquila

Rango: double

2.3.3. Propiedades de objeto

tiene Calle

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tieneCalle

Calle asociada a una Calle Tranquila.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: CalleTranquila

Rango: Calle

tipoUso

TRT.

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tipoUso

Identificador del tipo de uso que puede tener la calle. Se han definido 2 clases para ello:

- $\verb|http://vocab.ciudadesabiertas.es/kos/urbanismo-infraestructuras/calle/tipo-uso/CICLOCALLE|$
- http://vocab.ciudadesabiertas.es/kos/urbanismo-infraestructuras/calle/tipo-uso/PEATONAL

Dominio:

 $\operatorname*{CalleTranquila}_{-}$

Rango: concept

tieneIdEje

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tieneIdEje

Identificador de Eje asociado a una calle.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: CalleTranquila

Rango: idEje

tieneIdGrupo

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tieneIdGrupo

Identificador de Grupo asociado a una calle.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: CalleTranquila

Rango: idGrupo

tieneIdObject

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tieneIdObject

Identificador de Objeto asociado a una calle.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: CalleTranquila

Rango: idObject

${\bf tiene IdVia}$

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tieneIdVia

Identificador de calle asociado.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: Calle

Rango: Via

tiene Municipio

IRI: http:

//vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila#tieneMunicipio

Municipio que contiene una calle.

Definida por:

http://vocab.ciudadesabiertas.es/def/calle-bici/calle-tranquila

Dominio: CalleTranquila

Rango: municipio

${\bf nuDireccion}$

IRI: http://vocab.ciudadesabiertas.es/def/calle-bici#nuDireccion

Dirección y sentido de la calle o tramo de la calle a la que se refiere. Puede tomar los siguientes valores:

http://vocab.ciudadesabiertas.es/recurso/calle-bici#SOL - Dirección y sentido a Sol

http://vocab.ciudadesabiertas.es/recurso/calle-bici#paraleloSOL - Dirección paralela a Sol

http://vocab.ciudadesabiertas.es/recurso/calle-bici#contrarioSOL - Dirección a sol y sentido opuesto.

Esta definición esta basada en datos aproximados proporcionados por mapas y otros elementos que han podido verse en los ejemplos del dataset. Pueden ser erróneos o no reflejar con exactitud la realidad del dato ya que no existe una leyenda que poder consultar.

Definida por: http://vocab.ciudadesabiertas.es/def/calle-bici

Dominio: Calle

Rango

concept

calzada

IRI: https://datos.ign.es/def/btn100#calzada

Doble sentido o sentido único de una calle. Puede tomar los siguientes valores definidos en datos.ign.es [11]:

https://datos.ign.es/kos/transportes/calzada/convencional

https://datos.ign.es/kos/transportes/calzada/doble

https://datos.ign.es/kos/transportes/calzada/sentido-unico

Definida por: https://datos.ign.es/kos/transportes/calzada/

Dominio: Calle

 $\frac{\mathbf{Rango:}}{\mathbf{concept}}$

tipoVia

IRI: http://vocab.linkeddata.es/datosabiertos/def/
urbanismo-infraestructuras/callejero#tipoVia

Se ha reutilizado la definición de tipoVia proporcionada por vocab.linkeddata.es [12]. Tipo de vía, que será representado mediante la clasificación en SKOS de URI http://vocab.linkeddata.es/datosabiertos/kos/urbanismo-infraestructuras/tipo-via. Por ejemplo, estas serán las URIs correspondientes a calles y plazas http://vocab.linkeddata.es/datosabiertos/kos/urbanismo-infraestructuras/tipo-via/CL http://vocab.linkeddata.es/datosabiertos/kos/urbanismo-infraestructuras/tipo-via/PL

 $\underline{\mathbf{Definida\ por:}}\ \texttt{http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero}$

Dominio: Calle

Rango: concept

2.4. Transformaciones en los vocabularios: Fase 1

Partiendo de los datos proporcionados por el ayuntamiento de Madrid y con el fin de plasmar las estructuras antes definidas, se han realizados ciertos cambios con respecto al dataset original. Campos añadidos o modificaciones o transformaciones en los ya existentes son algunos de los motivos para realizarlos.

Con el fin de repetir el proceso en datasets iguales referentes a otros años o posibles similitudes con otros, a continuación se detallan los pasos que se han seguido en este caso para poder replicarlos:

- Transformar a codificación ISO-8859-1.
- Modificar palabras básicas como por ejemplo C/ por calle y demás casuística. Para ello se ha definido el siguiente código para aplicarlo a todos los registros del dataset por igual.

Listing 2.1: Función cambiosBasicos

```
import os
  3 from re import findall
     def cambiosBasicos(nombreCarpeta="", nombreSinCsv="", nRowNombre=-1, nFileFin = "-1"):
    with open(nombreCarpeta + "/" + nombreSinCsv + ".csv") as csvfile:
                      csvreader = csv.reader(csvfile, delimiter=";")
file = open(nombreCarpeta + "/" + nombreSinCsv + nFileFin + ".csv", "w")
                      primeraLinea = True
for row in csvreader:
 10
 \frac{11}{12}
                             if (primeraLinea):
                                     fila = ";".join(row)
primeraLinea = False
13
14
15
                                     16
17
18
19
20
21
                                     if(nombreVia.__contains__("S/N")): # P.ej. SAN FRANCISCO S/N, casos que generan error
    nombreVia = nombreVia.replace('S/N', "")
if (findall('([A-z, C!, Å', 0-9]{1})/[A-z]{1}', nombreVia) != []): # Escritura estandar de
23
                                     "Cruce"
nombreVia = nombreVia.replace("/", " / ")

# C/ MONASTERIO DE ARLANZA-AV. SANTUARIO DE VALVERDE

if (findall('(?\\d)[\-{1}](?\\d)', nombreVia) != []): # Evitar palabras con numeros como M-30

arr1 = nombreVia.replace("-", " - ").split()

if (getTipoVia(palablasMalEscritas(arr1[arr1.index("-")+1])) != ""): # Si es palabra clave

indicadora de nueva via, es Cruce

nombreVia = nombreVia.replace("-", " / ")

if (findall('(?!\d)[\-{1}](\d)', nombreVia) != []): # Palabras con numero como M-30

nombreVia = nombreVia.replace("-", "")
24
26
27
28
29
30
31
32
33
                                         ;19:10; "CAMINO EN ZONA DENOMIDADA COMO ""ESPACIO MÉXICO""";
34
35
                                     if (nombreVia.__contains__('"')): #Quitar comillas
   nombreVia = nombreVia.replace('"', '')
36
37
                                     if (nombreVia.__contains__("PASO ELEVADO")):
    nombreVia = nombreVia.replace('PASO ELEVADO', 'PASO_ELEVADO')
38
39
                                      elif (nombreVia.__contains__("SENDA CICLABLE")):
    nombreVia = nombreVia.replace('SENDA CICLABLE', 'SENDA_CICLABLE')
\frac{40}{41}
42
                                     nombreVia = palablasMalEscritas(nombreVia)
 43
                                      row[nRowNombre] = nombreVia
44
                                                          .join(row)
                             file.write(fila + os.linesep)
```

La finalidad de la función definida en Listing 2.1 es modificar ciertas palabras para que posteriormente se puedan entender mejor e inferir información a partir de ellas sin llegar a errores.

Es el caso por ejemplo de los cruces. Se pueden escribir de múltiples formas, pero en el tratamiento que vamos a realizar será del modo çalle1 / calle2". Para conseguir ese formato han de modificarse otros nombres como por ejemplo ÇRUCE calle1 CON calle1"para que posteriormente las funciones sean aplicables a estos casos.

También se eliminan elementos como "S/N" (Sur/Norte) que no tienen excesiva relevancia, no forman parte del nombre, pero en cambio si pueden llegar a producir errores graves.

Otras transformaciones serian en palabras que consideramos clave, por ejemplo Paso elevado o Senda Ciclable, a las cuales se les considerará tipo de via, que sean formadas como una única palabra, facilitando asi su posterior búsqueda y que no se cometa el error de incluirlas en las palabras clave del nombre de la via.

También, de nuevo debido a la importancia que tienen guiones o barras en la detección de elementos inusuales o cruces, para carreteras como M-30, M-40 o elementos como KM-0 se les eliminará el guión, considerandolos de esa forma una única palabra.

Por último, se llama a la función palabrasMalEscritas, definida en Listing 2.2, la cual es en parte una continuación de esta aunque para casos más concretos.

Listing 2.2: Función palabrasMalEscritas

```
import os
  3 from re import findall
  5 \ \det \ palablas \texttt{MalEscritas} (\texttt{nombreCalle = ""}): \quad \textit{\#Arreglar errores de codificación y abreviaturas}
              if(nombreCalle ==
                     return ""
  8
              nuevaPalabra = ""
             nuevaPalabra = ""
for palabra in nombreCalle.split():
   palabra = palabra.replace(" ", "").upper()
   if(palabra == "PNTE."):
        palabra = "PUBTE"
   elif (palabra == "CÑADA." or palabra == "CÑADA"):
        palabra = "CÑADA." or palabra == "AVDA." or palabra == "AVDA."
   elif (palabra == "AVDA." or palabra == "AVDA." or palabra == "AV." or palabra == "AVDA."
   elif (palabra == "JARDÍN" or palabra == "JARDÍNES"):
        palabra = "JARDÍN"
 11
 13
14
15
 16
17
                     palabra = "JARDIN" or palabra == "JARDINES
palabra = "JARDIN"
elif (palabra == "CUSTA." or palabra == "CUSTA"):
palabra = "CUESTA"
18
19
20
21
                     palabra = "CUESTA"
elif (palabra == "POLÍGONO"):
    palabra = "POLIGONO"
elif (palabra == "GALERÍA"):
    palabra = "GALERÍA"
22
23
24
                     palabra = "PLAZA."):
palabra = "PLAZA"

palabra = "PLAZA"

elif (palabra == "PISTA."):
25
26
27
28
29
                      palabra = "PISTA"
elif (palabra == "CMNO." or palabra == "CMNO"):
                      palabra = "CAMINO"
elif (palabra == "RONDA"
30
31
                                                        "RONDA."):
32
33
                      elif (palabra ==
                                                        "GTA." or palabra == "GTA"):
34
35
                      palabra = "GLORIETA"
elif (palabra == "CUSTA.
                                                         "CUSTA." or palabra == "CUSTA"):
                              palabra = "CUESTA"
36
37
                     palabra == "PQUE"

palabra = "PARQUE"

elif (palabra
                      elif (palabra ==
                                                        "PQUE." or palabra == "PQUE"):
38
                              (palabra == "CTRA."
palabra = "CARRETERA"
39
                                                        "CTRA." or palabra == "CTRA"):
40
                      elif (palabra = "AUTOV." or palabra == "AUTOV"):
palabra = "AUTOVIA"
elif (palabra == "CRUCE."):
41
42
                              palabra = "CRUCE"
\frac{44}{45}
                     paiarra = "CRUCE"
elif (palabra == "ANILLO.")
palabra = "ANILLO"
elif (palabra == "PASEO."):
palabra = "PASEO"
46
47
48
                     palabra = "PASEU"
elif(palabra == "TRVA."):
    palabra = "TRAVESIA"
elif(palabra == "ESTFE."):
    palabra = "ESTACION_FERROCARRIL"
49
50
51
52
                     elif (palabra == "P?"):
    palabra = "PLAZA"
elif (palabra == "PZA."):
    palabra = "PLAZA"
53
54
55
56
57
                                 (palabra == "CÃŹ" or palabra == "CALL." or palabra == "CALL" or palabra == "C/" or palabra ==
                     palabra = "CALLE"
elif (palabro
58
59
                               (palabra == "GRAL"):
60
61
                      palabra = "GENERAL"
elif (palabra == "STA." or palabra == "STA"):
62
                              palabra = "SANTA"
63
                          Mayoritariamente errores de codificación:
                     elif(palabra._contains__("Ý"):
    palabra = palabra.replace("Ý", "Í")
elif (palabra._contains__("Åa")):
```

En este caso se vuelven a transformar palabras de forma que sea más clara y fácil su utilización. En primer lugar se eliminan abreviaturas y se sustituyen por las palabras completas.

En segundo lugar se corrigen errores de codificación. Aunque esto a priori no debería ocurrir, el dataset de ciclocarriles no fue posible descargarlo y tratarlo con la correcta, por tanto muchos de sus elementos estaban corruptos. Aunque se tuvo que hacer algún cambio manual si fue posible determinar los elementos más comunes que habían sido modificados y de esta forma es posible hacer la gran mayoria de forma automática, y por si volviera a ocurrir con otro dataset.

• Cambios realizados de forma manual al conjunto de datos de ciclocarriles:

Debido a los errores de codificación en sus registros, además de los cambios ya expuestos anteriormente en Listing 2.2, se han tenido que realizar manualmente. Son los siguientes:

```
- ln 52: M/ndez ?lvaro->Méndez Álvaro
```

- ln 64-65: Men/ndez Pelayo ->Menéndez Pelayo
- ln 72-73: Ortega y Gasset -> Jose Ortega y Gasset (Igual al nombre del Callejero de Madrid)
- ln 103: Donoso Cort/s ->Donoso Cortés
- ln 112: Gral Moscardæ ->Gral Moscardó
- ln 114: Camilo Jos/Cela Azcona -> Camilo José Cela (Tambéin elminado Azcona ya que no esta previsto la existencia de cruces)
- ln 117: Gral Yag>e ->Gral Yagüe
- ln 125: MARQUÉS DE VIANA SOR ANGELA DE LA CRUZ ->Divido en 2 registros con caracteristicas similares.

Cambios realizados de forma manual al dataset de CallesTranquilas:

l
n 1292-1292: Calle de la Cooperativa El Úctrica –>Calle de la Cooperativa El
éctrica

l
n 1822-1823: Doctor Mart Ýn Ar Úvalo –>Doctor Martín Arévalo

- l
n 1919-1921: Errores en el formato del c
sv ${\bf o}$ de codificación. Mismo registro en varias lineas.
- l
n 2040: ENLACE CALLE AMERICIO CON MADRID RÍO ->CALLE AMERICIO

Para la realización de estos cambios se ha observado el mapa proporcionado por el ayuntamiento https://datos.madrid.es/egob/new/detalle/auxiliar/mapa.gata-detalle-tranquilas.kml y se ha determinado la mejor forma de representar los datos.

Vias añadidas al dataset del Callejero:

201600;CALLE;DEL;COMANDANTE ZORITA;AVIADOR ZORITA;6;1;59;2;50 -> Igual que el registro . Aviador Zorita" 334200;CALLE;DE;GENERAL YAGÜE;6;1;57;4;76 -> Igual que el registro "San German". Cambio de nombre de la via posterior a la realizacion de varios dataserts https://es.wikipedia.org/wiki/Calle_de_San_Germán.

331600;CALLE;DE;GENERAL MOSCARDO;GENERAL MOSCARDÓ;6;1;39;2;34 -> Igual que el registro .Edgar Neville". Cambio de nombre de la via posterior a la realizacion de varios dataserts https://www.elmundo.es/madrid/2017/05/31/592dbf00e2704ed5058b4688.html. 765800;RONDA;DE;RONDA VALENCIA;RONDA VALENCIA;1;;;2;18 -> Se considera nombre completo Ronda de Valencia", y no separado como muestra inicialmente

Estos cambios se realizan directamente en el dataset del callejero ya que pueden ser aplicables a todos los datasets. Elementos que se consideran básicos en casos concretos, calles nuevas o nuevos nombres (como es el caso de algunos referidos a personajes militares o políticos) cambiados en los ultimos años, deben añadirse por si no han sido actualizados en algunos casos, conservando ambos.

Para ello se ha seguido la lista proporcionada por El Pais en https://elpais.com/ccaa/2017/04/28/madrid/1493369660_675682.html y se han añadido tanto los cambios ya realizados, como los aprobados aun no actualizados en el dataset, para que estén ambos nombres.

96200; CALLE; BATALLA DE BELCHITE; BATALLA DE BELCHITE; 2;1;15;2;22

917;PASEO;DEL;DOCTOR VALLEJO-NAJERA;DOCTOR VALLEJO-NÁJERA;2;1;61;2;56

356700:PLAZA:DE LOS:HERMANOS FALCO Y ALVAREZ:HERMANOS FALCÓ Y ÁLVAREZ:21:1:25:2:24

526000; PASEO; DE; MUÑOZ GRANDES; MUÑOZ GRANDES; 11; 1; 53; 2; 64

329900;CALLE;DEL;GARCIA DE LA HERRANZ;GARCÍA DE LA HERRANZ;11;1;19;2;10

329700; CALLE; DEL; GENERAL FRANCO; GENERAL FRANCO; 11;1;15;2;12

73600;PLAZA;;ARRIBA ESPAÑA;ARRIBA ESPAÑA;5;1;13;2;12

123600; CALLE; CAIDOS DE LA DIVISION AZUL; CAÍDOS DE LA DIVISIÓN AZUL; 5;1;15;2;28

82000;PLAZA;;AUNOS;AUNÓS;5;1;11;2;10

328950;CALLE;DE LA;POETA ANGELA FIGUERA;POETA ÁNGELA FIGUERA;7;1;41;2;22

329400;CALLE;DE;GENERAL DAVILA;GENERAL DÁVILA;7;1;15;2;12

419300;CALLE;DE;JUAN VIGON;JUAN VIGÓN;7;1;25;2;10

332950;CALLE;DEL;GENERAL RODRIGO;GENERAL RODRIGO;7;1;17;2;12

417850;PLAZA;;JUAN PUJOL;JUAN PUJOL;1;1;1;;

 $402600; \mathrm{CALLE}; \mathrm{DE}; \mathrm{JOSE}$ LUIS DE ARRESE; JOSÉ LUIS DE ARRESE; 15; 1; 91; 2; 66

 $48900; CALLE; DEL; ANGEL \ DEL \ ALCAZAR; \'ANGEL \ DEL \ ALC\'AZAR; 15; 1; 7; 2; 8$

330300;CALLE;DEL;GENERAL KIRKPATRICK;GENERAL KIRKPATRICK;15;1;37;2;46

158300; PLAZA; DEL; CAUDILLO; CAUDILLO; 8;1;5;2;4

609700; CALLE;; PRIMERO DE OCTUBRE; PRIMERO DE OCTUBRE; 8; 1; 15; 2; 20

772400; PLAZA;
DEL; VEINTIOCHO DE MARZO; VEINTIOCHO DE MARZO; 8;1;11;2;10
 $\,$

 $137100; CALLE; DEL; CAPITAN\ CORTES; CAPITÁN\ CORTÉS; 16; 1; 13; 2; 14$

31000067;AVENIDA;DEL;ALCALDE CONDE MAYALDE;ALCALDE CONDE MAYALDE;8;;;

 $28150; CALLE; DEL; ALGABE\~NO; ALGABE\~NO; 16; 1; 125; 2; 192$

 $329500; AVENIDA; DEL; GENERAL\ FANJUL; GENERAL\ FANJUL; 10; 1; 185; 2; 144$

331250;CALLE;DEL;GENERAL MILLAN ASTRAY;GENERAL MILLÁN ASTRAY;10;1;81;2;72

 $333250; CALLE; DEL; GENERAL\ SALIQUET; GENERAL\ SALIQUET; 10; 1; 109; 2; 54$

325200;CALLE;DE;GARCIA MORATO;GARCÍA MORATO;10;5;9;22;26

329850;CALLE;DEL;GENERAL GARCIA ESCAMEZ;GENERAL GARCÍA ESCÁMEZ;10;3;27;2;52

 $333000; CALLE; DEL; GENERAL\ ROMERO\ BASART; GENERAL\ ROMERO\ BASART; 10; 1; 149; 2; 90$

67700; AVENIDA; DEL; ARCO DE LA VICTORIA; ARCO DE LA VICTORIA; 9;1;3;2;4

333200; PASEO;
DEL; GENERAL SAGARDÍA RAMOS; 9;1;7;2;24

31004081;GLORIETA;DE;RAMON GAYA;RAMÓN GAYA;9;;;;

 $144900; CALLE; DE; CARLOS\ RUIZ; CARLOS\ RUIZ; 9; 1; 3; 2; 10$

263650;PLAZA;DE;EMILIO JIMENEZ MILLAS;EMILIO JIMÉNEZ MILLAS;9;1;1;2;4

```
1887:CALLE:DEL:PHERTO DE LOS LEONES:PHERTO DE LOS LEONES:9:1:61:2:92
360800:CALLE:DE LOS:HEROES DEL ALCAZAR:HÉROES DEL ALCAZAR:13:::2:12
166500;CALLE;DEL;CERRO DE GARABITAS;CERRO DE GARABITAS;13;1;17;2;12
220600; CALLE; DEL; CRUCERO BALEARES; CRUCERO BALEARES; 13;1;25;2;16
338200;PLAZA;DEL;GOBERNADOR CARLOS RUIZ;GOBERNADOR CARLOS RUIZ;13;1;7;2;8
256300; CALLE; DE; EDUARDO AUNOS; EDUARDO AUNÓS; 4; 1; 41; 2; 56
331500;PASAJE;DEL;GENERAL MOLA;GENERAL MOLA;4;1;9;2;6
357000:CALLE:DE LOS:HERMANOS GARCIA NOBLEJAS:HERMANOS GARCÍA NOBLEJAS:15:::2:198
331800;CALLE;DEL;GENERAL ORGAZ;GENERAL ORGAZ;6;1;31;2;18
333900; CALLE; DEL; GENERAL VARELA; GENERAL VARELA; 6;1;37;2;38
328800;CALLE;DEL;GENERAL ARANDA;GENERAL ARANDA;6;1;55;2;98
328900;ESCALINATA;DEL;GENERAL ARANDA;GENERAL ARANDA;6;;;;
466800; CALLE; DE; MANUEL SARRION; MANUEL SARRIÓN; 6; 1; 13; 2; 12
137400; CALLE; ; CAPITAN HAYA; CAPITAN HAYA; 6; 1; 65; 2; 66
293200:PLAZA:DE:FERNANDEZ LADREDA:FERNÁNDEZ LADREDA:11:3:5::
293200;PLAZA;DE;FERNANDEZ LADREDA;FERNÁNDEZ LADREDA;12;1;1;2;2
```

 Añadir la propiedad esCruce únicamente al dataset de Accidentes.Inferida a partir del nombre de la via aplicando el siguiente código:

Listing 2.3: Función annadirEsCruceAccidentes

```
1 import csv
3 from re import findall
4 def annadirEsCruceAccidentes(nombreCarpeta = "", nombreSinCsv = "", nRowNombre = -1, nFileIni = "-1", nFileFin
       with open(nombreCarpeta + "/" + nombreSinCsv + nFileIni + ".csv") as csvfile:
           6
7
8
9
           for row in csvreader: if (primeraLinea):
10
                   fila = ";".join(row)
fila = fila + ";" + "ES_CRUCE"
primeraLinea = False
11
13
14
15
                   fila = ";".join(row)
16
17
                   if (row[nRowNombre].__contains__(" / ")):
    fila = fila + ";" + "1"
                   else:
fila = fila + ";" + "0"
19
               file.write(fila + os.linesep)
           file.close()
```

En el caso de que el accidente haya ocurrido en la intersección entre varias vias, se añaden al nombre siguiendo ciertos patrones (antes englobados en uno: calle1 / calle2) y en esta propiedad se detalla si se ha detectado o no que sea cruce.

Únicamente se realiza para el dataset de accidentes ya que los demás no dispondrán de esta propiedad.

 Añadir la propiedad Tipo de Via a los conjuntos de datos. Inferida a partir del nombre de la via aplicando el siguiente código:

Listing 2.4: Función annadirTipoVia

```
8  file = open(nombreCarpeta + "/" + nombreSinCsv + nFileFin + ".csv", "w")
9  primeraLinea = True
10  for row in csvreader:
11   if (primeraLinea):
12   fila = ";".join(row)
13   fila = fila + ";" + "TIPO_VIA"
14   primeraLinea = False
15   else:
16   tipoVia = getTipoVia(row[nRowNombre])
17   fila = ";".join(row)
18   fila = fila + ";" + tipoVia
19  file.write(fila + os.linesep)
20  file.close()
```

En el código mostrado en Listing 2.4 se observa la función annadirTipoVia. Aplcada a los datasets de CallesTranquilas y Accidentes, pero no a Ciclocarriles debido a la naturaleza de los nombres de sus calles, los cuales no contienen estas palabras que determinan el tipo que es (Solo disponen de la parte esencial del nombre).

Esta función recorre las filas del dataset analizando el nombre (ya formateado para que se más sencilla su lectura) y se añade la colunma tipoCalle con esta información. Para obtener el tipo se llamará a la función getTipoVia, la cual se muestra en Listing 2.5.

Listing 2.5: Función getTipoVia

```
1 import csv
    from re import findall
    def getTipoVia(nombreViaIni = ""):
           if(nombreViaIni == ""):
          return ""

arr = nombreViaIni.upper().split()
          # Para solo tener las primeras palabras y no por ejemplo Gran VIA nombreVia = arr[0]
10
          if (nombreVia.__contains__("CALLE")):
    return "Calle"
elif (nombreVia.__contains__("PASEO")):
    return "Paseo"
12
13
14
          elif (nombreVia.__contains__("PLAZA")):
    return "Plaza"
15 \\ 16 \\ 17 \\ 18
          elif (nombreVia.__contains__("GLORIETA")):
    return "Glorieta"
          elif (nombreVia.__contains__("RONDA")):
    return "Ronda"
elif (nombreVia.__contains__("CAMINO")):
19
20
21
22
23
                return "Camino"
          elif (nombreVia.__contains__("PISTA")):
    return "Pista"
24
25
26
          elif (nombreVia.__contains__("ANILLO")):
    return "Anillo"
27
28
          elif (nombreVia.__contains__("CRUCE")):
    return "Cruce"
          elif (nombreVia.__contains__("AUTOVIA")):
    return "Autovia"
29
30
          elif (nombreVia.__contains__("CARRETERA")):
    return "Carretera"
31
          elif (nombreVia.__contains__("PARQUE")):
    return "Parque"
33
35
          elif (nombreVia.__contains__("CUESTA")):
    return "Cuesta"
          return "Cuesta"
elif (nombreVia.__contains__("CAÑADA")):
    return "Cañada"
37
38
39
          elif (nombreVia.__contains__("AVENIDA")):
                            "Avenida"
          elif (nombreVia.__contains__("BULEVAR")):
    return "Bulevar"
41
42
43
          elif (nombreVia.__contains__("JARDIN")):
    return "Jardin"
\frac{44}{45}
          elif (nombreVia.__contains__("PARTICULAR")):
    return "Particular"
\frac{46}{47}
          elif (nombreVia.__contains__("POLIGONO")):
48
49
                 return "Poligono"
          elif (nombreVia.__contains__("GALERIA")):
    return "Galeria"
elif (nombreVia.__contains__("ESCALINATA")):
50
51
          return "Escalinata"
elif (nombreVia.__contains__("VIA")):
return "Via"
52
53
54
55
56
57
58
          elif (nombreVia.__contains__("PASARELA")):
          elif (nombreVia.__contains__("PASAJE")):
    return "Pasaje"
          elif (nombreVia.__contains__("PUENTE")):
```

```
return "Puente"
60
         elif (nombreVia.__contains__("COSTANILLA")):
    return "Costanilla"
62
         elif (nombreVia.__contains__("COLONIA")):
    return "Colonia"
64
         elif (nombreVia.__contains__("CARRERA")):
    return "Carrera"
66
67
         elif (nombreVia.__contains__("PLAZUELA")):
68
               return "Plazuela"
         elif (nombreVia.__contains__("ACCESO")):
    return "Acceso"
69
70
71
72
73
74
75
76
77
78
79
80
         elif (nombreVia.__contains__("POBLADO")):
    return "Poblado"
         elif (nombreVia.__contains__("PASADIZO")):
         return "Pasadizo"
elif (nombreVia.__contains__("TRASERA")):
               return "Trasera"
         elif (nombreVia.__contains__("SENDA")):
    return "Senda"
         elif (nombreVia.__contains__("ARROYO")):
    return "Arroyo"
81
82
         elif (nombreVia.__contains__("VALLE")):
    return "Valle"
83
84
         elif (nombreVia.__contains__("AEROPUERTO")):
               return "Aeropuerto"
85
86
         elif (nombreVia.__contains__("PASO_ELEVADO")):
    return "Paso Elevado"
87
88
         elif (nombreVia.__contains__("SENDA_CICLABLE")):
    return "Senda Ciclable"
              r (nombreVia.__contains__("PASAJE")):
return "Pasaje"
a:
          elif (nombreVia._
89
90
91
               return ""
```

A pesar de su simpleza actual esta función esta creada además para que si en futuras ocasiones se observasen nuevos tipos o abreviaturas sean sencillas da añadir sin necesidad de modificar el dataset desde el origen.

Crear una nueva propiedad llamada "palabras clave" que nos permita reducir al máximo los elementos del nombre de la via para asi comparar de forma más satisfactoria con los registros del Callejero de la ciudad.

Listing 2.6: Función annadirPalabrasClave

```
1 import csv
 3 from re import findall
 5 def annadirPalabrasClave(nombreCarpeta = "", nombreSinCsv = "", nRowNombre = -1, nFileIni = "-1", nFileFin =
        with open(nombreCarpeta + "/" + nombreSinCsv + nFileIni + ".csv") as csvfile:
             csvreader = csv.reader(csvfile, delimiter=";")
file = open(nombreCarpeta + "/" + nombreSinCsv + nFileFin + ".csv", "w")
             primeraLinea = True
for row in csvreader:
10
                  if (primeraLinea):
    fila = ";".join(row)
    fila = fila + ";" + "PALABRAS_CLAVE"
11
12
13
14
                        primeraLinea = False
\frac{15}{16}
                       palabrasClave = quitarConectores(row[nRowNombre])
                         palabrasClave = quitarPalabrasConflicto(palabrasClave)
                       palabrasClave - quite....
fila = ";".join(row)
fila = fila + ";" + palabrasClave
19
                   file.write(fila + os.linesep)
              file.close()
```

En el código de Listinf 2.6 se añade a este nuevo campo una palabra obtenida a partir del nombre a la que le han sido eliminados los conectores y palabras que se consideran genéricas o conflictivas (por ejemplo todas las referidas a tipos de via, elementos ya añadidos a este nuevo campo y que pueden estar en un dataset y no en otro por su caracter genérico).

Listing 2.7: Función quitarConectores

Listing 2.8: Función quitarPalabrasConflicto

```
1 import csv
 2 import os
 3 from re import findall
 6 def quitarPalabrasConflicto(nombreVia = ""):
         nombreVia = nombreVia.upper()
         if(nombreVia == ""):
               return
      listaPalabrasQuitar = ["CRUCE. ", "CRUCE ", "CALL. ", "CALL ", "CALLE ", "C/", "C/ ", "PASEO. ", "PASEO ", "PLAZA.", "PLAZA ", "GTA. ",
10
         "PLAZA.", "PLAZA ", "GTA. ",
"GTA ", "GLORIETA ", "RONDA. ", "RONDA ",
"CMNO. ", "CMNO ", "CAMINO ", "PISTA. ", "PISTA ", "AUTOVIA ", "CTRA. ",
"CTRA ", "CARRETERA ", "PQUE. ", "PQUE ", "PARQUE ", "CUSTA. ", "CUSTA ", "CUESTA ", "CÑADA. ", "CĂNDA ",
"CAÑADA ", "AVDA. ",
"AVDA ", "AVDA. ", "AVDA ", "AVENIDA ", "VIA ", "PASARELA ", "PASAJE ", "PUENTE ", "COSTANILLA ", "COLONIA ",
"CARRERA ",
"PLAZIETA ", "PUNTENTA ", "PONTONIA ", "PASALELA ", "PASAJE ", "PUENTE ", "COSTANILLA ", "COLONIA ",
11
12
13
14
                  DELA ", "BULEVAR ", "BULEV. ", "ESCALINATA ", "JARDÍN ", "JARDIN ", "JARDINES ", "PARTICULAR ", "POLÍGONO ", "POLIGONO ", "ACCESO ",
15
         "POBLADO ", "PASADIZO ", "TRASERA ", "SENDA ", "GALERÍA ", "GALERIA ", "ARROYO ", "VALLE ", "AEROPUERTO ", "PASO_ELEVADO", "SENDA_CICLABLE"]
16
17
         for a in listaPalabrasQuitar:
               if(nombreVia._contains_(a) and findall('[A-z]{1}' + a, nombreVia) == []): #Para evitar errores por ejemplo en SegoVIA
19
20
                     arr = nombreVia.split()
                    22
24
26
         # Podria ser M11 por ejemplo
if(nombreVia.strip()[0] == '/'):
28
              # Para eliminar descripciones al inicio y su / p.ej. SENDA_CICLABLE / AV.ROSALES / CARRET.VILLAV. A
                      VALLECA
30
               nombreVia = nombreVia.replace(',', '', 1)
         return nombreVia
```

En los códigos mostrado en Listing 2.7 y 2.8 se observan las funciones antes mencionadas. Los conectores han sido obtenidos de los propios datasets aquí utilizados por lo tanto pueden ocurrir errores puntuales en el caso de que apareciesen nuevos no esperados. Se ha optado por esta opción en vez de buscar palabras monosílabas por ejemplo debido a que da mayor seguridad en cuanto a errores, y es preferible detectar un conector sin omitir que una palabra eliminada que si era parte esencial.

En el caso de las palabras conflicto, la función está destinada a eliminar los tipos de vias como ya se mencionó anteriormente.

Cabe destacar que esta función es llamada por todos los datasets incluido el Callejero de Madrid. Ésto permite que todos por igual hayan recibido las mismas transformaciones para obtener sus palabras clave, de tal forma que sea más sencillo su cruce para obtener los identificadores.

Para finalizar esta primera fase, todos los vocabularios son comparados con el Callejero de Madrid. Para ello se comparan las palabras claves de sus nombres, anteriormente obtenidas, y se comprueba que sean las mismas para así asignarles el identificador. Éste identificador lo proporciona el ayuntamiento, por ello es importante este cruce con el

Callejero, del cual se obtendrán.

Listing 2.9: Función crearFichNombresId

```
1 import csv
 2 import os
3 import difflib
  4 import unicodedata
 10
          primeraLinea = True
          nCorrect = 0
nIncorrect = 0
12
13
          nIncorrect - v
contadorFila = 0
for row in csvreader:
    if (primeraLinea):
        fila = ";".join(row)
        fila = fila + ";" + "ID_CALLE"
\frac{14}{15}
\frac{16}{17}
18
                       primeraLinea = False
20
21
22
                       if (nRowEsCruce != -1 and row[nRowEsCruce] == "1"): # Si es cruce tendra varios ids
                              arrCallesCruce = getArrCalles(row[nRowPalabrasClave].upper())
24
25
                             arrCallesCruce = [row[nRowPalabrasClave].upper()]
26
                       listIds = []
nElem = 0
28
29
30
                       for nombre in arrCallesCruce: #TODO: quitar
31
                              ','if(not(nombre.__contains__("EUGENIA"))):
32
                                    continue,
                             #TODO: quitar
segundaVuelta = False
33
34
                             nVuelta = 0
while (nVuelta == 0 or segundaVuelta):
35
36
37
38
                                   nVuelta = nVuelta + 1
with open("callejeroCSV" + "/" + "nombres_IDs" + ".csv") as csvCallejero:
                                         csvreaderCallej = csv.reader(csvCallejero, delimiter=";")
primeraLinea = True
encontrado = False
for rowCallj in csvreaderCallej:
39
40
41
42
                                               if (primeraLinea):
    primeraLinea = False
43
44
                                                else:
45
46
                                                     try:
                                                            # TODO: quitar
'''if (rowCallj[11].__contains__("EUGENIA")):
\frac{47}{48}
                                                            print("OK2")'''
# TODO: quitar
if((nElem >0 or not(tieneTipoCalle) or
(segundaVuelta and esTipoCalleOmitible(row[nRowTipoVia].upper()))
or checkerPalabras(rowCallj[1].upper(), row[nRowTipoVia].upper(),False))
49
51
53
                                                             anu #Porque si es la segunda entrada no contiene el tipo de via checkearPalabras(rowCallj[11].upper(), nombre.upper(), segundaVuelta)):
54
                                                                  listIds.append(rowCallj[0])
encontrado=True
56
57
                                                                   nCorrect = nCorrect+1
58
59
60
                                                                   segundaVuelta=False
                                         break
except IndexError:
print("Error")
# Posible error de que haya una linea vacia extra al final
if (not(encontrado) and not(segundaVuelta)):
segundaVuelta = True
continua
61
62
63
64
65
66
                                                continue
67
68
                                          \quad \hbox{if (not (encontrado) and segundaVuelta):} \\
                                                segundaVuelta = False
                                               listIds.append("-1")
nIncorrect = nIncorrect + 1
69
70
\frac{71}{72}
                                   csvCallejero.close()
                                   nElem = nElem + 1
                      nElem = nblem . .
fila = ";".join(row)
fila = fila + ";" + listIds.__str__()
contadorFila = contadorFila+1
if(contadorFila % 20 == 0):
    print("Fila: " , contadorFila)
73
74
75
76
                print("Fila: " , cont
file.write(fila + os.linesep)
79
           file.close()
          print("Incorrectas: ", nIncorrect, " || Correctas: ", nCorrect)
```

En Listing 2.9 se muestra la funcion crearFichNombresId. Ésta funcion realiza numerosas vueltas para cada registro. Primero selecciona el elemento del dataset que se va a analizar

(que aun no tenga id). Segundo comprueba que no sea un cruce, y en caso de serlo llama a la funcion getArrCalles (Listing 2.10) para obtener las dos o más vias que lo componen (y comprobar los ids para cada una de ellas). Tercero recorre el dataset del callejero para comparar las palabras clave y obtener coincidencias. Para ello sigue los siguientes pasos:

- 1. Compara el tipo de via, en caso de no coincidir será distinto tipo de calle y por tanto diferente (esto no es válido para el caso de cruces ya que la via secundaria no tendría tipo, por tanto se omite).
- 2. En caso de no coincidir en ningún elemento, vuelve a dar otra vuelta al callejero rebajando las restricciones. No se comprueba que sea del mismo tipo de via y se aplica una fórmula que se detallará posteriormente para que se permitan letras añadidas, suprimidas o modificadas (por erratas, plurales, femeninos y masculinos...)

Por último se imprime por pantalla los conteos de los elementos que han coincidido y los que no para determinar la eficacia del proceso (aproximado, ya que pueden haberse cometido errores no detectados y que se hayan dado paso a identificadores mal asociados).

Listing 2.10: Función getArrCalles

```
1 import csv
 2 import os
 4 import unicodedata
 6 def getArrCallesRec(nombreDataset = ""):
          listaNombres = []
          if (nombreDataset._contains_('/')):
    n1 = nombreDataset.split().index('/')
10
                 n1 = nombreDataset.split().index('/')
txtRestante = " ".join(nombreDataset.split()[n1 + 1:])
nombre = " ".join(nombreDataset.split()[0:n1])
elemRecursiv = getArrCallesRec(txtRestante)
listaNombres = [nombre, elemRecursiv]
11
12
13
14
          else:
    listaNombres = [nombreDataset]
15
16
17
18
19
          return listaNombres
20 def getArrCalles(nombreDataset = ""):
            list1 = getArrCallesRec(nombreDataset)
21
           txt = list1._str__().replace('[', '').replace(']', '')
22
23
          list3 = []
nombre=""
25
26
          for elem in txt.split():
                 if(elem in txt.spiit();
if(elem __contains__("'") and nombre != "");
nombre = nombre + " " + elem.replace("'", "").replace(",", "")
27
                       list3.append(nombre)
nombre = ""
29
                 elif (elem.__contains__("'") and nombre == "" and elem.replace("'", "", 1).__contains__("'")):
    nombre = elem.replace("'", "").replace(",", "")
    list3.append(nombre)
31
33
                 nombre = """
elif (elem.__contains__(",") and nombre == ""):
    nombre = elem.replace(",", "").replace(",",
35
                 else:
                        nombre = nombre + " " + elem
          return list3
```

En este codigo se obtiene un array de calles a partir del nombre. Siguiendo el patrón "calle1 / calle2 " que se ha utilizado para los cruces, se separan sus elementos devolviendo un array con los nombres completos de cada una de ellas separados.

Listing 2.11: Función checkearPalabras

```
1 import csv
2 import os
3 import difflib
4 import unicodedata
5 def checkearPalabras(nombreCallej = "", nombreDataset = "", segundaVuelta = False):
6 # En la segunda vuelta también se omite el tipo de calle
```

```
8
10
11
12
                         return True
                13
15
                        return True
               18
20
22
                          # P.ej. COMANDANTE ZORITA
24
                         # P.eg. CUMANDANIE ZONII...
nombreDataset1 = nombreDataset
nombreCallej1 = nombreCallej
26
                         if (nombreDataset.__contains__("COMANDANTE")):
    nombreDataset1 = nombreDataset.replace("COMANDANTE", "AVIADOR")
28
                         29
30
32
33
34
                                 return True
35
                         "
if (nombreCallej.replace(" ", "").__contains__("M30") or
nombreCallej.replace(" ", "").__contains__("M40")):
return False #Si no la ha detectado ya es que hay problemas y las Bicicletas no pueden circular
36
37
                                              por ahi
39
40
                          # Por ejemplo FERNADO VI --> FERNANDO VI
                         longCadena1 = nombreDataset.__len__()
longCadena2 = nombreCallej.__len__()
if(longCadena1 < (longCadena2*0.7) or longCadena1 > (longCadena2 * 1.3)) :
42
43
44
                                   # Si el tamaño de la cadena difiere mucho
                                   return False
                         diff = difflib.ndiff(nombreCallej1.replace(" ", ""), nombreDataset1.replace(" ", ""))
diferenciastxt = ''.join(diff)
46
48
                         #Primero comprobar 1 solo error en total
if ((diferenciastxt.count('+') + diferenciastxt.count('-')) <= 1</pre>
50
                                   and not (nombreCallej._contains_("DARROS")) # BARRIOS :: B, and not (nombreCallej._contains_("OLIVAR")) # Bolivar - Olivar and not (nombreCallej._contains_("MANZANAR")) # MANZANARES ::
52
54
                                   print("Posible Conicidencia Simple: ", nombreDataset, " :: ", nombreCallej)
56
58
59
                         # Segundo comprobar 1 sustitucion if (diferenciastxt.count('-') <= 1 and diferenciastxt.count('-') <=1
60
                                  (diferenciastxt.count('+') <= 1 and diferenciastxt.count('-') <=1
and not(nombreDataset.__contains__("HORTALEZA")) #Hortaleza - Fortaleza
and not (nombreCallej.__contains__("GALENA")) # Gilena - Galena
and not (nombreCallej.__contains__("PEAL")) # Real - Peal
and not (nombreCallej.__contains__("OLIVAR")) # Bolivar - Olivar
and not (nombreCallej.__contains__("CRUCES")) # RUICES :: CRUCES
and not (nombreCallej.__contains__("HAYA")) # RAYA :: HAYA
and not (nombreCallej.__contains__("HAYA")) # ONA :: OCA
and not (nombreCallej.__contains__("MAYANAR")) # MANZANARES :: MANZANAR
and not (nombreCallej.__contains__("GANDILEJAS") and nombreCallej.__contains__("CANALEJAS")) #
CANDILEJAS :: CANALEJAS
and not (nombreCallej._contains__("LIMON")) # JUAN VIGON :: JUAN LIMON</pre>
61
62
63
64
65
66
67
68
69
                                 and not (nombreCallej._contains_("LIMON")) # JUAN VIGON :: JUAN LIMON
and not (nombreCallej._contains_("CENICIENTOS")) # CENICIENTA :: CENICIENTOS
and not (nombreCallej._contains_("CENICIENTOS")) # MANOTERAS :: CANTERAS
and not (nombreCallej._contains_("MANOTERAS")) # MANOTERAS :: SANTERAS
and not (nombreCallej._contains_("GOR")) # SIERRA GADOR :: SIERRA GOR
and not (nombreCallej._contains_("GADOR")) # SIERRA GADOR :: SIERRA GUDAR
and not (nombreCallej._contains_("ERASMO")) # SANZ RASO :: SAN ERASMO
and not (nombreCallej._contains_("ERASMO")) # SANZ RASO :: SAN TIAGO ALIO
and not (nombreCallej._contains_("ALIO")) # SANTIAGO AMON :: SANTIAGO ALIO
and not (nombreDataset._contains_("PARVILLAS")) # PARVILLAS :: MARAVILLA
and not (nombreDataset._contains_("MARMOLINA")) # MARMOLINA :: CAROLINA
and not (nombreCallej._contains_("SAMNIEGO")) # SAN DIEGO :: SAMANIEGO
and not (nombreCallej._contains_("BARROS")) # BARRIOS :: BARROS
and not (nombreCallej._contains_("BARROS")) # BARRIOS :: BARROS
and not (nombreDataset.replace(" ", "").__contains_("MADRIDRIO")) # MADRID RIO :: MADRE L
):

print("Posible Conicidencia Suctitation :: "
70
72
73
74
75
\frac{76}{77}
78
80
81
                                                                                                                                                                                       BARROS
O")) # MADRID RIO :: MADRE DIOS
82
84
                                   print("Posible Conicidencia Sustitucion: ", nombreDataset, " :: ", nombreCallej)
86
                                   return True
                         # Dependiendo de la longitud de la palabra acepta 1 error o más
if (diferenciastxt.count('+') <= (longCadenal/10+1) and diferenciastxt.count('-') <= (longCadenal/10+1)
and not (nombreDataset.__contains__("HORTALEZA")) #Hortaleza - Fortaleza
and not (nombreCallej.__contains__("EGALENA")) # Gilena - Galena
and not (nombreCallej.__contains__("PEAL")) # Real - Peal
and not (nombreCallej.__contains__("OLIVAR")) # Bolivar - Olivar
and not (nombreCallej._contains__("CRUCES")) # RUICES :: CRUCES
and not (nombreCallej._contains__("HAYA")) # RAYA :: HAYA
and not (nombreCallej._contains__("GOA")) # OÑA :: OCA
and not (nombreCallej._contains__("COA")) # OÑA :: OCA
and not (nombreCallej._contains__("CANDILEJAS") and nombreCallej.__contains__("CANDILEJAS")) #
88
89
90
92
93
94
95
96
                                   and not (nombreDataset.__contains__("CANDILEJAS") and nombreCallej.__contains__("CANALEJAS") ) #
```

```
CANDILEJAS :: CANALEJAS
                                    CANDILEJAS :: CANALEJAS

and not (nombreCallej._contains_("LIMON")) # JUAN VIGON ::
and not (nombreCallej._contains_("CENICIENTOS")) # CENICIENTA

and not (nombreCallej._contains_("CANTERAS")) # MANOTERAS ::
and not (nombreDataset._contains_("MANOTERAS")) # MANOTERAS ::
and not (nombreCallej._contains_("GADOR")) # SIERRA GADOR ::
and not (nombreCallej._contains_("GADOR")) # SIERRA GADOR ::
and not (nombreCallej._contains_("ERASNO")) # SANZ RASO ::
and not (nombreCallej._contains_("ERASNO")) # SANZ RASO ::
                                                                                                                                                                                                   JUAN LIMON
                                                                                                                                                                                                                  CENTCIENTOS
100
102
                                                                                                                                                                                                             SANTERAS
103
                                                                                                                                                                                                    SIERRA GOR
104
                                                                                                                                                                                                           STERRA GUDAR
                                     and not (nombreCallej.__contains__("ALIO")) # SANTIAGO AMON ::
and not (nombreDataset.__contains__("PARVILLAS")) # PARVILLAS ::
and not (nombreDataset.__contains__("MARMOLINA")) # MARMOLINA ::
and not (nombreCallej.__contains__("SAMANIEGO")) # SAN DIEGO ::
and not (nombreDataset.replace(" ", "").__contains__("SANDIEGO"))
                                                                                                                                                                                                          SANTIAGO ALIO
106
108
                                                                                                                                                                                                             CAROLINA
                                                                                                                                         __contains__("SANDIEGO"))
S")) # BARRIOS :: B
110
                                                                                                                                                                                                        # SAN DIEGO ::
                                                                                                                                                                                                                                                    SAN DACTO
                                                        (nombreCallej.__contains__("BARROS"))
                                     and not
                                    and not (nombreDataset.replace(" ", "").__contains__("MONTEAYA")) # MONTE AYA :: MON and not (nombreDataset.replace(" ", "").__contains__("MONTEAYA")) # MONTE AYA :: MON not (nombreDataset.replace(" ", "").__contains__("PALOROSA")) # PALO ROSA :: ALC and not (nombreDataset._contains__("SACROMONTE")) # SACROMONTE :: CERRO MONTE - and not (nombreDataset.replace(" ", "").__contains__("EDUARDOAUNOS")) # EDUARDO AUNOS EDUARDO UROSA -
112
                                                                                                                                                                                                                                                    MONTANA -
114
                                                                                                                                                                                                                                                    ALCORISA
116
117
                                     and not (nombreDataset.replace(" ", "").__contains__("MADRIDRIO")) # MADRID RIO :: MADRE DIOS
                                     # Si se han añadido o quitado 2 o menos letras, se puede considerar igual
119
120
                                     print("-----Posible Conicidencia: ", nombreDataset, " :: ", nombreCallej)
121
                                     return True
                 return False
123
```

En esta función primero se sustituyen los elementos como tildes y se comparan para obtener coincidencias.

En caso de que ésta no se diese, el programa daría una segunda vuelta, a la cual la función actuaria de diferente manera. Primero igual, ya que podría darse coincidencia por haber eliminado el tipo de via de la restricción. Si tampoco se diese el caso se comprobaría primero con una letra añadida o eliminada, despues con 2 (lo que permitiría letras cambiadas) y despúes con una diferencia variable dependiendo de su dimensión.

Se han añadido numerosas restricciones observadas en el proceso y también se imprime por pantalla los cambios realizados para que puedan ser vistos y modificados en caso de errores.

Listing 2.12: Función esTipoCalleOmitible

```
import os
 3 import difflib
 4 import unicodedata
 6 def esTipoCalleOmitible(nombreVia = ""):
      if(nombreVia == ""):
    return True
      "E").replace("Í"
10
      if (nombreVia.__contains__("CALLE")):
11
          return True
      elif (nombreVia.__contains__("PASEO")):
13
14
15
           return True
      elif (nombreVia.__contains__("PLAZA")):
16
17
18
19
       elif (nombreVia.__contains__("GLORIETA")):
      elif (nombreVia.__contains__("RONDA")):
20
21
      elif (nombreVia.__contains__("CAMINO")):
      elif (nombreVia. contains ("PISTA")):
23
24
25
       elif (nombreVia.__contains__("ANILLO")):
26
27
      elif (nombreVia.__contains__("CRUCE")):
28
29
      elif (nombreVia.__contains__("AUTOVIA")):
           return False
      elif (nombreVia.__contains__("CARRETERA")):
31
32
33
       elif (nombreVia.__contains__("PARQUE")):
34
35
           return False
       elif (nombreVia.__contains__("CUESTA")):
36
37
           return True
       elif (nombreVia.__contains__("CAÑADA")):
           return False
       elif (nombreVia.__contains__("AVENIDA")):
           return True
       elif (nombreVia.__contains__("BULEVAR")):
```

```
return True
elif (nombreVia.__contains__("JARDIN")):
43
44
             return False
        elif (nombreVia.__contains__("PARTICULAR")):
46
             return False
        elif (nombreVia.__contains__("POLIGONO")):
        return False elif (nombreVia.__contains__("GALERIA")):
48
49
50
51
52
53
54
55
56
57
58
59
60
        return False
elif (nombreVia.__contains__("ESCALINATA")):
             return False
        elif (nombreVia.__contains__("VIA")):
             return True
        elif (nombreVia.__contains__("PASARELA")):
        return True
elif (nombreVia.__contains__("PASAJE")):
             return True
        elif (nombreVia.__contains__("PUENTE")):
             return False
61
62
        elif (nombreVia.__contains__("COSTANILLA")):
             return False
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
        elif (nombreVia.__contains__("COLONIA")):
             return False
        elif (nombreVia.__contains__("CARRERA")):
             return True
        elif (nombreVia.__contains__("PLAZUELA")):
             return False
        elif (nombreVia.__contains__("ACCESO")):
             return True
        elif (nombreVia.__contains__("POBLADO")):
             return False
        elif (nombreVia.__contains__("PASADIZO")):
    return False
        elif (nombreVia.__contains__("TRASERA")):
             return False
        elif (nombreVia.__contains__("SENDA")):
return True
        elif (nombreVia.__contains__("ARROYO")):
80
             return False
        elif (nombreVia.__contains__("VALLE")):
    return False
81
82
        elif (nombreVia.__contains__("AEROPUERTO")):
    return False
83
84
85
        elif (nombreVia.__contains__("PASO_ELEVADO")):
86
87
88
89
             return False
        elif (nombreVia.__contains__("SENDA_CICLABLE")):
        return True
```

Esta función es llamada en la segunda vuelta del bucle por si el tipo de via al que hace referencia no pudiese ser obviado. Esto es por ejemplo en Parque, Polígono... No es posible comparar un parque con una calle que tengan el mismo nombre y aceptar que es válido, en cambio una avenida con una calle si. Para ello se debe especificar los tipos que pueden ser comparables y los que no.

2.5. Transformaciones en los vocabularios: Fase 2

En esta segunda fase de transformaciones se realizarán cambios menores. Una vez ya se han realizado los cambios necesarios para obtener las palabras clave y los identificadores, se realizarán cambios de nombres en las propiedades de los datasets y se añadirán nuevos elementos como por ejemplo municipio.

Estos cambios se realizarán a partir de lo generado en la Fase1.

En esta sección los cambios serán diferentes por cada Dataset por tanto se detallarán por separado.

2.5.1. Cambios relativos a CicloCarriles

Los relativos a CicloCarriles serán los siguientes:

• Resumen de cambios:

out: CAPA; TX_NOMBRE; MaxSimpTol; MinSimpTol; FECHA; Longitud; TIPO_VIA; PALABRAS_CLAVE; ID_CALLE

in: tipoUso; nombreVia; distMaxExclusBici; longitud; TIPO_VIA; PALABRAS_CLAVE; ID_CALLE; carrilExclusBici; municipio

Para el caso de tipoUso, nombreVia, distMaxExclusBici y longitud no se realizará ninguna modificación en los registros. Únicamente se ha cambiado el nombrado.

Se han eliminado las propiedades MinSimpTol y FECHA, junto con sus datos, debido a que no aportan información añadida y por ejemplo en el caso de la distancia minima, se sustituirá por la propiedad carrilExclusBici que dará la misma información de forma más clara con otros tipos de datos.

Se han añadido carrilExclusBici y municipio.

Se modificarán los datos de Longitud, ya que están en kilómetros y es preferible, para la longitud de las calles en una ciudad, que se representen en metros por ser el formato estándar.

Cambios en los nombres y añadir carrilExclusBici y municipio:

Listing 2.13: Función variosCambiosEnCicloCarriles

En el código mostrado primero se cambian los nombres de las propiedades a tipoUso, nombreVia, distMaxExclusBici y longitud. Posteriormente se añaden las de carrilExclusBici y municipio. Para su cálculo, en el primer caso se multiplica por 1.000 la longitud (para convertir kilometros en metros y en el segundo se añade la palabra Madrid, ya que siempre será el municipio para los datos aqui utilizados.

Listing 2.14: Función eliminarColumnaCompleta

```
1 import csv
3 from re import findall
5 def eliminarColumnaCompleta(nombreCarpeta="", nombreSinCsv="", nFileIni = "-1", nFileFin = "-1", nombreColumna
        with open(nombreCarpeta + "/" + nombreSinCsv + nFileIni + ".csv") as csvfile:
            csvreader = csv.reader(csvfile, delimiter=";")
file = open(nombreCarpeta + "/" + nombreSinCsv + nFileFin + ".csv", "w")
             primeraLinea = True
nRowColumna = -1
10
             for row in csvreader:
11
12
                 if (primeraLinea):
    nRowColumna = row.index(nombreColumna)
13
14
15
                      del row[nRowColumna]
                      fila = ";".join(row)
16
                      primeraLinea = False
                      del row[nRowColumna]
19
                 fila = ";".join(row)
file.write(fila + os.linesep)
             file.close()
```

En el codigo de Listing 2.14 se muestra una función que se reutilizará para todos los datasets. En ella se elimina una columna completa, tanto su nombre como sus datos en todos los registros.

En el caso del dataset de CicloCarriles se eliminarán las columnas FECHA y MinSimplTol, como se observa en las siguientes llamadas a esta función.

Listing 2.15: Llamadas a eliminarColumnaCompleta

```
1 eliminarColumnaCompleta(nombreCarpeta, nombreSinCsv, "2", "3", "FECHA")
2 eliminarColumnaCompleta(nombreCarpeta, nombreSinCsv, "3", "4", "MinSimpTol")
```

2.5.2. Cambios relativos a Accidentes de Bicicletas

Los relativos a Accidentes de Bicicletas serán los siguientes:

• Resumen de cambios:

out: Nº EXPEDIENTE; FECHA; HORA; CALLE; NÚMERO; DISTRITO; TIPO ACCIDENTE; ESTADO METEREOLÓGICO; TIPO VEHÍCULO; TIPO PERSONA; RANGO EDAD; SEXO; LESIVIDAD*;esCruce;tipoVia;palabrasClave;idVia

in: numero Expediente; fecha; hora; nombre Via; portal; distrito; tipo Accidente; meteorologia; tipo Vehiculo; tipo Persona Afectada; typical AgeRange; gender; lesividad; esCruce; tipo Via; palabras Clave; idVia; municipio

Se modifican el nombre de las columnas N°Expediente, FECHA, HORA, CALLE, NÚMERO, DISTRITO, TIPO ACCIDENTE, ESTADO METEREOLÓGICO, TIPO VEHÍCULO, TIPO PERSONA, RANGO EDAD, SEXO, LESIVIDAD* para que sea más entendible y coincida con la definición antes desarrollada.

Se han modificado los datos de las columnas typicalAgeRange (RANGO EDAD) para que seguir el formato definido por schema.org (añoInicio - añoFin) (p.ej. 10-12)

Se han mofificado los datos de gender (SEXO) de para que también concuerden con los definidos en schema.org. En este caso el dataset solo proporciona los valores Hombre y Mujer, por lo tanto solo se sustituirán por Male y Female.

Se añade la columna municipio asignandole a cada registro el valor "Madrid"para esta propiedad.

Listing 2.16: Función variosCambiosEnAccidentesBici

```
1 import csv
 3 from re import findall
 4

def variosCambiosEnAccidentesBici(nombreCarpeta="", nombreSinCsv="", nFileIni = "-1", nFileFin = "-1"):

with open(nombreCarpeta + "/" + nombreSinCsv + nFileIni + ".csv") as csvfile:

csvreader = csv.reader(csvfile, delimiter=";")

file = open(nombreCarpeta + "/" + nombreSinCsv + nFileFin + ".csv", "w")
                  file = open(nombreCarpeta + primeraLinea = True
                   nColTypicalAge =
11
                   nColGender =
                   for row in csvreader:
13
                          if (primeraLinea):
                                 fila = ";".join(row)
nColTypicalAge = row.index("RANGO EDAD")
nColGender = row.index("SEXO")
15
16
                                 # Cambio nombres: N^O Expediente, FECHA, HORA, CALLE, NÚMERO, DISTRITO, TIPO ACCIDENTE, # ESTADO METEREOLÓGICO, TIPO VEHÍCULO, TIPO PERSONA, RANGO EDAD, SEXO,
17
18
                                          I.ESTVTDAD*
                                                      replace("N° EXPEDIENTE", "numeroExpediente").replace("FECHA", "fecha")\
.replace("HORA", "hora").replace("CALLE", "nombreVia").replace("NÚMERO", "portal")\
.replace("DISTRITO", "distrito").replace("TIPO ACCIDENTE", "tipoAccidente")\
.replace("ESTADO METEREOLÔGICO", "meteorologia").replace("TIPO VEHÍCULO",
                                 fila = fila.replace("\mathbb{N}^{\circ}
19
21
22
                                                      "tipoVehiculo")\
.replace("TIPO PERSONA", "tipoPersonaAfectada").replace("RANGO EDAD",
    "typicalAgeRange")\
23
24
25
                                 # Añadir municipio:
                                                       .replace("SEXO", "gender").replace("LESIVIDAD*", "lesividad")
26
27
                                 primeraLinea = False
28
29
                                 # Modificar el Genero
30
31
                                 generoEspan = row[nColGender]
                                 if (generoEspan == "Hombre"):
32
33
                                        row[nColGender] = "Male"
(generoEspan == "Mujer"):
                                 elif(generoEspan ==
\frac{34}{35}
                                        row[nColGender] = "Female"
                                 # Modificar TypicalAgeRange
row[nColTypicalAge] = getTypicalAgeRangeOk(row[nColTypicalAge])
36
37
39
                                 fila = ";".join(row)
                                 # Añadir municipio:
fila = fila + ";" + "Madrid"
40
```

```
42 file.write(fila + os.linesep)
43 file.close()
```

En 2.16 se muestra el codigo de los cambios en el dataset de Accidentes de Bicicleta. Primero se cambian los nombres antes mencionados por los nuevos para que coincidan con los definidos en el vocabulario.

Segundo se realizan las transformaciones para typical AgeRange y para gender. Para Gender únicamente el dataset tiene 2 valores posibles, Hombre y Mujer, por lo que se ha optado por considerar únicamente esas posibilidades y sustituirlos por sus nombres en inglés, lo definido por schema.org. Si se diese el caso de nuevas posiblidades en esta propiedad se deberían añadir. En [1] https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=20f4a87ebb65b510VgnVCM1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default se puede consultar la documentación asociada a este dataset y hasta el momento solo puede haber estas 2 opciones y no asignado.

Para typicalAgeRange se ha creado una nueva función a la que llama para obtener el rango de edad en el formato estandar definido por schema.org. Se muestra en el siguiente código:

Listing 2.17: Función getTypicalAgeRangeOk

```
1 import csv
    import os
from re import findall
    def getTypicalAgeRangeOk(txtOld = "")
         if(txtOld.replace(" ",
    return ""
         if(txtOld.upper().__contains__("DESCONOCIDA")):
         arrPal = txtOld.split()
# "DE 30 A 34 AÑOS"
10
11
12
               fIni = arrPal[1]
fFin = arrPal[3]
\frac{13}{14}
         if (fIni > fFin):
    fIni, fFin = fFin, fIni
txtFin = fIni + "-" + fFin
except IndexError:
\frac{15}{16}
17
18
19
               txtFin = ""
print("Distinto formato Rango Edad: ", txtOld)
20
         # En caso de que no siga el formato estandar if(findall('([0-9]{1})-([0-9]{1})', txtFin) == []):
21
22
23
               fIni = -1
fFin = -1
24
25
26
               for elem in arrPal:
               27
28
29
30
31
                    if(fIni > fFin):
fIni, fFin = fFin, fIni
return fIni + "-" + fFin
32
33
35
         return txtFin
```

En 2.17 comprueba primero el formato en el que vienen la totalidad de los datos de esta columna (De 10 A 20 AÑOS) y lo devuelve en el formato estandar de schema.org (10-20). En caso de que el dataset contenga la información con distinta representación, se comprueba de forma básica si es posible encontrar en todo el texto dos números separados. En este caso se considerará que el primero es la fecha inicio y el segundo la final. Si es posible esta combinación se devolverán con el formato antes mencionado.

Por si se diera el caso de que la fecha estuviese cambiada de sentido, se comprueba que el inicio sea menor que el fin, en caso contrario se invierten para seguir el estandar.

2.5.3. Cambios relativos a Calles Tranquilas

Los relativos a Calles Tranquilas serán los siguientes:

• Resumen de cambios:

out: OBJECTID; ID_EJE; ID_GRUPO; ID_TIPO; TX_CAPA; NU_DIRECCI; NU_DOBLE_S; TX_NOMBRE; TX_NOMBRE_; Shape_Leng; tipoVia; palabras-Clave; idVia

in: idObject; idEje; idGrupo; tipoUso; nuDireccion; nuDobleS; nombreVia; nombreViaReduc; longitud; tipoVia; palabrasClave; idVia; municipio

Se modifican el nombre de las columnas OBJECTID, ID_EJE, ID_GRUPO, TX_CAPA, NU_DIRECCI, NU_DOBLE_S, TX_NOMBRE, NU_DOBLE_S y Shape_Leng para que sean más entendibles y coincida con la definición antes desarrollada. En este caso los datos de la longitud si están en metros por lo tanto no es necsario cmabiarlos.

Se ha eliminado la columna ID_TIPO ya que representa con un codigo numérico lo mismo que TX_CAPA (llamada a partir de ahora tipoUso). Se ha optado por esta última propiedad ya que comparte valores con tipoUso del dataset CicloCarriles, y por lo tanto se entiende que es más utilizado en otros dataset que la primera propiedad y que, al representar lo mismo, es mñas recomendable conservar si la segunda. Se añade la columna municipio asignandole a cada registro el valor "Madrid"para esta propiedad.

Listing 2.18: Función variosCambiosEnCallesTranquilas

```
1 import csv
 3 from re import findall
 5 def variosCambiosEnCallesTranquilas(nombreCarpeta="", nombreSinCsv="", nFileIni = "-1", nFileFin =
        with open(nombreCarpeta + "/" + nombreSinCsv + nFileIni + ".csv") as csvfile:
             file = open(nombreCarpeta + "/" + nombreSinCsv + nFileFin + ".csv", as carried to csvreader = csv.reader(csvfile, delimiter=";")
              primeraLinea = True
10
              for row in csvreader:
                                ":".join(row)
                        fila =
                       14
16
                                       "idGrupo").replace("TX_CAPA","tipoUso").replace("NU_DIRECCI","nuDireccion")\
.replace("NU_DOBLE_S", "nuDobleS").replace("TX_NOMBRE_","nombreViaReduc")\
.replace("TX_NOMBRE","nombreVia").replace("Shape_Leng","longitud")
                        # Añadir municipio:
                                              + "municipio"
                        fila = fila + ";
                        primeraLinea = False
                        fila = ";".join(row)
                  # Añadir municipio:
fila = fila + ";" + "Madrid"
file.write(fila + os.linesep)
26
```

En 2.18 se muestra el codigo de los cambios en el dataset de Calles Tranquilas En este caso se cambian los nombres antes mencionados por los nuevos para que coincidan con los definidos en el vocabulario y se añade la columna municipio con el valor "Madrid.^{en} todos los registros.

En el siguiente código se muestra la llamada a la funcion eliminarColumnaCompleta

?? para suprimir del dataset ID_TIPO.

 ${\bf Listing~2.19:~Llamadas~a~eliminar Columna Completa}$

 ${\tt 1} \ {\tt eliminarColumnaCompleta(nombreCarpeta, nombreSinCsv, "2", "3", "ID_TIPO")}$

Capítulo 3

Resultados y conclusiones

Resumen de resultados obtenidos en el TFG. Y conclusiones personales del estudiante sobre el trabajo realizado.

Bibliografía

- [1] http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero/index-en.html
- [2] http://schema.org
- [3] https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=20f4a87ebb65b510VgnVCM1000001d4a900aRCRD&vgnextchannel= 374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default
- [4] https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/
 ?vgnextoid=435a7cd5de319410VgnVCM1000000b205a0aRCRD&vgnextchannel=
 374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default
- [5] https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=a320f5ac548f4410VgnVCM1000000b205a0aRCRD&vgnextchannel= 374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default
- [6] https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=a320f5ac548f4410VgnVCM1000000b205a0aRCRD&vgnextchannel= 374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default
- [7] https://datos.madrid.es/sites/v/index.jsp?vgnextoid= b3c41f3cf6a6c410VgnVCM2000000c205a0aRCRD&vgnextchannel= 374512b9ace9f310VgnVCM100000171f5a0aRCRD
- [8] https://datos.madrid.es/egob/new/detalle/auxiliar/mapa.jsp?geoUrl=/egob/catalogo/205115-4-calles-tranquilas.kml
- [9] https://datos.ign.es/def/btn100/index-es.html#calzada
- [10] https://www.ine.es/daco/daco42/codmun/codmunmapa.htm
- [11] https://datos.ign.es/def/btn100/index-es.html#calzada
- [12] http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero/index-en.html#tipoVia
- [13] http://vocab.linkeddata.es/datosabiertos/def/sector-publico/territorio/index-en.html#Municipio
- [14] http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero/index-en.html#Via
- [15] https://schema.org/gender
- [16] https://lists.w3.org/Archives/Public/public-schemaorg/20190ct/0013.html

- $[17] \ \mathtt{https://schema.org/typicalAgeRange}$
- $[18] \ \, \texttt{http://vocab.linkeddata.es/datosabiertos/def/urbanismo-infraestructuras/callejero/index-en.html\#Portal}$

Anexo

Este capítulo es opcional, y se escribirá de acuerdo con las indicaciones del Tutor.