

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



“INTEGRACIÓN DE DATOS DE LA CONVENCION CONSTITUCIONAL CON FUENTES EXTERNAS”

JOAQUÍN ALEJANDRO CASTILLO TAPIA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Carlos Buil-Aranda
Profesor Correferente: José Luis Martí

Julio - 2023

DEDICATORIA

Dedicado a quienes ven en los datos una hermosa oportunidad de crear conocimiento.

AGRADECIMIENTOS

Quiero agradecer a Félix Tapia, mi abuelo paterno y la persona más inteligente que conozco, por recordarme en cada palabra que no se deben dejar escapar las oportunidades, y por enseñarme que a pesar de las adversidades de la vida, siempre se puede reír y que la formación académica no limita tus habilidades y capacidades de aprendizaje.

A Bernarda Tapia, mi madre, por ser el soporte principal en mi vida, entenderme, cobijarme y estar siempre presente en lo que necesito. Por escucharme sin entender nada de lo que le hablo, por aconsejarme cuándo se lo pido, pero por sobre todo, soportarme en mis peores momentos de estrés y fatiga académica.

También, a Juan Castillo, mi padre, quién lo único que siempre me ha pedido es llegar más lejos de lo que él ha podido, gracias por creer en mí sin dudar, por alimentar mi autoestima cuando lo necesité, pero aún más, por enseñarme los principios que hoy forman mi carácter y me han permitido llegar a mis propios finales.

Agradecer también, a María Paz Morales, por ser quien me enseñó y demostró que uno puede ser un factor de cambio en la vida de los demás, por acompañarme durante todos estos años universitarios, aconsejarme y ser la fuente de motivación principal en esta dura y compleja etapa. Gracias por enseñarme a buscar lo que hay dentro de mí, ayudarme a crecer como persona y como profesional, y sobre todo, gracias por permitirme contar contigo estos años, y los que vienen.

Por último, quiero agradecer a todos mis amigos que estuvieron presentes en de distintas formas en esta etapa, no necesariamente de forma presencial, sino también acompañándome a la distancia, y aún así ser un espacio de distracción y confianza.

RESUMEN

Resumen— A mediados del 2022, se realizó el plebiscito de salida de un nuevo proceso constituyente, donde se votó la aceptación del trabajo realizado por la Convención Constitucional. Instituciones como la Biblioteca del Congreso Nacional crearon plataformas exclusivas para el proceso, pero estas no integran información de fuentes externas e históricas sobre procesos similares. Con el objetivo de generar visualizaciones y análisis de datos que permitan informar a la población, se propone la definición, diseño e integración de datos de la Convención Constitucional con Comparative Constitute Project siguiendo recomendaciones estándar propuestas por la W3C e implementado una ontología de grafos RDF.

Palabras Clave— Convención Constitucional; Web Semántica; Integración de datos; Comparative Constitute Project.

ABSTRACT

Abstract— In mid 2022, the exit plebiscite of a new constitutional process was held, where the acceptance of the work done by the Constitutional Convention was voted. Institutions such as the Biblioteca del Congreso Nacional created exclusive platforms for the process, but these do not integrate information from external and historical sources on similar processes. With the objective of generating visualizations and data analysis to inform the population, is proposed the definition, design and integration of data from the Constitutional Convention with the Comparative Constitute Project following standard recommendations proposed by the W3C and implementing an RDF graph ontology.

Keywords— Constitutional Convention; Semantic Web; Data integration; Comparative Constitute Project.

GLOSARIO

- **BCN:** Biblioteca del Congreso Nacional.
- **CC:** Convención Constitucional.
- **CCP:** Comparative Constitute Project.
- **FAIR:** Findable, Accessible, Interoperable and Reusable.
- **IMFD:** Instituto Milenio Fundamento de los Datos.
- **IRI:** Internationalized Resource Identifier.
- **ISBN:** International Standar Book Number.
- **ISO:** International Organization for Standardization.
- **RDF:** Resource Description Framework.
- **OWL:** Web Ontology Language.
- **URI:** Uniform Resource Identifier.
- **URL:** Uniform Resource Locator.
- **URN:** Uniform Resource Name.
- **UTFSM:** Universidad Técnica Federico Santa María.
- **W3C:** World Wide Web Consortium.
- **XML:** Extensible Markup Language.

ÍNDICE DE CONTENIDOS

| | |
|---|----------|
| RESUMEN | IV |
| ABSTRACT | IV |
| GLOSARIO | V |
| ÍNDICE DE FIGURAS | VIII |
| ÍNDICE DE TABLAS | XI |
| INTRODUCCIÓN | 1 |
| CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA | 2 |
| 1.1 Plataformas en torno a la Convención Constitucional | 2 |
| 1.1.1 Biblioteca del Congreso Nacional | 2 |
| 1.1.2 Plataforma Telar | 3 |
| 1.2 Definición del Problema | 4 |
| 1.3 Solución Propuesta | 5 |
| 1.4 Objetivos | 6 |
| 1.4.1 Objetivo General | 6 |
| 1.4.2 Objetivos Específicos | 6 |
| 1.5 Recapitulación de la Problemática | 6 |
| CAPÍTULO 2: MARCO CONCEPTUAL | 8 |
| 2.1 Principios de datos FAIR | 8 |
| 2.1.1 <i>Findable / Encontrable</i> | 8 |
| 2.1.2 <i>Accessible / Accesible</i> | 8 |
| 2.1.3 <i>Interoperable / Interoperabilidad</i> | 9 |
| 2.1.4 <i>Reusable / Reutilizable</i> | 9 |
| 2.2 Definición de Ontología | 9 |
| 2.2.1 Ontología Filosófica | 10 |
| 2.2.2 Ontología de dominio | 10 |
| 2.2.3 Ontología formal | 10 |
| 2.2.4 Ontología aplicada | 10 |
| 2.3 Web Semántica | 11 |
| 2.3.1 Resource Description Framework | 11 |
| 2.3.2 URIs, URLs, URNs e IRIs | 12 |
| 2.3.3 Vocabularios dentro de RDF | 13 |
| 2.4 SPARQL | 14 |
| 2.4.1 Definición | 14 |
| 2.5 Base de Datos Relacional a Grafo RDF | 15 |
| 2.5.1 Patrones de mapeo de tablas: | 16 |

| | |
|--|-----------|
| 2.5.2 Patrones de mapeo de atributos: | 16 |
| 2.6 Faceted Browsing | 18 |
| CAPÍTULO 3: PROPUESTA DE SOLUCIÓN | 20 |
| 3.1 Análisis preliminar de datos | 20 |
| 3.1.1 Análisis a los datos de la BCN | 20 |
| 3.1.2 Análisis a los datos de CCP | 25 |
| 3.2 Planteamiento de la Ontología | 27 |
| 3.2.1 Estructura para los datos CCP | 28 |
| 3.2.2 Estructura de integración | 31 |
| 3.3 Procesamiento de datos | 32 |
| 3.3.1 Datos de Países | 33 |
| 3.3.2 Datos de Constituciones | 34 |
| 3.3.3 Características de constituciones | 35 |
| 3.3.4 Integración | 37 |
| 3.4 Sistema de consultas SPARQL | 38 |
| 3.5 Creación de Visualizaciones | 39 |
| CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN | 41 |
| 4.1 Cumplimiento del estándar | 41 |
| 4.2 Consultas con técnicas Facetadas | 41 |
| 4.3 Visualizaciones | 43 |
| 4.3.1 Convención Constituyente | 43 |
| 4.3.2 Comparative Constitution Project | 47 |
| 4.3.3 Integración | 52 |
| CAPÍTULO 5: CONCLUSIONES | 57 |
| 5.1 Trabajo Futuro | 59 |
| 5.2 Reflexiones | 59 |
| ANEXOS | 60 |
| A Datos de la Biblioteca del Congreso Nacional | 60 |
| B Datos de Comparative Constitution Project | 63 |
| C Estructura de integración | 66 |
| D Procesamiento de datos | 69 |
| REFERENCIAS BIBLIOGRÁFICAS | 73 |

ÍNDICE DE FIGURAS

| | | |
|----|--|----|
| 1 | Árbol del problema. | 7 |
| 2 | Ejemplo de representación gráfica de un <i>triple</i> | 11 |
| 3 | Ejemplo de representación gráfica de un grafo RDF con 2 <i>triples</i> | 12 |
| 4 | Partes de una consulta SPARQL. | 15 |
| 5 | Filtros según facetas en Amazon e Ebay respectivamente. | 19 |
| 6 | Filtros según facetas en Solotodo, en donde se indican la cantidad de coincidencias. | 19 |
| 7 | Representación gráfica del primer subgrafo. | 23 |
| 8 | Representación gráfica del segundo subgrafo. | 24 |
| 9 | Representación gráfica del subgrafo de tópicos CCP mostrando sólo el primer nivel de profundidad. En la Figura 48 en los Anexos B se muestra la representación completa. | 28 |
| 10 | Representación gráfica del subgrafo de países y grupos propuesto. | 30 |
| 11 | Representación gráfica del subgrafo de eventos propuesto para dar soporte a los datos base de una constitución. | 31 |
| 12 | Estructura general de los datos del archivo <i>countries.xml</i> | 33 |
| 13 | Estructura general de los datos del archivo <i>countries.ttl</i> | 34 |
| 14 | Estructura general de los datos del archivo <i>chronology.xml</i> | 34 |
| 15 | Estructura general de los datos del archivo <i>eventos.ttl</i> | 35 |
| 16 | Estructura general de los datos del archivo <i>datosDuros.ttl</i> | 36 |
| 17 | Estructura general de los datos del archivo <i>integracion.ttl</i> | 37 |
| 18 | Carga de estructura y datos en la aplicación <i>Jena Fuseki</i> | 38 |
| 19 | Ejemplo de consulta SPARQL en la aplicación <i>Jena Fuseki</i> | 39 |
| 20 | Consulta SPARQL para la obtención de cantidad de constituyentes por partido político. | 40 |

| | | |
|----|--|----|
| 21 | Código en Python para el procesamiento y visualización de cantidad de constituyentes por partido político. | 40 |
| 22 | Porcentaje de constituyentes por partido político. | 40 |
| 23 | Definición de prefijos para ejemplo de búsqueda facetada. | 42 |
| 24 | Primera consulta para ejemplo de búsqueda facetada. | 42 |
| 25 | Segunda consulta para ejemplo de búsqueda facetada. | 43 |
| 26 | Prefijos utilizados para las consultas de validación. | 43 |
| 27 | Consulta SPARQL para obtener las edades de todo los candidatos a la CC. . . . | 44 |
| 28 | Distribución de edades para Candidatos y Constituyentes de la CC. | 44 |
| 29 | Consulta SPARQL para obtener el Top 20 de tópicos hablados en los discursos de la CC. | 45 |
| 30 | Cantidad de apariciones de tópicos en discursos de la CC. | 45 |
| 31 | Consulta SPARQL para obtener la cantidad de discursos según tópico y fecha en la CC. | 46 |
| 32 | Cantidad de discursos totales y con tópico <i>Pueblos Originarios</i> a lo largo del tiempo en al CC. | 46 |
| 33 | Consulta SPARQL para obtener el año de la última constitución de cada país. Se incluye un comentario en caso de que se quiera considerar las enmiendas y reformas. | 47 |
| 34 | Año de la creación de la última constitución a nivel mundial. | 48 |
| 35 | Año de la última actualización de constitución a nivel mundial. | 48 |
| 36 | Consulta SPARQL para obtener los 10 tópicos más recurrentes según región del mundo. Se considera la última actualización de constitución de cada país. . | 49 |
| 37 | Top 10 tópicos más recurrentes en las constituciones actuales de América. . . . | 50 |
| 38 | Top 10 tópicos más recurrentes en las constituciones actuales de Europa. . . . | 50 |
| 39 | Consulta SPARQL para obtener los eventos constitucionales en América del Sur desde el año 2000. | 51 |
| 40 | Cantidad de eventos constitucionales en América del Sur desde el año 2000. . . | 51 |

| | | |
|----|--|----|
| 41 | Consulta SPARQL para obtener los países que actualmente implementan el tópico <i>Pueblos Originarios</i> en sus constituciones. | 52 |
| 42 | Países que actualmente implementan el tópico <i>Pueblos Originarios</i> en sus constituciones. | 53 |
| 43 | Consulta SPARQL para obtener la cantidad de apariciones de los tópicos de la CC en las constituciones actuales del mundo. | 54 |
| 44 | Cantidad de apariciones de los tópicos de la CC en las constituciones actuales del mundo. | 54 |
| 45 | Consulta SPARQL para obtener la cantidad de tópicos de la CC relacionados a la constitución actual de cada país en una región en específico. | 55 |
| 46 | Distribución de la cantidad de tópicos de la CC relacionados a la constitución actual de cada país en Europa y América. | 56 |
| 47 | Representación gráfica de la ontología utilizada para almacenar los datos de la BCN. | 60 |
| 48 | Representación gráfica de la estructura del subgrafo de los tópicos de CCP. | 65 |
| 49 | Representación gráfica de la ontología propuesta. Se consideró sólo un nivel de profundidad para los tópicos CCP. | 68 |
| 50 | Pseudocódigo utilizado para la extracción de datos de los archivos <i>countries.xml</i> y <i>chronology.xml</i> | 69 |
| 51 | Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 1 en el archivo <i>ccpcnc_v4.csv</i> | 69 |
| 52 | Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 2 en el archivo <i>ccpcnc_v4.csv</i> | 70 |
| 53 | Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 3 caso 1 en el archivo <i>ccpcnc_v4.csv</i> | 70 |
| 54 | Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 3 caso 2 en el archivo <i>ccpcnc_v4.csv</i> | 71 |
| 55 | Pseudocódigo utilizado para la creación de declaraciones de integración a partir del archivo <i>datosDuros.ttl</i> | 72 |

ÍNDICE DE TABLAS

| | | |
|---|---|----|
| 1 | Taxonomía de los tópicos identificados en los discursos de la CC. Fuente: elaboración propia. | 61 |
| 2 | Taxonomía de los tópicos de CCP con dos niveles de profundidad. Fuente: elaboración propia. | 63 |
| 3 | Homologación de tópicos en el que el tópico CCP es igual o más general que el tópico BCN. Fuente: elaboración propia. | 66 |
| 4 | Homologación de tópicos en el que el tópico BCN es más general que el tópico CCP. Fuente: elaboración propia. | 67 |

INTRODUCCIÓN

El crecimiento de los datos de la web aumenta de forma exponencial, las estimaciones indican que el 90% de estos datos se han creado en los últimos años, y que sólo en 2020 se crearon más de 64 zettabytes de datos [Nations, 2021]. Esto plantea desafíos no sólo de almacenamiento, sino también de análisis e integración de estos datos, desafíos que de ser resueltos resultarían en un crecimiento de conocimientos útiles, que aplican a todo ámbito [Tascón, 2013]. Sumado a esto, no se cuenta con tecnologías y estructuras necesarias que permitan el análisis de los datos, y las que existen, son aplicables sólo a contextos en específico, dificultando la reutilización [Borja y Pérez, 2019]. Esto impacta directamente en la toma de decisiones a nivel empresarial, y a nivel persona, tanto en decisiones de negocio, como en decisiones políticas y sociales.

Uno de estos casos es el proceso Constitucional realizado en Chile entre 2021 y 2022, que por primera vez en la historia del país se desarrolló bajo una Convención Constitucional con integrantes escogidos por votación popular [ChileAtiende, 2022]. Este proceso contó con el desarrollo de varias plataformas de seguimiento y entrega de información oficiales, como las desarrolladas por la Biblioteca del Congreso Nacional ¹. Pero no contó con una integración con fuentes internacionales que permitan analizar desde diferentes puntos de vista este proceso, y así entregar al ciudadano información útil para la evaluación y la toma de decisiones en las próximas votaciones relacionadas.

En este trabajo se plantea una estructura capaz de integrar los datos provenientes de la Biblioteca del Congreso Nacional sobre la Convención Constitucional, con *Constitute Comparative Project*, proyecto internacional que funciona como repositorio de todas las constituciones del mundo, permitiendo realizar búsquedas históricas y comparaciones a nivel mundial [Elkins y Ginsburg, 2022]. Para este propósito se implementa la Web Semántica, que permite, además de almacenar e identificar recursos web, darle fácilmente significado semántico a los datos [Berners-Lee *et al.*, 2001], con el objetivo final de permitir la consulta y construcción de visualizaciones sobre la integración, y así entregar conocimiento no trivial al alcance del ciudadano.

La estructura del documento se basa en los siguientes capítulos: Capítulo 1, en este capítulo se da contexto en detalle el problema presentado, además de introducir la solución propuesta, los objetivos y alcances de esta. En el Capítulo 2, se especifican y definen formalmente las metodologías y estándares a seguir en la construcción de la solución. En el Capítulo 3, se analizan las fuentes de datos y se construye la ontología a implementar, además de explicar los procesos de extracción, carga, consulta y visualización de los datos. Luego, en el Capítulo 4, se verifica el cumplimiento de estándares, metodologías propuestas y capacidad de consulta sobre la ontología propuesta. Por último, en el Capítulo 5, se realiza un análisis general del trabajo realizado, y se proponen mejoras de actualización de datos y nuevas integraciones para acercar aún más esta integración al usuario común.

¹Véase en <https://www.bcn.cl/porta1/>

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

En nuestra historia como país, nos hemos vistos envueltos en varios procesos constituyentes con el objetivo de redactar la carta fundamental que nos rige como nación [BCN, 2020]. Por ejemplo, el proceso constituyente para la elaboración de la constitución de 1980, que nos rige actualmente, se basó en la formación de un agente externo al gobierno, en específico, una comisión de expertos conformada por abogados y profesionales expertos en las materias de derechos constitucionales o ciencias sociales. Por otra parte, en 2019, en consecuencia de manifestaciones ciudadanas, se inició una serie de acuerdos y reformas constitucionales, y el 25 de octubre de 2020 se celebró un plebiscito histórico en Chile, dando pie a un proceso nunca antes visto en el país, la elaboración de una nueva constitución mediante el mecanismo de Asamblea Constituyente [ChileAtiende, 2022], donde el órgano encargado es conformado por ciudadanos escogidos por sufragio popular, quienes en representación del resto de la nación se encargan de discutir y escribir la nueva constitución, documento que luego debe ser aprobado mediante un plebiscito de salida.

Bajo este contexto, se crearon plataformas y servicios para acompañar el proceso, ya sean repositorios para unificar los documentos oficiales de la Convención Constitucional (CC), o análisis de datos de fuentes externas de todo lo que rodea al proceso. A continuación, en la subsección 1.1 se describen las principales plataformas de datos en torno a la CC, en la subsección 1.2 se define el problema que estas presentan. Luego, en la subsección 1.3 se define el impacto del problema y se introduce la solución propuesta. En la subsección 1.4 se presentan los objetivos generales y específicos de esta memoria. Finalmente, en la subsección 1.5 se realiza una recapitulación de la problemática.

1.1. Plataformas en torno a la Convención Constitucional

Entre las plataformas entorno a la CC se destacan principalmente 2: la Biblioteca del Congreso Nacional y la Plataforma Telar. Ambas ofrecen datos del proceso en base a distintos enfoques. A continuación, se presenta la descripción de ambas, resaltando sus principales características.

1.1.1. Biblioteca del Congreso Nacional

La Biblioteca del Congreso Nacional de Chile (BCN) es un servicio común del poder Legislativo, principalmente de las cámaras que componen el congreso. Su misión es apoyar a los integrantes del parlamento en su función, a través de la generación de información, conocimiento y asesorías. Así como también, liderar instancias de vinculación con la ciudadanía,

poniendo a disposición el repositorio bibliográfico, documental y jurídico de la historia política legislativa del país [BCN, 2014].

En este contexto, la BCN desarrolló una plataforma que pone a disposición los documentos oficiales de la CC, como el histórico de la sesiones, los hitos importantes ocurridos, las temáticas sociales abordadas y los discursos realizados por los distintos integrantes de esta ². Entre las herramientas que ofrece se encuentran un lector online de la constitución vigente de 1980, metodologías, reglamento y documentos de los encuentros auto convocados y cabildos celebrados en 2016, una línea de tiempo para visualizar el avance del proceso, un repositorio con las noticias en medios nacionales oficiales sobre el proceso, videos explicativos y registros visuales de distintos eventos, entre otros. Aún así, una de las principales herramientas dispuestas es el comparador de constituciones. Que permite revisar las constituciones vigentes de 194 países alrededor del mundo, además de un apartado para identificar cómo están definidas ciertas temáticas en estas constituciones, como por ejemplo protección al medio ambiente y la naturaleza o derecho a la vivienda, entre otras ³.

Esta herramienta fue construida en alianza con Constitute Project, una plataforma internacional de comparación de constituciones que ofrece un repositorio histórico de las constituciones de cada país. Su estructura permite comparar las constituciones actuales, anteriores, e incluso borradores, de todos los países independientes del mundo, a través de una clasificación por tópicos ⁴.

1.1.2. Plataforma Telar

El Instituto Milenio de Investigación sobre los Fundamentos de los Datos (IMFD) es un centro multidisciplinario en América Latina que reúne a expertos del área de la informática, matemáticas, estadísticas, ciencias políticas y comunicación, con el objetivo de hacer frontera en ciencia de datos y así hacer frente a las distintas problemáticas que existen hoy en estos campos [Hevia, 2020].

La comunidad de este instituto vio en la CC una oportunidad de poner todos sus conocimientos al servicio de la sociedad. Así es como nace la Plataforma Telar, una plataforma que se centra en la extracción, transformación y análisis de datos de distintas fuentes que rodean a la CC, basándose principalmente en las RRSS, traduciendo sus resultados en la presentación de visualizaciones estadísticas para acercarle la información a la ciudadanía [PUC, 2021].

Entre los principales análisis presentados por la plataforma, se destaca la integración con redes sociales, noticias de los medios y por sobre todo los resultados de encuestas a la ciudadanía sobre la percepción que estos han tenido sobre la CC ⁵.

²Véase en <https://www.bcn.cl/procesoconstituyente/index.html>

³Véase en <https://www.bcn.cl/procesoconstituyente/comparadordeconstituciones/home>

⁴Véase en <https://www.constituteproject.org/content/about?lang=es>

⁵Véase en <https://www.cnnchile.com/tag/plataforma-telar/page/1/>

1.2. Definición del Problema

El comparador de constituciones, disponible en la plataforma de la BCN, está basado en la estructura planteada por *Constitute Project*, pero no implementa todas sus funcionalidades. La herramienta de la BCN sólo considera las constituciones y reformas vigentes de cada país, dejando de lado toda la información histórica en este ámbito, entre ella los datos de cuantos procesos constituyentes han sido desarrollado en cada país, cómo han evolucionado las distintas constituciones, qué constituciones fueron restauradas, entre otros. Además, cuenta con una limitada visualización de los datos en cuanto a este mismo contexto, sólo permitiendo visualizar qué países abordan cierto tema, impidiendo aplicar más filtros. Por último, esta herramienta, a pesar de ser desarrollada por la BCN, no está integrada a los demás datos que ofrece, como los discursos en las sesiones de la CC, o la asociación de las temáticas planteadas dentro de la CC con las constituciones internacionales, dificultando el análisis en profundidad de los datos generados por este proceso.

En cuanto a la Plataforma Telar, sufre de la misma problemática, si bien entrega más visualizaciones en base a fuentes externas, estas no están relacionadas en su totalidad con las temáticas abordadas dentro de la CC y no permiten hacer una comparación en profundidad de estas con fuentes externas e históricas. Pero, sí entrega información de los candidatos y Constituyentes que participaron del proceso. Entregando datos desde un enfoque interesante a ser considerado.

Por otra parte, la plataforma de *Constitute Project* nos entrega datos de todas las constituciones del mundo, pero presenta limitadas visualizaciones que permiten un análisis visual de los datos entregados, dificultando el entendimiento de estos con otros puntos de vista. Además, esta plataforma no cuenta con datos actualizados del proceso constitucional, que le permita relacionar temáticas presentes en las constituciones del mundo con las temáticas abordadas dentro de la CC.

A modo de resumen, el problema que se presenta en este contexto es la falta de un mecanismo que integre los datos sobre la CC con fuentes externas como lo es la plataforma *Constitute Project*, y permita relacionar lo hablado dentro de la CC con fuentes históricas. Además, la falta de instrumentos capaces de generar visualizaciones a partir de esta integración de datos, visualizaciones que permitan realizar un análisis visual y más profundo. La complejidad principal del problema recae en la falta de instrumentos que extraigan y procesen los datos, y la falta de un diseño estandarizado que permita la integración. Pues dado que el proceso se desarrolló por primera vez en Chile, aún no se han creado y diseñado los mecanismos necesarios. Por último, el gran volumen de datos actuales, y que aumentarán a medida que avancen los nuevos procesos, impiden la extracción e integración manual de estos.

1.3. Solución Propuesta

La falta de una solución a este problema impacta directamente en los ciudadanos, dado que son ellos los que al final del proceso deben aceptar o rechazar la nueva constitución, y para ello necesitan estar informados a fondo de qué es lo que ocurre dentro de la CC y evaluar todas las aristas posibles. Estudios realizados en los últimos años, muestran que en promedio más del 50 % de la ciudadanía vota desinformada⁶, y estudios recientes indican que las fuentes principales de incidencia en los votos son los debates, redes sociales y programas políticos de televisión⁷, indicando que las fuentes de información más influyentes son visuales. Esto, hace necesaria la construcción de visualizaciones al alcance y entendimiento de todos abordando la mayor cantidad de aristas posibles.

Dado el gran porcentaje de votantes desinformados, varios organismos han iniciado campañas y programas para incentivar y difundir la información de las distintas elecciones⁸, pero estas han tenido poca participación ciudadana resultando en un bajo impacto según últimos estudios realizados⁹. Es decir, los bajos índices de participación y los altos índices de desinformación son problemas latentes en la sociedad y se han abordado desde varias perspectivas, siendo la entrega de material visual la más efectiva.

Por lo tanto, la solución al problema, a modo general, requiere de la creación de instrumentos de extracción de grandes volúmenes de datos, diseño de estructuras de orden y clasificación, y una integración que permita la construcción de visualizaciones. Para esto, se propone:

1. La construcción de instrumentos que sigan principios de extracción e integración estándar.
2. El diseño de una estructura que permita realizar una integración entre la Biblioteca del Congreso Nacional y *Constitute Project*.
3. Construcción de visualizaciones, como resultado de la integración, para un posterior análisis tanto de las fuentes externas como de la integración en sí.

⁶Datos basados en las publicaciones de La Tercera en 2016, T13 en 2021 y estudio TuInfluyes de marzo de 2021

⁷Dato obtenido del estudio TuInfluyes realizado en octubre de 2021

⁸Programa por un voto informado de la Universidad Católica en octubre de 2021., y distintos llamados del Serval a ser partícipes de las elecciones.

⁹Estudios realizados por TuInfluyes accesibles en <https://www.tuinfluyes.com/data-influye/>

1.4. Objetivos

En base al problema presentado, y a la solución propuesta, a continuación se presentan los objetivos de esta memoria.

1.4.1. Objetivo General

Definir, diseñar e implementar la integración de datos que generó la Convención Constitucional con fuentes externas, permitiendo así realizar un análisis en profundidad del proceso a través de visualizaciones.

1.4.2. Objetivos Específicos

- Diseñar una ontología de estructura y comunicación para los datos provistos por la Biblioteca del Congreso Nacional sobre la Convención Constitucional.
- Integrar datos de la Convención Constitucional con los datos de la plataforma *Constitute Project* a través de un grafo RDF.
- Generar visualizaciones a partir de la integración con técnicas de *Faceted Browsing* que permitan realizar un análisis en profundidad.

1.5. Recapitulación de la Problemática

A continuación, en la Figura 1 se presenta un resumen de la problemática planteada y los objetivos de esta memoria.

En resumen, el problema principal a abordar en esta memoria es la falta de un mecanismo que permita la integración de datos de la CC con fuentes externas. Las causas de este problema se basan en 2 tópicos: El volumen de datos, que impide realizar una integración manual, y lo reciente que es el proceso, que dificulta la integración ya que aún no hay mecanismos para esto y las plataformas actuales no está diseñadas para este propósito. Los efectos de este problema, es un aumento en la falta de información al alcance de la población encargada de decidir sobre el resultado del proceso, pues no es posible realizar comparaciones de los datos de la CC con fuentes históricas, y esto a su vez, imposibilita la creación de material visual respecto a este tema dificultando el análisis y la entrega de información.

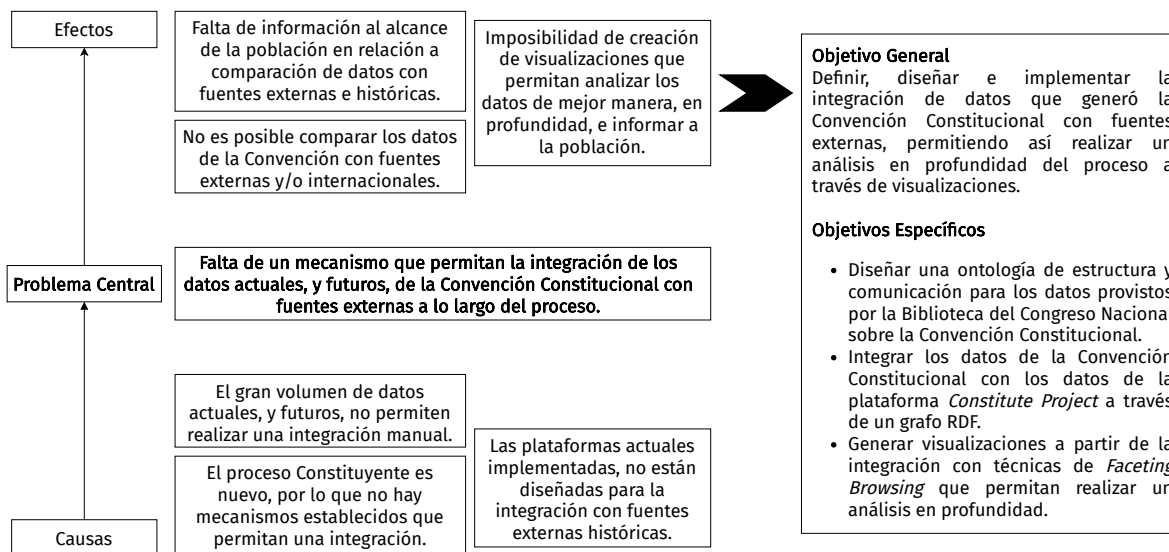


Figura 1: Árbol del problema.

Fuente: elaboración propia.

La desinformación es un problema latente en nuestra sociedad, y una de las principales causas es la dificultad en el acceso a las distintas fuentes de datos y a cómo éstas se relacionan entre sí, sobre todo, cuando se trata de grandes volúmenes de datos que impiden una integración manual o a simple vista. Es decir, se vuelve completamente necesaria la creación de mecanismos de integración y análisis de Big Data, y por sobre todo, la creación de estructuras robustas que permitan realizar estos procesos de forma sencilla. Con estas integraciones, no sólo se busca acercar los datos a la población, sino también, tener un espectro general de cómo se han planteado las distintas temáticas a nivel histórico en el resto del mundo, y así, en un mundo cada vez más globalizado, ser capaces de evaluar, analizar y entender procesos, nacionales e internacionales, teniendo conocimiento de todas las aristas posibles.

CAPÍTULO 2

MARCO CONCEPTUAL

2.1. Principios de datos FAIR

Uno de los principales desafíos a la hora de integrar fuentes de datos es cómo obtener y procesar estos datos, sobre todo porque no siguen un estándar establecido para su almacenamiento y disposición. Con el objetivo de construir una estructura que disminuya la complejidad de este desafío en integraciones futuras se decide seguir el estándar de datos FAIR.

Los principios de datos FAIR proponen un estándar de almacenamiento, construcción y manejo de datos y algoritmos que permite la reutilización e integración, en todas sus etapas, de la extracción de información. Estos principios son aplicables, independiente del dominio, a un gran número de tareas en la recolección y procesamiento de los datos y metadatos [Gvishiani *et al.*, 2021]. Los lineamientos de los principios son detallados a continuación:

2.1.1. *Findable / Encontrable*

- **F1.** A los (meta)datos se les asigna un identificador persistente y único global.
- **F2.** Los datos se describen con metadatos enriquecidos.
- **F3.** Los metadatos incluyen de forma clara y explícita el identificador de los datos que describen.
- **F4.** Los (meta)datos están registrados o indexados en un recurso de búsqueda.

2.1.2. *Accessible / Accesible*

- **A1.** Los (meta)datos son recuperables por su identificador utilizando un protocolo de comunicaciones estandarizado.
 - **A1.1** El protocolo es abierto, gratuito y universal.
 - **A1.2** El protocolo permite un procedimiento de autenticación y autorización, cuando sea necesario.
- **A2.** Los metadatos son accesibles, incluso cuando los datos ya no están disponibles.

2.1.3. *Interoperable / Interoperabilidad*

- I1. Los (meta)datos utilizan un lenguaje formal, accesible, compartido y de amplia aplicación para la representación del conocimiento.
- I2. Los (meta)datos utilizan vocabularios que siguen los principios FAIR.
- I3. Los (meta)datos incluyen referencias calificadas a otros (meta)datos.

2.1.4. *Reusable / Reutilizable*

- R1. Los (meta)datos se describen detalladamente con una multitud de atributos precisos y relevantes.
- R1.1 Los (meta)datos se publican con una licencia de uso de datos clara y accesible.
- R1.2 Los (meta)datos están asociados con la procedencia detallada.
- R1.3 Los (meta)datos cumplen con los estándares de la comunidad del dominio concreto.

Estos principios no proponen tecnologías o implementaciones estándar de ningún tipo. Sólo proveen lineamientos generales para las personas y/u organizaciones interesadas en publicar y entregar acceso a sus tecnologías y métodos que permiten recuperar, almacenar y reutilizar datos. Muchas de las herramientas y/o repositorios científicos actuales implementan todos o parte de los principios utilizando diferentes tecnologías [Gvishiani *et al.*, 2021].

2.2. Definición de Ontología

La plataforma *Constitute Project* basa su ontología de ordenamiento en un grafo que, a modo general, permite la clasificación de las constituciones en distintos tópicos, ontología que la BCN usó como base para el desarrollo de su plataforma de comparación de constituciones. Esta ontología define la estructura de un grafo, por lo tanto, es relevante estudiar cómo lograr una integración de datos mediante grafos. Pero antes, es importante definir qué es una ontología.

Veléz León en 2015 realizó una investigación exhaustiva acerca de las distintas definiciones que se pueden encontrar en la literatura acerca de las ontologías [León, 2015]. Entre todas ellas, logró agrupar y definir cuatro tipos de ontologías: ontología filosófica, ontología de dominio, ontología formal y ontología aplicada. A continuación, se dará una breve descripción de cada una de ellas:

2.2.1. Ontología Filosófica

Definición más tradicional de las cuatro, y quizás la primera definición que se le dio al concepto. La ontología filosófica se centra en el estudio de las estructuras y clases de los objetos que observamos, analizando sus propiedad, acontecimientos, procesos y las distintas relaciones en la realidad. En este contexto se puede considerar “ontoloigía” como sinónimo de “metafísica”.

2.2.2. Ontología de dominio

A diferencia de la ontología filosófica, donde la base de análisis es todo lo que observamos y la realidad en sí, la ontología de dominio parte de la base de un dominio reducido de la realidad, como lo pueden ser las artes, la ciencia, las leyes, etc. Por lo tanto, una ontología de dominio tiene un vocabulario controlado y estructurado, con el objetivo que permita a los humanos investigar, anotar y detallar información que posteriormente será procesada por sistemas informáticos. La ontología de dominio se ve favorecida por la ontología formal, ya que esta última se centra en hacer posible la comunicación dentro de un dominio, y así proveer un lenguaje común y una estructura formal que permita el razonamiento.

2.2.3. Ontología formal

Tal como se mencionó, la ontología formal es una disciplina que permite la comunicación externa e interna de un dominio, entregando un lenguaje estándar y una estructura formal que permita el razonamiento. Esta comunicación se hace posible al clasificar las entidades en distintas categorías y al describir qué relaciones hay entre categorías, intra categorías y por medio de otras categorías. Este tipo de ontología se aplica últimamente en la bio-informática, análisis de inteligencia, ciencia administrativa, entre otros campos de la ciencia y el comercio, donde es usa como base para la clasificación, organización y razonamiento automático.

2.2.4. Ontología aplicada

Es el tipo de ontología más complejo de definir, pues no hay un consenso entre las áreas del conocimiento sobre qué es una ontología. Las distintas definiciones dadas a lo largo de la literatura son construidas respecto a lo que se está trabajando y a lo que se busca. Así, hay definiciones según tamaño, estructura, conceptualización, grado de dependencia, entro otras características. Aún así, todas coinciden en considerar a una ontología como un sistema de organización del conocimiento.

Para propósito de esta memoria, y siguiendo la definición de ontología formal, se entien-

de que una ontología es la caracterización del ordenamiento y clasificación, que permite la comunicación y la relación dentro de un dominio.

2.3. Web Semántica

En un principio el desarrollo de la *World Wide Web* fue pensado para el consumo humano, por lo tanto el significado de la información sólo era posible ser procesada por seres pensantes [Allemang y Hendler, 2008]. En estas últimas décadas el aumento de la cantidad de datos ha hecho casi imposible el análisis y procesamiento por parte de los humanos de toda la información almacenada en la Web, siendo necesaria la intervención de sistemas automatizados. Con el fin de proveer estándares para la creación de ambientes de trabajo en donde aplicaciones, empresas y comunidades puedan acceder a los datos almacenados en la nube, la *World Wide Web Consortium* (W3C), define lo que hoy se conoce como Web Semántica [W3C, 1994], una estructura que permite la lectura de datos desde un software y la realización de las mismas tareas que un usuario común [Berners-Lee *et al.*, 2001].

Para el manejo de el gran volumen de datos, mantener la disponibilidad y permitir la comunicación con nuevas fuentes a futuro, la W3C propuso un único formato de estructuración de los datos, el *Resource Description Framework* (RDF) [Allemang y Hendler, 2008].

2.3.1. Resource Description Framework

Resource Description Framework es un método para representar fuentes Web, entregando un estándar común para expresar metadatos y así intercambiar datos entre aplicaciones sin pérdida de información, y permitiendo el procesamiento semántico sin la interacción humana [van der Ham *et al.*, 2006]. La estructura básica propuesta es el *triple*, que es la conexión de un **sujeto** (el recurso a ser descrito), la **propiedad** (característica del sujeto a ser descrita) y **objeto** (El valor de la propiedad según la declaración) [van der Ham *et al.*, 2006]. La unión de estas entidades se realiza considerando al sujeto y al objeto como nodos, y la propiedad como arista que los une. Por ejemplo, si quisiéramos representar una constitución y al país al que pertenece tendríamos el siguiente *triple*:

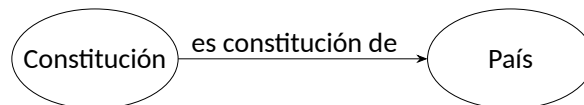


Figura 2: Ejemplo de representación gráfica de un *triple*.

Fuente: elaboración propia.

El objeto, en el ejemplo el País, no siempre es un dato, ya que también puede representar a otro sujeto, permitiendo así indicar propiedades de este. Por ejemplo, si quisiéramos indicar

el continente al que pertenece el País, tendríamos la siguiente estructura:

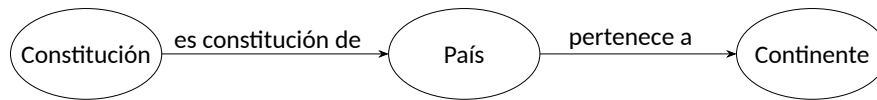


Figura 3: Ejemplo de representación gráfica de un grafo RDF con 2 *triples*.
Fuente: elaboración propia.

Así, la unión de *triples* al considerar un objeto como un nuevo sujeto, permite la construcción de grafos RDF, que con una gran cantidad de metadatos y sin una ontología clara pueden aumentar enormemente su complejidad [van der Ham *et al.*, 2006, Schewe y Thalheim, 2011]. Es decir, se vuelve fundamental la definición previa de una ontología de orden y comunicación antes de la integración de los datos, ya que de esta dependerá la estructura del grafo y la posterior extracción de información.

2.3.2. URIs, URLs, URNs e IRIs

Los objetos y sujetos dentro de un grafo RDF representan recursos únicos dentro de la web, por lo tanto, necesitan un identificador único que permita reconocerlos [Allemang y Hendler, 2008]. Para este propósito, la W3C recomienda el uso de las URIs. Una URI (Uniform Resource Identifier) es un *string* compacto de caracteres que permite identificar un recurso abstracto o físico [Berners-Lee *et al.*, 1998], por lo tanto, es útil para identificar toda entidad existente.

Las características principales de una URI están dadas por su propia definición:

- **Uniforme:** Al ser uniforme, permite la estandarización de la identificación de recursos en la web. Favoreciendo la identificación de recursos dentro de un mismo contexto, como también, la integración de recursos de contextos diferentes.
- **Recurso:** Un recurso puede ser todo lo que tenga identidad, incluyendo documentos, imágenes, archivos, servicios, colecciones de recursos, lugares, etc. Destacar que no todos los recursos pueden ser representados dentro de la web, como por ejemplo los libros en una biblioteca, o los seres humanos.
- **Identificador:** Un objeto que actúa como referencia a un elemento con identidad. En específico para las URI, este identificador tiene una sintaxis restringida (código ASCII).

Dentro de las URIs podemos identificar 2 tipos, no excluyentes entre sí:

- **URL:** Uniform Resource Locator, es el tipo de URI que se centra en identificar recursos según **cómo** se accede a ellos y **dónde** se encuentran. Si el recurso deja de estar disponible en dónde indica la URL, ya no podrá ser identificado [Berners-Lee *et al.*, 1998].

Por ejemplo, el URI “<https://www.chileconvencion.cl/wp-content/uploads/2022/07/Texto-Definitivo-CPR-2022-Tapas.pdf>” es un URL, dado que indica específicamente cómo obtener el recurso y dónde obtenerlo.

- **URN:** Uniform Resource Name, es el tipo de URI que se centra en identificar recursos según el nombre, o en muchos casos, un conjunto de caracteres alfanuméricos. Permite identificar el recurso independiente de dónde esté almacenado o cómo se accede a él.

Por ejemplo el URI “*ISBN:9789560015921*”, que hace referencia el mismo recurso del ejemplo anterior, es un URN, ya que entrega un identificador único permanente, dentro de un contexto específico, que se mantendrá independiente de dónde esté disponible el recurso [Berners-Lee *et al.*, 1994].

La sintaxis restringida en la definición de URIs, dado que sólo acepta codificación ASCII, presenta un problema cuando se desea utilizar este mecanismo de forma internacional, ya que idiomas como el Chino y el Japonés no son soportados por esta codificación. Es por ello, que en 2005, la W3C recomendó un nuevo estándar [Dürst y Suignard, 2005]. El IRI (Internationalized Resource Identifier) implementa una codificación UNICODE (ISO 10646). Así, toda URI es una IRI, pero no todas las IRIs son URIs y en estos casos es necesaria una normalización y así evitar problemas de intemporalidad.

2.3.3. Vocabularios dentro de RDF

Dado el gran volumen de datos, estructuras y sintaxis dentro de la web, se hizo necesaria la estandarización de conjuntos de datos para este propósito y así crear un vocabulario común. Entre las propuestas por la W3C, por ende entendidas como estándar, se destacan tres conjuntos de datos o vocabularios [Cyganiak *et al.*, 2014]:

- **rdf:** Contiene las definiciones estándar del vocabulario incorporado en RDF. Algunos ejemplos son la definición de *type*, *Property*, *List*, entre otros. Su IRI de acceso es: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
- **rdfs:** Contiene las definiciones estándar del vocabulario del esquema de RDF. Algunos ejemplos son al definición de *Class*, *subClassOf*, *label*, entre otros. Su IRI de acceso es: <http://www.w3.org/2000/01/rdf-schema#>.
- **xsd:** Contiene las definiciones estándar del vocabulario del tipo de datos compatibles con RDF. Algunos ejemplos son la definición de *string*, *integer*, *date*, entre otros. Su IRI de acceso es: <http://www.w3.org/2001/XMLSchema#>.

Dadas estas IRIs de vocabulario, es posible hacer referencia a las definiciones sin necesidad de indicar su IRI completa. Por ejemplo, si quisiéramos hacer referencia a la definición de *Class*, lo haríamos como *rdfs:Class* en vez de <http://www.w3.org/2000/01/rdf-schema#Class>.

2.4. SPARQL

Una vez establecida la estructura de *triples* y los grafos RDF como estándar en la construcción de estructuras semánticas en la web, surgió un nuevo desafío: cómo consultar los datos almacenados en estas estructuras. Esta discusión fue el foco de investigación a inicios de los años 2000, apareciendo varias soluciones, en las que se destacan: RQL (RDF Query Language), SeRQL (Sesame RDF Query Language), TRIPLE y RDQL (RDF Data Query Language). Todos distintos entre sí con sus ventajas y desventajas, tal como lo especificó Peter Heese en la conferencia internacional de la web semántica en 2004 [Haase *et al.*, 2004]. A su vez, la W3C también dedicó sus esfuerzos en la construcción de un lenguaje de consulta que permita aprovechar todas las propiedades de esta nueva estructura. Así, en 2004 dieron a conocer al público SPARQL [Zou, 2018, DuCharme, 2013], lenguaje de consultas RDF que rápidamente ganó popularidad, y el 15 de enero de 2008 fue publicada como recomendación oficial de la W3C [Prud'hommeaux y Seaborne, 2008], volviéndose el estándar para este propósito. A continuación, se definirá formalmente SPARQL y se identificarán sus principales características.

2.4.1. Definición

El nombre SPARQL es un acrónimo recursivo para *SPARQL Protocol and RDF Query Language* propuesto en 2004 por la W3C. Es un lenguaje estructurado destinado a consultar y manipular datos almacenados en formato RDF [Zou, 2018], y dado que RDF es una estructura de datos basada en un grafo dirigido, SPARQL es en esencia un lenguaje de consultas de patrones sobre grafos [Arenas *et al.*, 2019]. Estas consultas están divididas en tres partes:

- **Coincidencia de patrones:** Parte principal de la consulta, es aquí donde se ingresa el patrón de datos a consultar.
- **Modificación de respuesta:** Ya obtenida la respuesta al patrón ingresado, se le puede agregar una sentencia de orden, agrupamiento, búsqueda de valores distintos, etc.
- **Formato de salida:** Aquí es donde se seleccionan las variables que se desean mostrar y el formato de cómo se desean mostrar. Los formatos válidos para la última versión (SPARQL 1.1) son: XML, JSON, CSV y TSV.

Además, permite la definición de prefijos, para indicar vocabularios o conjuntos de datos y así facilitar las consultas. También permiten la definición del grafo RDF en el que se quiere ejecutar la consulta, esto es útil cuando se trabaja con más de un grafo RDF.

A modo de ejemplo, en la Figura 4¹⁰ se muestra una consulta indicando las partes anteriormente mencionadas:

¹⁰El resultado de la consulta es el nombre (en lenguaje natural) y el IRI de todos los recursos establecidos como `rdf:Class`, ordenados alfabéticamente.

```
# Prefijos (opcional)
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

# Formato de salida
SELECT ?nombre ?IRI
# Grafo RDF (opcional)
FROM <http://www.w3.org/People/Berners-Lee/card>
# Patron de coincidencias
WHERE {
    ?IRI rdf:type rdfs:Class;
        rdfs:label ?nombre .
}
# Modificaciones (opcional)
ORDER BY ?nombre
```

Figura 4: Partes de una consulta SPARQL.

Fuente: elaboración propia.

2.5. Base de Datos Relacional a Grafo RDF

Dado el gran volumen de datos con el que se trabajará, es necesario un sistema que permita almacenar los *triples* y llevarlos a su representación de grafo de forma eficaz y eficiente. Si bien hay sistemas NoSQL que solucionan este problema, según artículos empresariales, de opinión y de investigación [Sahatqija *et al.*, 2018, Anderson y Nicholson, 2022, Bartholomew, 2010], estos sistemas presentan grandes desafíos en su implementación y uso, ya que no siguen un estándar en específico y la implementación dependerá del contexto en donde se utiliza, lo que se aleja de los principios de datos FAIR que se persiguen en el desarrollo de esta memoria. Por el contrario, las bases de datos relacionales son ampliamente utilizadas, teniendo implementaciones estándar ya sea de almacenamiento o de consulta de datos, lo que facilita la disponibilidad y el acceso de los datos.

En este contexto, se necesitan herramientas estándares que permitan, en lo posible, automatizar el proceso de construcción del grafo luego de la definición de la ontología. Es decir, se necesitan herramientas y/o metodologías que permitan transformar una base de datos relacional en un grafo RDF. En la literatura se pueden encontrar varios patrones de mapeo, Juan Sequeda en 2012 [Sequeda *et al.*, 2012a] reunió en un documento los patrones más comunes, ya sea para mapear tablas de la base de dato como clases de la ontología, o atributos de las tablas como propiedades de un *triple*. Los distintos patrones reunidos por Sequeda se muestran a continuación:

2.5.1. Patrones de mapeo de tablas:

- **Uno a Uno:** Es el caso más básico con el que nos podemos encontrar. Se busca que cada instancia en la tabla de la base de datos sea una instancia de una clase de la ontología.
- **Uno a Muchos:** Este caso se da cuando en la ontología hay dos o más clases que se refieren al mismo sujeto, como por ejemplo un *estudiante* que es de la clase *estudiantes* y a la vez es de la clase *personas*. Por lo tanto, las instancias de la tabla serán instancias de ambas clases.
- **Muchos a Uno:** De forma similar al patrón anterior, pero en este caso son dos tablas en la base de datos que tienen el mismo significado semántico en una de las clases de la ontología. Como por ejemplo, *profesor* y *estudiante* que semánticamente son *personas*. Por lo tanto, todas las instancias de las tablas serán instancias de la misma clase.
- **Muchos a Muchos:** Este patrón es una combinación de los patrones uno a muchos y muchos a uno, en donde se generaliza y permite la transformación de varias instancias en 2 o más tablas de la base de datos en instancias de 2 o más clases de la ontología.

2.5.2. Patrones de mapeo de atributos:

- **Uno a Uno:** Al igual que en el mapeo de tablas, este es el caso más básico. Se busca mapear una de las columnas de la tabla en la base de datos a una de las propiedades de una clase en la ontología.
- **Uno a Muchos:** En este caso una de las columnas de una tabla en la base de datos tiene más de un significado semántico en la ontología, como por ejemplo el *apellido*, que puede ser considerado como *apellido* y como *nombre de la familia* del sujeto descrito.
- **Muchos a Uno:** Es un caso común dado que una propiedad en una clase de la ontología puede tener varios valores, por ejemplo las columnas *numero de celular* y *número telefónico* en una tabla de la base de datos, se transforman en dos instancias de la propiedad *número de contacto*.
- **Muchos a Muchos:** Al igual que en los patrones de mapeo de tablas, este es un caso combinado de uno a muchos y muchos a uno, en donde se desea transformar varias columnas de una tabla en una propiedad de la clase, y a la vez hay columnas de la tabla que tienen más de un significado en la ontología, por lo que se transforman en varias propiedades.
- **Concatenar propiedades:** Uno de los patrones más útiles a utilizar, en donde se busca unir dos columnas de la tabla y el resultado llevarlo a sólo una propiedad de la ontología. Por ejemplo, unir las columnas *nombre* y *apellido* de la tabla, para que el resultado se instancie en la propiedad *nombre completo* de la ontología.

No sólo debemos preocuparnos de la construcción de del grafo RDF, sino también en el lenguaje que permite definir el mapeo entre la base de datos y el grafo. Claus Stadler, en 2015, comparó los distintos lenguajes de mapeo y propuso un nuevo lenguaje, que simplifica la definición de las distintas instancias, y que es posible transformarla a lenguajes estándar y viceversa [Stadler *et al.*, 2015]. Así, también existen metodologías completas de transformación, que plantean un mapeo específico para cada tabla y tipo de relación, además de lenguajes de mapeo y lenguajes de consultas en los grafos RDF generados [Thuy *et al.*, 2014, Wardani y Kiing, 2014, W3C, 2012].

Entre los artículos de la literatura, destaca lo propuesto por W3C. En su metodología propuesta [W3C, 2012], introducen un mapeo exhaustivo para cada uno de los casos posibles, e incluso de cómo construir manualmente el grafo RDF, lo que permite hacer modificaciones manuales a la estructura sin necesidad de realizar el proceso nuevamente, o en caso de ser necesario, deshacer *triples* o subgrafos por completo. A partir de esta metodología, se han creado varias nuevas que buscan automatizar aún más el proceso, aplicar nuevas tecnologías, o mapear las bases de datos a otras estructuras, ejemplo de esto son los trabajos [Sequeda y Miranker, 2015, Thapa y Giese, 2021, Chen *et al.*, 2014, Sequeda *et al.*, 2012b].

Sequeda y Miranker crearon una herramienta semiautomática que agiliza el proceso. En su documento [Sequeda y Miranker, 2015] presentan un paso a paso de cómo utilizar la herramienta y explican cómo esta ayuda a realizar el proceso con la mínima interacción humana. Thapa y Giese, en 2021 propusieron una modificación a la metodología de W3C [Thapa y Giese, 2021], en la que buscan la reescritura de las reglas y restricciones de la base de datos en el grafo RDF disminuyendo el costoso trabajo de validación. Similar a lo que hicieron Y. Chen, Z. Xu, Y. Ni, G. Cao, y S. Zhang en 2014 [Chen *et al.*, 2014], quienes modificaron la metodología de W3C para orientar el mapeo desde un punto de vista más cercano a la base de datos, a través de la definición de nuevas reglas de mapeo y un nuevo prototipo de mapeo, el cuál probado con casos de estudio prueba que sí es posible su realización y es altamente efectivo. Por último, J. F. Sequeda, M. Arenas, y D. P. Miranker además de proponer una automatización, propusieron un mapeo al estándar OWL (Web Ontology Language), un lenguaje de marcado utilizado para publicar y compartir datos usando ontologías, que se centra en codificar los construido en RDF en un archivo XML (Extensible Markup Language) [Sequeda *et al.*, 2012b].

A pesar de los grandes beneficios que ofrecen estos trabajos al mejorar lo propuesto por W3C, su implementación implica la utilización de una ontología construida de forma automática, sin la opción de modificarla o de proponer una propia, lo que dificulta y limita mucho el trabajo a realizar. Es decir, estas modificaciones al estándar no son recomendables para el desarrollo del trabajo y objetivos propuestos.

2.6. Faceted Browsing

En los puntos anteriores ya se ha visto cómo los grafos RDF nos permite almacenar y estructurar los grandes volúmenes de datos presentes en la web, además de darles sentido semántico a las distintas relaciones entre los recursos. Con SPARQL se vio cómo recuperar estos datos desde los grafos RDF y las principales opciones que se tienen a la hora de realizar las consultas. Pero, como ya se planteó en un inicio, es fundamentalmente necesaria la definición de una ontología que facilite el planteamiento de patrones de búsqueda. Bajo este motivo, existen muchas opciones que permiten guiar la construcción de una ontología, pero pocas que permitan la búsqueda en base a las mismas propiedades y/o relaciones que presentan los recursos a buscar [Fagan, 2010].

La Navegación Facetada (Faceted Browsing) es una técnica de exploración de datos basada en facetas. La característica principal de esta técnica es permitir la aplicación de filtros de forma dinámica, permitiendo quitar o agregar estos filtros en el orden que se desee [Simonini y Zhu, 2015]. Esto permite obtener el mismo resultado desde distintos puntos de partida, y evitar la duplicidad de elementos en el universo de datos.

Esta técnica es común en sistemas de e-commerce para facilitar la búsqueda de los productos [Moreno-Vega y Hogan, 2018]. Ejemplo de esto son Amazon¹¹ e Ebay¹² (ver Figura 5). En estos casos, los filtros se actualizan de forma dinámica dependiendo de los productos que cumplen con los filtros ya aplicados. Incluso, en algunos sitios webs como Solotodo¹³ (ver Figura 6), indican la cantidad de productos que cumplen con el filtro a aplicar.

En la navegación facetada, el espacio de información se divide utilizando las dimensiones conceptuales de los datos. Estas dimensiones son denominadas facetas y representan las principales características de los elementos del espacio. Cada faceta tiene múltiples valores de restricción, y el usuario selecciona uno de estos valores de restricción para restringir elementos relevantes en el espacio. El mapeo de estos conceptos a los elementos de una sentencia RDF es directo: los elementos del espacio son los sujetos, las facetas son las propiedades y los valores de restricción son los objetos [Oren *et al.*, 2006].

¹¹Véase en <https://www.amazon.com/>

¹²Véase en <https://www.ebay.com/>

¹³Véase en <https://www.solotodo.cl/>

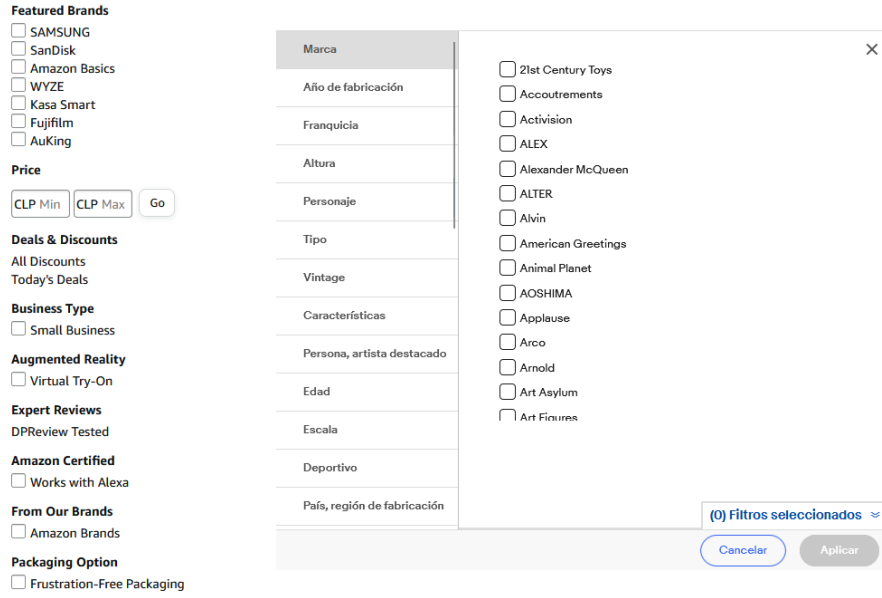


Figura 5: Filtros según facetas en Amazon e Ebay respectivamente.
Fuente: elaboración propia.

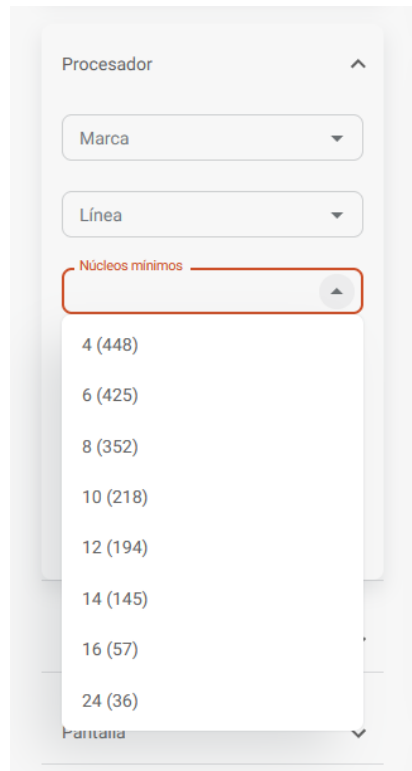


Figura 6: Filtros según facetas en Solotodo, en donde se indican la cantidad de coincidencias.
Fuente: elaboración propia

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

Dada la falta de una estructura que permita el manejo de grandes volúmenes de datos y la integración con conjuntos de datos externos, es que se propone la construcción de una ontología de estructura y comunicación para la integración de los datos de la CC, provistos por la BCN, con los datos históricos de las constituciones del mundo, provistos por Comparative Constitute Project (CCP), organización libre de lucro a encargada de la plataforma *Constitute Project*¹⁴. Para facilitar la integración con futuros conjuntos de datos, el acceso y la utilización de esta ontología, se seguirán las recomendaciones de la W3C, entendidas como el estándar de desarrollo en este contexto. Esta propuesta también incluye un sistema de consultas, basada en el estándar y técnicas de búsqueda, que permita la fácil construcción de visualizaciones de los datos integrados.

A continuación, se describe el proceso de construcción de la ontología propuesta y qué herramientas se utilizaron para este propósito. Primero se describirán los datos de cada fuente utilizada, su formato de almacenamiento y estructura. Luego, se presentará la ontología propuesta indicando sus principales características, para luego indicar cómo se lleva a cabo la integración entre las fuentes de datos a través de esta estructura. También se describirá cómo fueron procesados los datos y llevados a un formato útil para el almacenamiento en la estructura planteada. Finalmente, se presentará cómo se consulta sobre la ontología y qué métodos se utilizaron para llevar las respuestas a representaciones gráficas.

3.1. Análisis preliminar de datos

Una parte fundamental antes de construir la ontología es conocer y analizar con qué datos se cuentan para este propósito. Las dos fuentes de datos están estructuradas de forma distinta, pero siguen los principios estandarizados de la W3C. A continuación, se dará una descripción detallada del formato y estructura de los datos provenientes de la BCN y CCP.

3.1.1. Análisis a los datos de la BCN

La BCN, en su misión de transparencia y acercar todo lo que ocurre dentro de las cámaras y procesos legislativos a la ciudadanía, tiene a disposición toda la información del proceso constituyente llevado a cabo entre 2020 y 2022. En específico, en esta plataforma encontramos:

¹⁴Véase en <https://comparativeconstitutionsproject.org/>

- Candidatos a Constituyentes (nombre, edad y género)
- Constituyentes electos (nombre, edad y género)
- Profesión de los Constituyentes
- Redes sociales de los Constituyentes (Instagram, Facebook, Twitter y/o Pagina Personal)
- Cargos públicos que ha tenido el Constituyente
- Distrito al que representa el Constituyente
- Partido político del Constituyente
- Pacto político del Constituyente
- Comisiones en las que participó el Constituyente
- Discursos en los que participó el Constituyente
- Contenido, duración, número de sesión y fecha de cada uno de los discursos

Estos datos, almacenados en las bases de datos de la BCN, no tienen una estructura pública definida y sólo son accesibles desde la plataforma web de la BCN.

A inicios de 2021, Carina Flores, estudiante de la Universidad Técnica Federico Santa María (UTFSM), en conjunto con el IMFD, desarrolló un proceso de análisis de texto sobre los discursos de las sesiones de la CC, con el fin de establecer qué tópicos se hablan en cada uno de ellos. Este proceso lo llevó a cabo a través de técnicas de Web Scrapping¹⁵ y técnicas de procesamiento de texto, finalizando en la identificación de 55 tópicos. Además de este análisis, incluyó toda la información relacionada a los candidatos y constituyentes presente en la plataforma de la BCN, esto incluye todos los puntos antes comentados¹⁶. Toda la información obtenida la estructuró en una ontología RDF, con el objetivo de poner a disposición una estructura que unifique estos datos, además de permitir la consulta de estos.

La ontología propuesta por Carina se divide en dos partes principales: Los datos personales de cada integrante del proceso (aquí se incluyen redes sociales, partidos políticos, etc) y los datos de cada integrante en el contexto de la CC (aquí se incluyen comisiones, discursos, etc).

El primer subgrafo tiene como sujeto principal a la Persona, y contiene las siguientes clases:

- **Persona:** representación estándar de Persona. Se indica el nombre, edad y genero.
IRI: <http://xmlns.com/foaf/0.1/Person>

¹⁵Proceso de extracción de datos directamente de paginas webs mediante software automatizado.

¹⁶Los archivos fuentes utilizados en este proceso se pueden acceder desde el siguiente enlace: https://github.com/CarinaColoane/graph_cc/

- **Candidato:** subclase de Persona, representa a todas las personas identificadas también como candidatos a constituyente.
IRI: <http://www.imfd.cl/ontologia/convencion#Candidato>
- **Cargo:** cargo público que ha ejercido la persona, si es que corresponde.
IRI: <http://www.imfd.cl/ontologia/convencion#Cargo>
- **Profesión:** título o títulos profesionales que tiene la persona, si es que corresponde.
IRI: <http://www.imfd.cl/ontologia/convencion#Profesion>
- **Distrito:** distrito electoral por el que postula y/o representa la persona.
IRI: <http://www.imfd.cl/ontologia/convencion#Distrito>
- **Red Social:** redes sociales con las que cuenta la persona. Se incluye Instagram, Facebook, Twitter y paginas personales.
IRI: <http://www.imfd.cl/ontologia/convencion#RedSocial>
- **Partido Político:** partido político al que pertenece la persona, si es que corresponde. Se indica el nombre del partido.
IRI: <http://www.imfd.cl/ontologia/convencion#PartidoPolitico>
- **Pacto:** asociación de varios partidos políticos con objetivos en común. Se indica el nombre del pacto.
IRI: <http://www.imfd.cl/ontologia/convencion#Pacto>

Las clases del primer subgrafo presentado están conectadas según las siguientes propiedades:

- Candidato **subClassOf** Persona
- Persona **haTrabajado** Cargo
- Persona **haFinalizado** Profesion
- Persona **postuloPorDistrito** Distrito
- Persona **tieneRedSocial** RedSocial
- Persona **perteneceAPartidoPolitico** PartidoPolitico
- Pacto **tienePartidoPolitico** Pacto

A continuación, en la Figura 7 se presenta una representación gráfica del primer subgrafo:

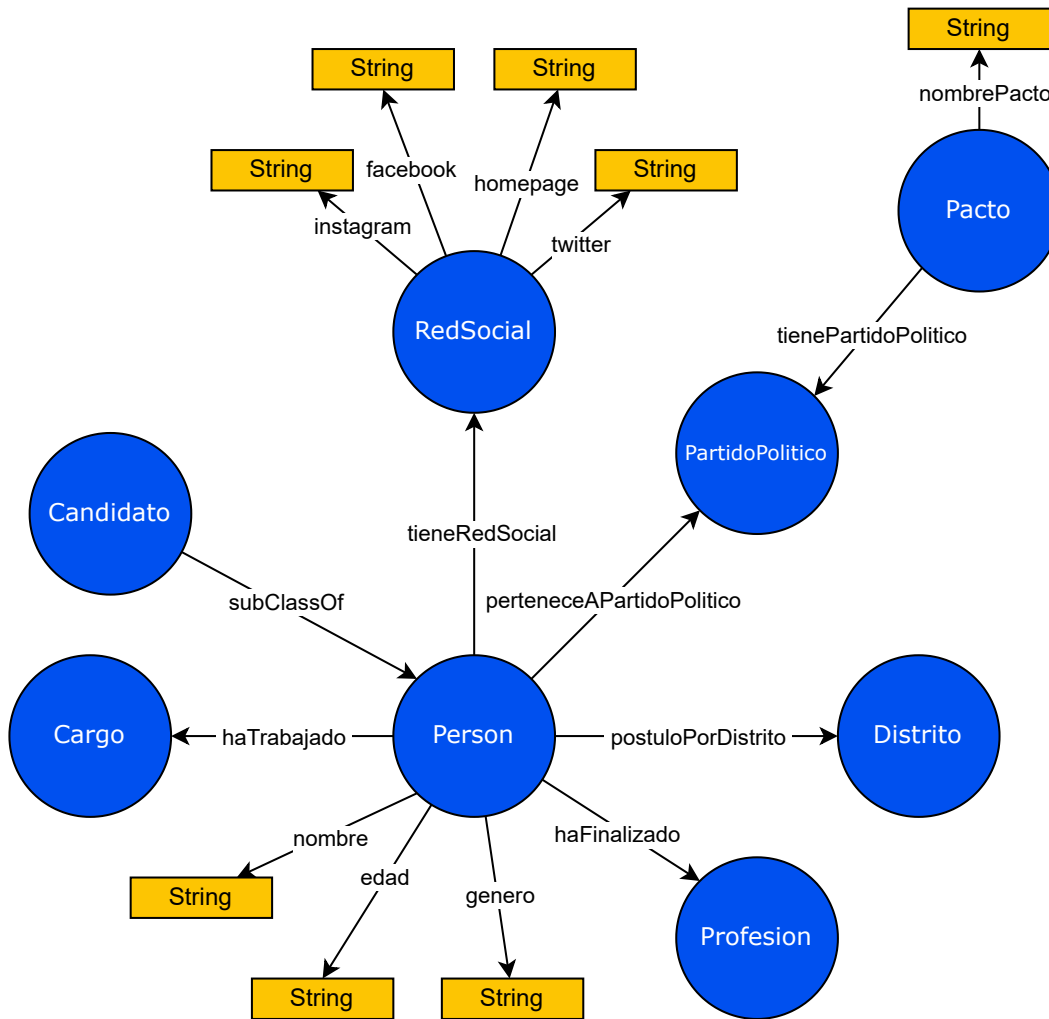


Figura 7: Representación gráfica del primer subgrafo.
Fuente: elaboración propia

El segundo subgrafo tiene como sujeto principal al Constituyente, y contiene las siguientes clases:

- **Constituyente:** subclase de Persona, representa a todas las personas identificadas también como Constituyentes.
IRI: <http://www.imfd.cl/ontologia/convencion#Constituyente>
- **Comisión:** grupos de constituyentes dedicados a tratar temas en específico. Se incluye el nombre de la comisión.
IRI: <http://www.imfd.cl/ontologia/convencion#Comision>

- **Discurso:** exposiciones orales dadas por los constituyentes a lo largo de las sesiones de la CC. Se incluye el tipo de discurso, la duración, el numero de sesión, la fecha en que fue dado y la transcripción respectiva.
IRI: <http://www.imfd.cl/ontologia/convencion#Discurso>
- **Tópico:** tema relacionado a los discursos a través de procesamiento de texto. Se incluye el nombre del tópico.
IRI: <http://www.imfd.cl/ontologia/convencion#Topico>

Las clases del segundo subgrafo presentado están conectadas según las siguientes propiedades:

- Constituyente **perteneceAComision** Comision
- Constituyente **participaEnDiscurso** Discurso
- Discurso **tieneTopico** Topico

A continuación, en la Figura 8 se presenta una representación gráfica del segundo subgrafo:

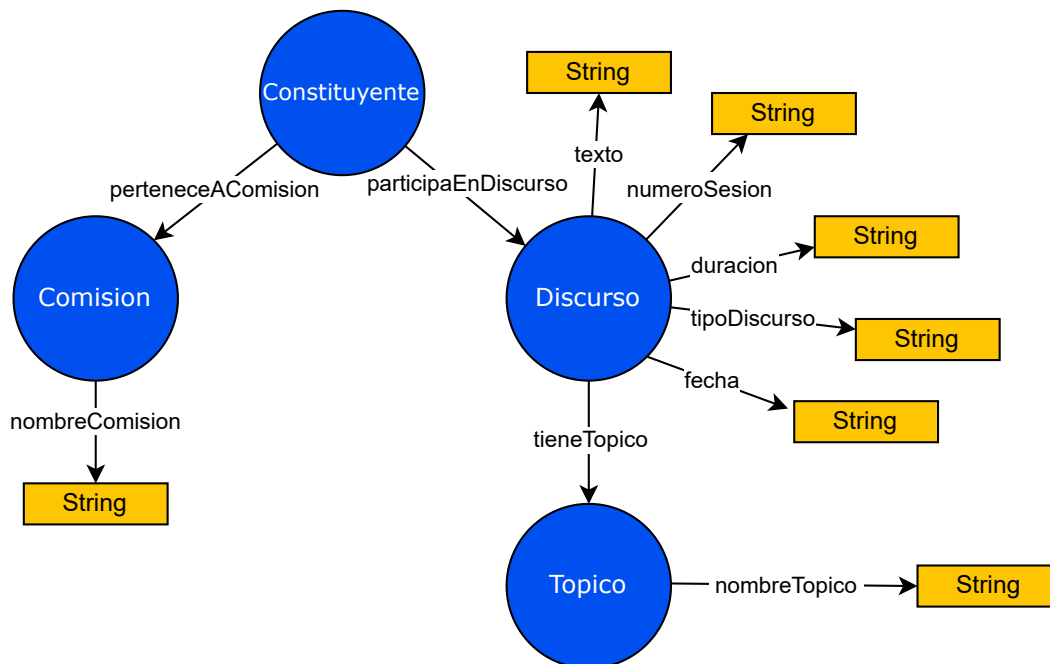


Figura 8: Representación gráfica del segundo subgrafo.
Fuente: elaboración propia

Ambos subgrafos se unen mediante la relación entre la clase Constituyente y Persona. El grafo completo de la ontología propuesta por Carina se puede ver en la Figura 47 de los Anexos A.

En la Tabla 1 del Anexo A se encuentra en detalle la taxonomía de los tópicos utilizados para clasificar cada discurso. Para aclarar, en cada discurso puede participar uno o más constituyentes, y puede estar relacionado a uno o más tópicos.

La estructura y los datos trabajados por Carina, se concentran en 2 archivos:

- **convencion_constituyente.owl**

Este archivo contiene las sentencias RDF que definen la ontología de estructura y comunicación que definió Carina en su trabajo. A nivel visual, es lo mostrado anteriormente en ambos subgrafos descritos.

- **grafo.ttl**

La extensión de archivos *ttl* es para definir los archivos *Turtle*, archivos de representación de grafos RDF en texto plano. Ideales para la escritura de *triples* sin aumentar demasiado el tamaño del archivo [Beckett *et al.*, 2014]. Es decir, en este archivo están los *triples* de los datos duros recopilados y construidos por Carina, que se guían por la ontología previamente definida.

3.1.2. Análisis a los datos de CCP

El *Comparative Constitution Project*, nacido en 2005, tiene la finalidad de proporcionar datos sistemáticos e históricos de las constituciones del mundo, y así, facilitar el análisis y comprensión de estos datos en el contexto tanto de estudios de derecho, como en el estudio básico de las constituciones de otros países actuales e históricas ¹⁷. El conjunto de datos, disponibles en la página oficial del proyecto ¹⁸, son utilizados para la construcción de la plataforma web *Constitute Project* ¹⁹. Si bien, para la construcción de esta plataforma fue necesaria la construcción de una estructura a través del planteamiento de una ontología, esta no está disponible oficialmente al público, en cambio nos encontramos con 6 archivos, descritos a continuación:

- **topic.xml**

Las constituciones presentes en el conjunto de datos están clasificadas bajo distintos tópicos, estos tópicos están estructurados como clases RDF. El archivo *topic.xml* contiene la taxonomía de estos tópicos, y los datos base de estos como por ejemplo el

¹⁷Véase en <https://comparativeconstitutionsproject.org/about-ccp/>

¹⁸Véase en <https://comparativeconstitutionsproject.org/download-data/>

¹⁹Véase en <https://constituteproject.org/>

nombre y descripción en distintos idiomas. La taxonomía describe a un total de 399 tópicos divididos en 3 niveles de profundidad. Por motivos de espacio, en la Tabla 2 del Anexo B sólo se presentan 2 niveles de profundidad ²⁰.

■ **countries.xml**

Cada constitución está asociada a un país en específico. En este archivo se definen todas las áreas territoriales y grupos y/u organizaciones a nivel países, tal como países, regiones continentales, continentes, grupos económicos, sociales o gubernamentales, entre otros. En cuanto a los países, se incluyen datos como el IRI, el nombre oficial, códigos de identificación como ISO, UN, UNDP, entre otros. Además, se indica a qué región continental y continente pertenece, así como también pertenencia a organizaciones y tratados internacionales.

■ **chronology.xml**

Este archivo contiene el histórico de todas las constituciones consideradas en el conjunto de datos. Cada constitución es clasificada como un evento, en el que se incluyen 5 datos relevantes: el año del evento, el tipo de evento (Nuevo, Reforma, Interna, etc.), el país asociado, un identificador y el IRI correspondiente. En particular el IRI contiene el URL y el URN, siendo el URN una combinación del nombre del país y el año del evento. Desde este punto, cuando se hable de evento se considerará como símil a hablar de evento constitucional.

■ **metadata.xml**

Es una extensión del archivo anterior, contiene datos más detallados de la constitución y el país al que pertenece. Además de indicar las relaciones RDF que unen el IRI del evento con el IRI del país, indica el largo en palabras de la constitución, el tipo de evento, el identificador de evento, identificadores del país como nombre, códigos ISO, UN, entre otros. En específico, en el conjunto de datos presentados por CCP hay documentados 4226 eventos.

■ **codebook.pdf**

Para analizar las constituciones, y su respectiva clasificación, CCP elaboró un sistema de preguntas y respuestas. Este documento presenta la descripción del método de recolección de estos datos, y presenta las preguntas más significativas del estudio, es decir, las que entregan mayor riqueza de datos (aproximadamente 2/3 del total de preguntas ²¹). Se destacan 3 tipos de pregunta: las preguntas que permiten como respuesta sí o no, preguntas que permiten respuestas en base a un conjunto predefinido, y las preguntas que permiten seleccionar más de una respuesta. Cada una de las preguntas está asociada a al menos uno de los tópicos presentados en primer archivo presentado, esto, a través del URN correspondiente. Específicamente, sólo los tópicos

²⁰La taxonomía completa está disponible en el siguiente archivo: Taxonomía tópicos CCP

²¹El cuestionario completo, con las 669 preguntas utilizadas y sus posibles respuestas están indicadas en el siguiente documento: <https://www.comparativeconstitutionsproject.org/files/surveyinstrument.pdf>

del más bajo nivel de jerarquía son asociados a alguna pregunta, los demás tópicos son utilizados para definir la taxonomía de la estructura.

- **ccpcnc_v4.csv**

Es el archivo con mayor cantidad de datos de todos, ya que contiene las respuestas de cada evento constitucional a las preguntas presentadas en el archivo anterior. Su formato csv permite asociar cada valor a cada una de las preguntas a través del URN indicado en estas. El nombre del archivo es un acrónimo para *Comparative Constitutions Project Characteristic of National Constitutions, versión 4*, que contiene, tal como dice, en detalle las características de cada constitución considerada dentro del conjunto de datos. Esta versión corresponde a la última disponible a día de hoy, actualizada al 24 de octubre de 2022.

Si bien, tal como se comentó al inicio, la ontología utilizada por CCP no está disponible, los archivos que disponen sí entregan una clasificación y orden que es útil al momento de construir una nueva estructura. Dado que el archivo de tópicos ya presenta una estructura de clases RDF no habrá mayor problema para llevarlo a un grafo RDF, pero los demás archivos presentan un desafío en cuando a qué datos considerar y a cómo llevar estos a la ontología que se planteará.

Según las descripciones de ambos conjuntos de datos, es necesario dar soporte estructural a 4 subconjuntos: Los datos de la BCN, los tópicos de clasificación de CCP, los países indicados por CCP y los eventos que representan a cada constitución.

3.2. Planteamiento de la Ontología

Para la construcción de la ontología se utilizó *Protégé*, software de código abierto para la construcción de ontologías. Respaldado por académicos, gobiernos y usuarios corporativos que construyen soluciones para áreas como la biomedicina, comercio electrónico y modelado organizacional. Además de tener una comunidad activa que aporta con una amplia documentación y nuevos complementos, también da soporte a los estándares propuestos por la W3C [Musen, 2015]. Permitiendo así, cumplir con los objetivos de estandarización de esta memoria.

Utilizando como base lo planteado por Carina en cuanto a los datos de la BCN, es necesario implementar una ontología que de soporte a los datos de CCP y además permita la integración entre estas dos fuentes. En esta sección dividiremos la discusión en dos partes: la estructura para los datos CCP y la estructura de integración.

3.2.1. Estructura para los datos CCP

Dado que el archivo que contiene los tópicos ya contiene una estructura RDF, estos se agregan a la ontología de la BCN mediante herramientas del software utilizado, que permite fácilmente indicar la jerarquía y luego indicar el nombre de las clases y su respectiva descripción. Es decir, se agrega el subgrafo correspondiente a los tópicos de CCP. En la Figura 9 se presenta de forma visual el subgrafo indicando sólo el primer nivel de profundidad por razones de espacio y orden, que de otra forma dificultarían el entendimiento de la estructura.

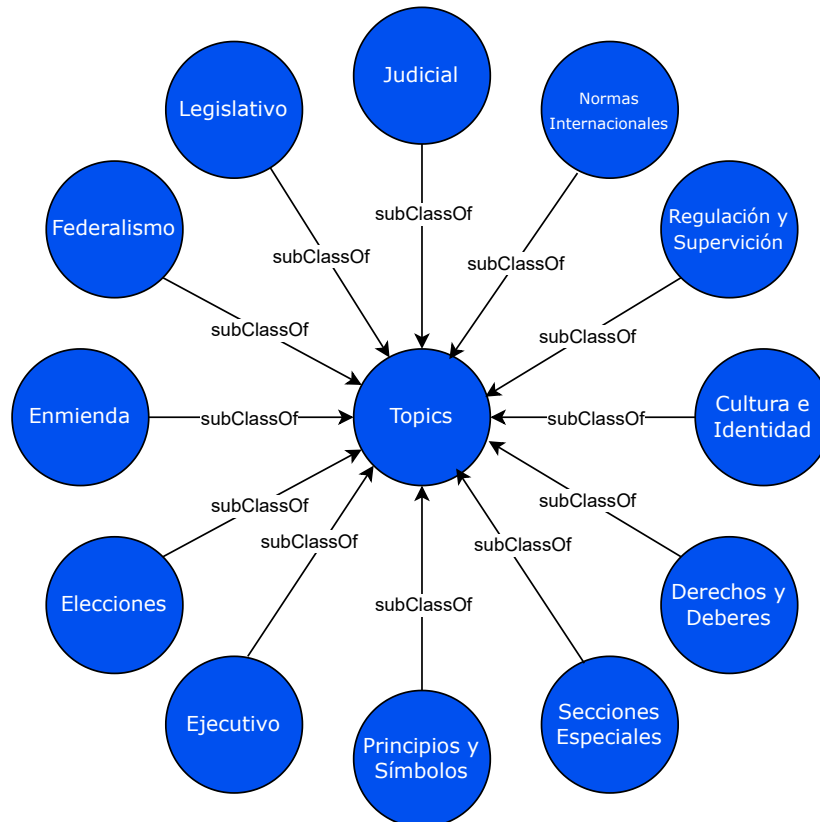


Figura 9: Representación gráfica del subgrafo de tópicos CCP mostrando sólo el primer nivel de profundidad. En la Figura 48 en los Anexos B se muestra la representación completa.

Fuente: elaboración propia.

Para darle estructura a los países, no sólo se debe considerar los datos de estos, sino también los datos de las áreas geográficas a las que pertenece como lo son continentes y regiones del mundo. Así también, se considerarán organizaciones y tratados internacionales, con el objetivo de abarcar la mayor cantidad de datos posibles. Para esto, y en consideración de los datos provistos, se proponen las siguientes clases:

- **Lugar:** clase principal de este subgrafo, indicará los atributos base de cada uno de los

elementos a definir. Se incluirá el nombre en español e inglés de estas entidades dados por los atributos *nameListES* y *nameListEN* respectivamente.

IRI: <http://www.imfd.cl/ontologia/convencion#Place>

- **Grupo:** subclase de Lugar, contendrá a los tipos de agrupaciones de países, ya sean organizaciones internacionales o tratados económicos.
IRI: <http://www.imfd.cl/ontologia/convencion/group>
- **Territorio:** subclase de Lugar, contendrá a los tipos de territorios en base a lugares geográficos. En específico es la previa definición de países.
IRI: <http://www.imfd.cl/ontologia/convencion/territory>
- **Región Económica:** subclase de Grupo, indica los tipo de agrupaciones y/u organizaciones basadas en temas económicos.
IRI: http://www.imfd.cl/ontologia/convencion/economic_region
- **Región Geográfica:** subclase de Grupo, indica los tipo de agrupaciones y/u organizaciones basadas en temas geográficos.
IRI: http://www.imfd.cl/ontologia/convencion/geographical_region
- **Organización:** subclase de Grupo, indica los tipos de agrupaciones y/u organizaciones basadas en temas políticos.
IRI: <http://www.imfd.cl/ontologia/convencion/organization>
- **Grupo Especial:** subclase de Grupo, indica los tipos de agrupaciones y/u organizaciones no presentes en las categorías mencionadas anteriormente.
IRI: http://www.imfd.cl/ontologia/convencion/special_group
- **Auto Gobierno:** subclase de Territorio, indica las superficies auto gobernadas que en la práctica corresponde a la mayoría de países.
IRI: http://www.imfd.cl/ontologia/convencion/self_governing
- **Disputado:** subclase de Territorio, indica las superficies que se encuentran en disputa entre 2 o más estados soberanos.
IRI: <http://www.imfd.cl/ontologia/convencion/disputed>
- **Otro:** subclase de Territorio, indicar las superficies que no coinciden con las clasificaciones anteriores.
IRI: <http://www.imfd.cl/ontologia/convencion/other>

También, para estructurar las distintas relaciones que pueden existir entre estas clases, se proponen las siguientes propiedades:

- Territorio **hasBorderWith** Territorio
- Territorio **isInGroup** Grupo

■ Grupo **hasMember** Territorio

La representación visual de este subgrafo se presenta en la Figura 10, en la que se incluyen también los atributos de las clases, según corresponda.

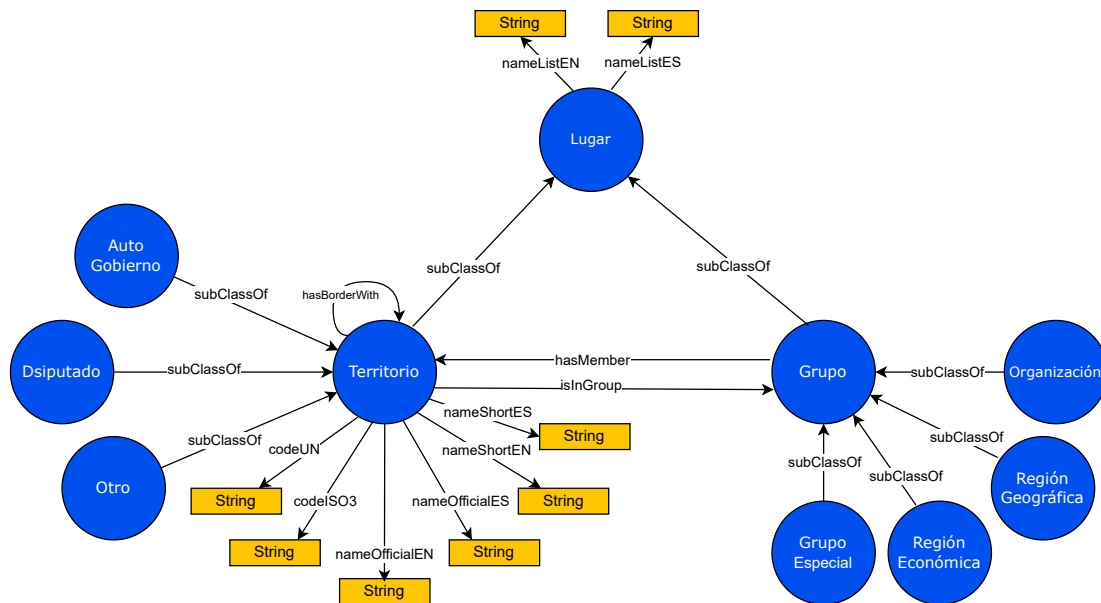


Figura 10: Representación gráfica del subgrafo de países y grupos propuesto.
Fuente: elaboración propia.

En cuanto a los eventos, se seguirá la implementación estándar propuesta por la W3C, que proporciona clases y atributos para identificar este tipo de entidades. En específico, se utilizará la clase *Evento* (IRI: <http://www.w3.org/1999/02/22-rdf-syntax-ns#event>) y los atributos *event_type* y *event_year* [Gandon y Schreiber, 2014], a los que se le sumarán *id* y *country*. Dado que cada evento representa un documento constitucional, y estos pertenecen a un país, se propone la implementación de la propiedad ***isConstitutionOf***, que va desde la clase *Evento* hasta la clase *Territorio*, para representar esta asociación.

En la Figura 11 a continuación, se visualiza la representación gráfica del subgrafo para dar soporte a las constituciones:

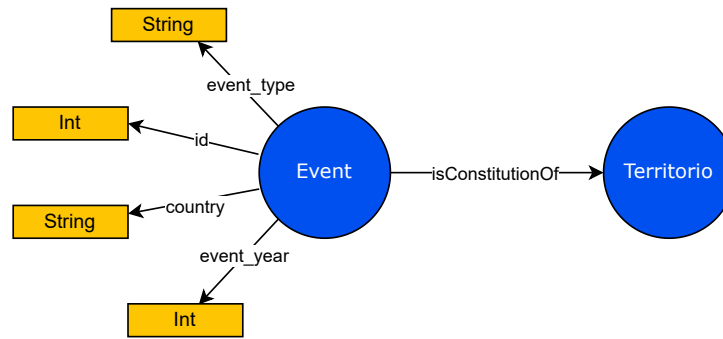


Figura 11: Representación gráfica del subgrafo de eventos propuesto para dar soporte a los datos base de una constitución.

Fuente: elaboración propia.

3.2.2. Estructura de integración

Como hemos evidenciado en los puntos anteriores, ambos conjuntos de datos a integrar implementan una clasificación en tópicos, ya sea para discursos por parte de la BCN, como para constituciones por parte de CCP. Esta similitud es la base fundamental de la integración a implementar.

Si bien hay tópicos de la BCN que se pueden homologar directamente con tópicos de CCP, como por ejemplo Anti Corrupción de la categoría de Comisiones en BCN y la clase Comisión anti-corrupción de CCP, hay otros en que la relación no es directa, como por ejemplo el tópico Símbolos de la categoría Cultura en BCN y las clases Himno Nacional y Bandera Nacional de CCP. Este problema se da, principalmente, porque el símil del tópico puede conceptualmente referirse a tópicos más específicos. En el ejemplo anterior, Himno Nacional es más específico que Símbolos, en otras palabras, el tópico Himno Nacional se puede considerar como una subcategoría del tópico Símbolos. Por lo tanto, la homologación de tópicos se divide en 2 casos: el caso en cuando el tópico CCP es igual o más general que el tópico BCN, y el caso en que el tópico BCN es más general que el tópico CCP.

En la Tabla 3 del Anexo C se presentan los casos en donde los tópicos CCP son igual o más generales que los tópicos de la BCN.

Por el contrario, en la Tabla 4 del Anexo C se presentan los casos donde los tópicos BCN son más generales que los tópicos CCP.

A pesar de la definición de tópicos de ambas fuentes de datos, estos están estructurados de forma distinta. En BCN los tópicos son individuos/entidades de la clase Tópico, en cambio en CCP, los tópicos son clases que forman parte de la estructura del grafo RDF para clasificar

constituciones. En otras palabras, en la BCN cada individuo de la clase Discurso está relacionado con otro individuo de la clase Tópico, y en CCP, cada individuo de la clase Evento es también individuo de la clase Tópico. Esta diferencia impide relacionar directamente las definiciones de tópicos entre las fuentes de datos. Se propone entonces, relacionar los individuos de la clase Evento en CCP con los individuos de la clase Tópico en BCN en base a las siguientes propiedades:

- Tópico **incluidoEn** Evento: esta propiedad apunta a dar soporte en los casos en que el tópico CCP es igual o más general que el tópico BCN.
- Evento **incluyeParteDe** Tópico: esta propiedad apunta a dar soporte en los casos en que el tópico BCN es más general que el tópico CCP.

Notar que estas propiedades serán condicionadas por parte de CCP. Por ejemplo, en el caso ya comentado entre los tópicos Anti Corrupción de la categoría Comisiones en BCN y Comisión anti-corrupción de CCP, el tópico Anti Corrupción estará incluido en todos los eventos que pertenezcan a la clase Comisión anti-corrupción. Y de manera similar, todos los eventos que pertenezcan a las clases Bandera Nacional o Himno Nacional de CCP incluirán parte del tópico Símbolos de BCN.

La estructura del grafo RDF de la ontología propuesta se muestra en la Figura 49 del Anexo C. Se pueden evidenciar 2 componentes conexas, y cómo la interacción entre la clase Eventos de CCP y Tópicos de BCN permiten la unión de las estructuras de ambas fuentes de datos. Se evidencia también, que los tópicos CCP no participan directamente de la estructura, pues cómo se ha indicado anteriormente, estos sólo cumplen la función de clasificación y ordenamiento.

3.3. Procesamiento de datos

Para poblar la estructura planteada es necesario extraer y transformar los datos de CCP, y dado que los datos de la BCN están almacenados en un archivo *ttl*, con el objetivo de mantener un proceso uniforme y estandarizado, los datos de CCP también serán almacenados en este tipo de archivos. Los datos son técnicamente los individuos de cada una de las clases definidas y todos los atributos asociados, y se dividen en 3 conjuntos: países, constituciones y características de las constituciones. Es decir, es necesario traducir el contenido de los archivos *countries.xml*, *chronology.xml*, y *ccpcnc_v4.csv* al formato *ttl*.

Para este propósito se utiliza el lenguaje de programación *Python* y las librerías *pandas* y *regex*. Ya que al ser un lenguaje de alto nivel, sumado a su baja complejidad y amplia documentación, permite lograr el objetivo con pocas líneas de código [Van Rossum y Drake, 2009].

3.3.1. Datos de Países

Los datos de países están concentrados en el archivo *countries.xml*, y no sólo contiene los países, sino también grupos y asociaciones internacionales relacionados a ellos. En general, los datos en el archivo están estructurados tal como se muestra en la Figura 12 a continuación:

```
<NamedIndividual rdf:about="http://www.constituteproject.org/ontology/Bolivia">
  <rdf:type rdf:resource="http://www.constituteproject.org/ontology/self_governing"/>
  <hasBorderWith rdf:resource="http://www.constituteproject.org/ontology/Argentina"/>
  <hasBorderWith rdf:resource="http://www.constituteproject.org/ontology/Brazil"/>
  <hasBorderWith rdf:resource="http://www.constituteproject.org/ontology/Chile"/>
  <hasBorderWith rdf:resource="http://www.constituteproject.org/ontology/Paraguay"/>
  <hasBorderWith rdf:resource="http://www.constituteproject.org/ontology/Peru"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/Americas"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/CAN"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/FAO"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/FAO_2006"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/LLDC"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/MERCOSUR_1991"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/World"/>
  <isInGroup rdf:resource="http://www.constituteproject.org/ontology/southern_America"/>
  <codeISO3 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">BOL</codeISO3>
  <codeUN rdf:datatype="http://www.w3.org/2001/XMLSchema#string">068</codeUN>
  <nameListEN rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Bolivia (Plurinational State of)</nameListEN>
  <nameListES rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Bolivia (Estado Plurinacional de)</nameListES>
  <nameOfficialEN rdf:datatype="http://www.w3.org/2001/XMLSchema#string">the Plurinational State of Bolivia</nameOfficialEN>
  <nameOfficialES rdf:datatype="http://www.w3.org/2001/XMLSchema#string">el Estado Plurinacional de Bolivia</nameOfficialES>
  <nameShortEN rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Bolivia</nameShortEN>
  <nameShortES rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Bolivia (Estado Plurinacional de)</nameShortES>
</NamedIndividual>
```

Figura 12: Estructura general de los datos del archivo *countries.xml*.

Fuente: elaboración propia a partir de [Elkins y Ginsburg, 2022].

La extracción se basa principalmente en obtener la información necesaria para construir los distintos *triples*. Para ello, a través de código se identificó al sujeto (IRI del individuo a describir), y se fueron variando las propiedades y objetos según corresponda. El resultado es almacenado en el archivo *countries.ttl* y sigue la estructura mostrada en la Figura 13.

El pseudocódigo del proceso desarrollado en Python se muestra en la Figura 50 del Anexo D. Para este caso, la definición de propiedades está basada en las relaciones que tiene la clase Lugar y sus subclases, ya sean atributos, propiedades de relación con otras clases o propiedades que indican el tipo de clase que es el sujeto. El archivo de destino es *countries.ttl*.

```

<http://www.constituteproject.org/ontology/Bolivia> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/self_governing>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/hasBorderWith>
<http://www.constituteproject.org/ontology/Argentina>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/hasBorderWith>
<http://www.constituteproject.org/ontology/Brazil>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/hasBorderWith>
<http://www.constituteproject.org/ontology/Chile>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/hasBorderWith>
<http://www.constituteproject.org/ontology/Paraguay>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/hasBorderWith>
<http://www.constituteproject.org/ontology/Peru>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/Americas>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/CAN>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/FAO>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/FAO_2006>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/LLDC>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/MERCOSUR_1991>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/World>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/isInGroup>
<http://www.constituteproject.org/ontology/southern_America>.
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/codeISO3>
"BOL".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/codeUN>
"068".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/nameListEN>
"Bolivia (Plurinational State of)".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/nameListES>
"Bolivia (Estado Plurinacional de)".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/nameOfficialEN>
"the Plurinational State of Bolivia".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/nameOfficialES>
"el Estado Plurinacional de Bolivia".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/nameShortEN>
"Bolivia".
<http://www.constituteproject.org/ontology/Bolivia> <http://www.constituteproject.org/ontology/nameShortES>
"Bolivia (Estado Plurinacional de)".
    
```

Figura 13: Estructura general de los datos del archivo *countries.ttl*.
Fuente: elaboración propia.

3.3.2. Datos de Constituciones

De forma similar, para la extracción de datos de las constituciones se trabajó sobre el archivo *chronology.xml* que contiene los datos base de los individuos de la clase Evento. La estructura de este archivo se muestra en la Figura 14 a continuación:

```

<event id="212" rdf:about="http://www.constituteproject.org/constitution/Chile_1980" country="Chile">
  <event_type>New</event_type>
  <event_year>1980</event_year>
</event>
    
```

Figura 14: Estructura general de los datos del archivo *chronology.xml*.
Fuente: elaboración propia a partir de [Elkins y Ginsburg, 2022].

Se observa que a pesar de ser similar a la estructura de *countries.xml*, los datos están estructurados de otra forma. Pues, el IRI del sujeto está presente en la misma entrada que el id del evento y el país al que pertenece. Si bien, se aplica el mismo pseudocódigo que en el caso anterior (Figura 50 en Anexo C), la identificación de propiedades requiere de un trabajo más elaborado, con ayuda de operaciones sobre strings y expresiones regulares. El archivo de destino toma el nombre de *eventos.ttl*, y sigue la estructura indicada en la Figura 15 a continuación:

```
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#event>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.constituteproject.org/constitution/id>
"212".
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.constituteproject.org/constitution/country>
"Chile".
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.constituteproject.org/constitution/isConstitutionOf>
<http://www.constituteproject.org/ontology/Chile>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#event_type>
"New".
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#event_year>
"1980".
```

Figura 15: Estructura general de los datos del archivo *eventos.ttl*.
Fuente: elaboración propia.

3.3.3. Características de constituciones

El archivo *ccpcnc_v4.csv* contiene 21342 registros y 2030 atributos para cada uno de ellos, pero según la extracción de eventos en el punto anterior, hay 4230 eventos. Según se explica en el documento *codebbok.pdf*, los registros no sólo describen eventos, sino también, procesos internos que ocurrieron dentro de un mismo año en cada país, es decir, que hay registros que se refieren al mismo evento. De forma similar, hay eventos en los que no se cuenta con el documento para poder hacer un análisis de características, ya sea porque son muy antiguos, muy recientes, o temas de derechos y acceso a estos documentos, por lo tanto no tendrán registros asociados en este archivo. En cuanto a los atributos, estos representan las respuestas a las 669 preguntas, y se incluyen datos adicionales y comentarios para cada una de ellas, si es que corresponde. En esta extracción sólo se considerará los registros pertenecientes a eventos, que corresponden a 4115 registros, y los atributos relacionadas a las preguntas que más riqueza entregan a los datos y que tengan un tópico CCP asociado, que corresponden a cerca de 340 atributos.

Dado el tipo de archivo, para la extracción se utiliza la librería *pandas*, que transforma el archivo en una tabla, en donde los registros toman las filas y los atributos las columnas. Las columnas se dividen en 2 subconjuntos, las que entregan datos para identificar el evento, y las que entregan las respuestas a cada una de las preguntas. Del primer subconjuntos se utiliza específicamente el identificador de evento para asociarlo a su IRI correspondiente. En cuanto al segundo subconjunto, el propio nombre de las columnas indica el URN de la pregunta asociada, que es el mismo URN de la clase CCP que representa.

Para procesar de mejor manera las respuestas, se clasificaron en 3 tipos:

- Tipo 1: respuestas a preguntas cerradas. Estas preguntas tienen en general sólo 2 respuestas, Sí o No. Un ejemplo es la pregunta asociada a la clase Bandera Nacional: ¿Contiene la constitución disposiciones relativas a la bandera nacional?
- Tipo 2: respuestas a preguntas de alternativas múltiples. Estas preguntas tienen más de una respuesta afirmativa, ya que indican cómo es implementado o qué define en específico la constitución. Un ejemplo es la pregunta asociada a la clase Inmunidad del Jefe de Gobierno: ¿Dispone el Jefe de Gobierno de inmunidad procesal? Donde las respuestas son: 1. Sí, inmunidad absoluta; 2. Sí, inmunidad limitada y 3. No, específicamente denegada.
- Tipo 3: respuestas a preguntas de selección múltiple. Estas preguntas tienen más de una respuesta afirmativa y pueden ser varias a la vez, por lo que cuentan con más de una columna de respuesta. En algunos casos cada respuesta representa una clase (caso 1), o en el caso más común para este tipo, cada respuesta está relacionada a la misma clase. Un ejemplo del caso 1 es la pregunta: ¿Para cuál de los siguientes tribunales especializados contiene disposiciones la constitución? y sus respuestas están asociadas a las clases Establecimiento de tribunales administrativos, Establecimiento del Tribunal Constitucional, Establecimiento de tribunales de amparo, etc. Un ejemplo del caso 2 es la pregunta: ¿Qué restricciones adicionales impone la constitución sobre la elegibilidad para servir como miembro de los tribunales ordinarios? y sus respuestas están asociadas a la clase Requisitos de los jueces de tribunales ordinarios.

Dada esta estructura, cada tipo de respuesta necesitaba un proceso distinto, siendo el Tipo 1 el más sencillo y el Tipo 3 el más complejo. En las Figuras 51, 52, 53 y 54 de los Anexos D se encuentra el pseudocódigo para los Tipo 1, Tipo 2, Tipo 3 caso 1 y Tipo 3 caso 2 respectivamente. El archivo de salida *datosDuros.ttl* es común para los 3 tipos, y sigue la estructura indicada en la Figura 16 a continuación:

```
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/dignity>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/solid>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/anthem>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/flag>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/amend>.
<http://www.constituteproject.org/constitution/Chile_1980> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.constituteproject.org/ontology/execnum>.
```

Figura 16: Estructura general de los datos del archivo *datosDuros.ttl*.

Fuente: elaboración propia.

3.3.4. Integración

Para crear el archivo *ttl* con los datos relacionados a la estructura de integración propuesta, se utilizó el archivo ya creado en el punto anterior *datosDuros.ttl*, ya que la existencia de las propiedades dependen de a qué clase pertenece cada evento constitucional. Este proceso se divide en dos tipos:

- Tipo A: Propiedad **IncluidoEn**, si el evento pertenece a alguno de los Tópicos CCP indicados en la Tabla 3, entonces se crea la declaración *Tópico IncluidoEn Evento* con el o los Tópicos BCN correspondientes.
- Tipo B: Propiedad **IncluyeParteDe**, si el evento pertenece a alguno de los Tópicos CCP indicados en la Tabla 4, entonces se crea la declaración *Evento IncluyeParteDe Tópico* con el o los Tópicos BCN correspondientes.

En la Figura 55 del Anexo D se presenta el pseudocódigo utilizado, que incluye ambos tipos descritos. El archivo de destino es *integracion.ttl* y sigue la estructura presentada en al Figura 17 a continuación:

```
<http://www.constituteproject.org/constitution/Togo_2019> <http://www.imfd.cl/ontologia/convencion#incluyeParteDe>
<http://www.example.cl/topico/comunicacion> .
<http://www.example.cl/topico/igualdad> <http://www.imfd.cl/ontologia/convencion#incluidoEn>
<http://www.constituteproject.org/constitution/Fiji_2013> .
<http://www.example.cl/topico/religion> <http://www.imfd.cl/ontologia/convencion#incluidoEn>
<http://www.constituteproject.org/constitution/Chile_1977> .
<http://www.constituteproject.org/constitution/Senegal_1970> <http://www.imfd.cl/ontologia/convencion#incluyeParteDe>
<http://www.example.cl/topico/territorio> .
<http://www.constituteproject.org/constitution/Sudan_2016> <http://www.imfd.cl/ontologia/convencion#incluyeParteDe>
<http://www.example.cl/topico/economia> .
<http://www.constituteproject.org/constitution/Albania_1976> <http://www.imfd.cl/ontologia/convencion#incluyeParteDe>
<http://www.example.cl/topico/derecho_social> .
<http://www.example.cl/topico/eleccion> <http://www.imfd.cl/ontologia/convencion#incluidoEn>
<http://www.constituteproject.org/constitution/Nicaragua_2004> .
<http://www.example.cl/topico/estado> <http://www.imfd.cl/ontologia/convencion#incluidoEn>
<http://www.constituteproject.org/constitution/Papua_New_Guinea_2010> .
<http://www.example.cl/topico/historia> <http://www.imfd.cl/ontologia/convencion#incluidoEn>
<http://www.constituteproject.org/constitution/United_Kingdom_1996> .
<http://www.example.cl/topico/eleccion> <http://www.imfd.cl/ontologia/convencion#incluidoEn>
<http://www.constituteproject.org/constitution/Malta_1997> .
```

Figura 17: Estructura general de los datos del archivo *integracion.ttl*.

Fuente: elaboración propia.

3.4. Sistema de consultas SPARQL

Con un total de 571.871 declaraciones en los archivos *ttl* construidos ²², es fundamental disponer de una herramienta que permita cargar la estructura, poblarla y realizar consultas para la extracción de datos. En este contexto, se puede seguir trabajando con *Protégé*, pero esta aplicación, basada en java, no está optimizada para este propósito cuando se cuenta con grandes volúmenes de datos [Musen, 2015]. Una alternativa, validada internacionalmente por organizaciones y académicos, es *Jena Fuseki*. Herramienta, basada en *Apache* [Miller *et al.*, 2010], que levanta un servidor local y permite procesar estructuras RDF a través de la carga de datos en formato *TTL*, realizar consultas SPARQL, y almacenar los datos de forma persistente a través de almacenamiento de *triples* en una base de datos relacional, aplicando técnicas estándar de transformación [Carroll *et al.*, 2004]. Utilizada por organizaciones de la industria como IBM [McCarthy, 2005], y organizaciones gubernamentales como la Unión Europea [EuropeanCommission, 2021].

La carga de datos con las herramientas provistas de la aplicación, se ven en la Figura 18 a continuación:

Files to upload

+ select files upload all

| name | size | speed | status | actions |
|-----------------|----------|------------|------------------------------------|----------------------|
| countries.ttl | 641.3kb | 641.53kb/s | 100.00 Triples uploaded: 4382 | upload now remove |
| datosDuros.ttl | 70.06mb | 26.41mb/s | 100.00 Triples uploaded: 457494 | upload now remove |
| eventos.ttl | 3.47mb | 894.38kb/s | 100.00 Triples uploaded: 25380 | upload now remove |
| grafo.ttl | 7.86mb | 3.23mb/s | 100.00 Triples uploaded: 41918 | upload now remove |
| integracion.ttl | 13.06mb | 2.03mb/s | 100.00 Triples uploaded: 84615 | upload now remove |
| ontologia.ttl | 533.35kb | 533.58kb/s | 100.00 Triples uploaded: 3629 | upload now remove |

Figura 18: Carga de estructura y datos en la aplicación *Jena Fuseki*.
Fuente: elaboración propia a partir de [Carroll *et al.*, 2004].

Notar que se incluyen los 4 archivos *TTL* construidos, el archivo con los datos de la BCN (*grafo.ttl*) y la estructura de la ontología propuesta (*ontologia.ttl*), que fue obtenida mediante la herramienta de exportación de *Protégé* [Musen, 2015].

²²countries.ttl = 4.382, eventos.ttl = 25.380, datosDuros.ttl = 457.494 e integracion.ttl = 84.615

Para realizar las consultas, la aplicación dispone de un espacio dinámico con definición de prefijos estándar. Un ejemplo de consulta se muestra en la Figura 19 a continuación:

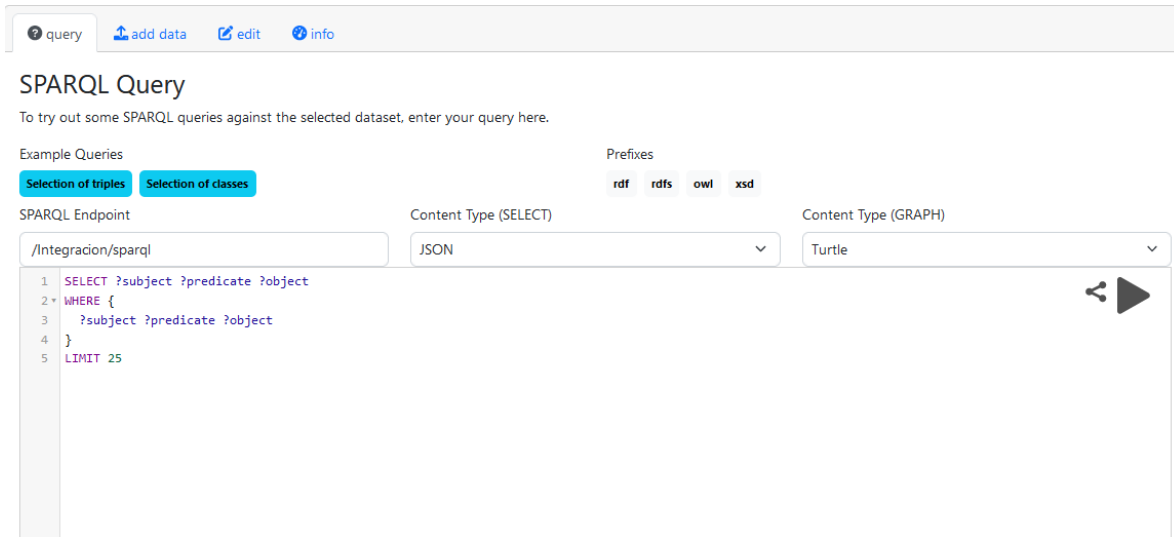


Figura 19: Ejemplo de consulta SPARQL en la aplicación *Jena Fuseki*.
Fuente: elaboración propia a partir de [Carroll *et al.*, 2004].

3.5. Creación de Visualizaciones

Para el procesamiento de las respuestas a las consultas SPARQL, al igual que en el procesamiento de datos, se utilizará el lenguaje de programación Python junto a la librería *Plotly*, que facilita la construcción de visualizaciones, y permite personalizar todos sus atributos de forma amplia en caso de ser necesario. Además de contar con una gran comunidad, y amplia documentación y casos de ejemplo [Inc., 2015].

Una de las ventajas de utilizar SPARQL 1.1 y *Jena Fuseki* es la posibilidad de extraer fácilmente el resultado de una consulta en un archivo *csv*, que se puede cargar como tabla en Python, procesarla y usarla como fuente para las visualizaciones en *Plotly*. Por ejemplo, para la consulta SPARQL mostrada en la Figura 20, se utiliza el código de Python mostrado en la Figura 21, y resulta la visualización mostrada en la Figura 22.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX cc: <http://www.imfd.cl/ontologia/convencion#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?nombrePartido (COUNT(?s) as ?constituyentes)
WHERE {
  ?s rdf:type cc:Constituyente;
     rdf:type ?person .

  ?person cc:perteneceAlPartidoPolitico ?partido .
  ?partido cc:nombrePartido ?nombrePartido.
}
GROUP BY ?nombrePartido
ORDER BY ?nombrePartido
```

Figura 20: Consulta SPARQL para la obtención de cantidad de constituyentes por partido político.
Fuente: elaboración propia.

```
import plotly.graph_objects as go
import pandas as pd

df = pd.read_csv("Preguntas/Constituyentes/Constituyentes por partido político.csv")
fig = go.Figure(data=[go.Pie(labels=df['nombrePartido'],
                             values=df['constituyentes'],
                             direction='clockwise',
                             sort=True)])
fig.update_traces(textinfo='percent')
fig.update_layout(width=700, height=400, margin={'l':0,'t':0,'b':0})
fig.write_image("Preguntas/Visualizaciones/Constituyentes por partido político.pdf")
fig.show()
```

Figura 21: Código en Python para el procesamiento y visualización de cantidad de constituyentes por partido político.
Fuente: elaboración propia.

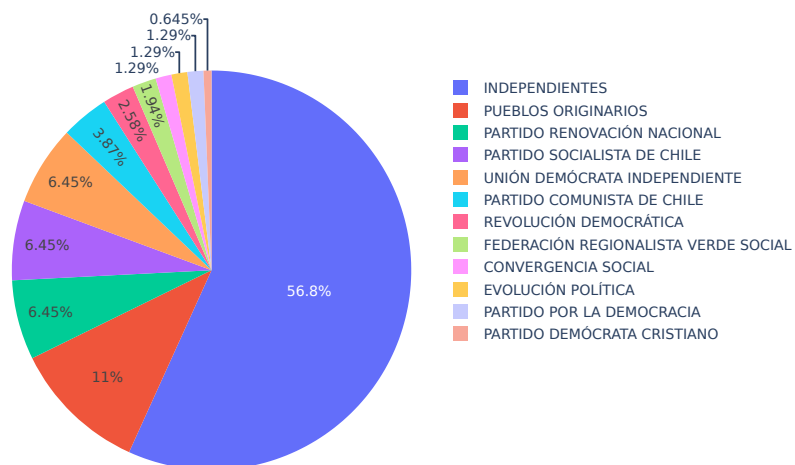


Figura 22: Porcentaje de constituyentes por partido político.
Fuente: elaboración propia.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

La validación de la solución se basa en la evaluación de la ontología respecto a tres puntos base: que la ontología propuesta siga los estándares internacionales, que la ontología permita realizar consultas aplicando técnicas de búsqueda facetada, y por último, que la ontología permita realizar consultas que aporten valor agregado a partir de las fuentes integradas y la integración en sí.

4.1. Cumplimiento del estándar

Tal como se como se comentó en un inicio, uno de los desafíos a la hora de integrar datos es la falta de una estructura común o estandarizada, que permita la reducción de complejidad en tareas como extracción y procesamiento. Los principios de datos FAIR, entregan lineamientos para estandarizar el almacenamiento y publicación de datos y así, apoyar este objetivo. Estos lineamientos se basan en 4 puntos principales: Encontrables, Accesibles, Interoperables y Reutilizables.

Dado que en este trabajo no se están generando datos, si no dándoles otro sentido a través de la unión de dos conjuntos, basta con que cada conjunto cumpla con estos lineamientos para que lo desarrollado cumpla con el estándar buscado. En base a los archivos disponibles y analizados de CCP, además de lo que se ha hecho con ellos en esta memoria, demuestran de forma práctica el cumplimiento de cada uno de los apartados de los principios FAIR. En cuanto a la BCN, desde 2011 han trabajado con entes internacionales para mejorar sus procesos y servicios al Congreso y a los ciudadanos, logrando que sus proyectos sean los primeros en América Latina en cumplir con clasificación de 5 estrella de los datos abiertos²³ [BCN, 2011]. Certificando así, el cumplimiento de los lineamientos.

4.2. Consultas con técnicas Facetadas

Las técnicas de búsqueda facetada se basan en la aplicación de filtros de forma dinámica, permitiendo obtener el mismo resultado desde puntos de partida totalmente diferentes, tal como se explicó en el marco conceptual de esta memoria. En cuanto a los grafos RDF, estos soportan esta técnica de forma nativa, pues cada una de las propiedades y atributos de una clase RDF son las facetas asociadas a ese elemento, y al ir aplicando filtros o considerando nuevos sujetos que aumentan el espacio, estas facetas aumentan o disminuyen. Considerar

²³Clasificación propuesta por Tim Berners-Lee, inventor de la World Wide Web. Descrito en <https://datos.bcn.cl/es/informacion/las-5-estrellas>

también, que una misma faceta permite la aplicación de más de un valor a filtrar. A continuación, en las Figuras 24 y 25 se presentan 2 caminos distintos para llegar al mismo resultado.:

Antes de comenzar, se definen los prefijos a utilizar:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cc: <http://www.imfd.cl/ontologia/convencion#>
PREFIX cpo: <http://www.constituteproject.org/ontology/>
PREFIX cpc: <http://www.constituteproject.org/constitution/>
```

Figura 23: Definición de prefijos para ejemplo de búsqueda facetada.

Fuente: elaboración propia.

El primer camino, en la Figura 24 parte desde la búsqueda de eventos de año mayores a 2010, que pertenezcan a países que tengan como vecinos a Chile o España, y que el evento represente una constitución Nueva.

```
SELECT ?constitucion
WHERE {
  ?constitucion rdf:type rdf:event ;
               rdf:event_year ?año ;
               cpc:isConstitutionOf ?pais ;
               rdf:event_type ?tipo .

  ?pais cpo:hasBorderWith ?paisBorde .

  ?paisBorde cpo:nameShortES ?nombrePaisBorde

  FILTER(?año > "2010")
  FILTER(?nombrePaisBorde IN ("Chile", "España"))
  Filter(?tipo = "New")
}

# Resultado CSV:
#  constitucion
#  http://www.constituteproject.org/constitution/Morocco_2011
```

Figura 24: Primera consulta para ejemplo de búsqueda facetada.

Fuente: elaboración propia.

En cambio, en la segunda consulta se buscan las constituciones de Marruecos que contengan el derecho a la libertad académica:

```

SELECT ?constitucion
WHERE {
  ?constitucion cpc:isConstitutionOf ?pais ;
    rdf:type cpo:acfree .

  ?pais cpo:nameShortES ?nombre .

  Filter(?nombre = "Marruecos")
}

# Resultado CSV:
# constitucion
# http://www.constituteproject.org/constitution/Morocco_2011
    
```

Figura 25: Segunda consulta para ejemplo de búsqueda facetada.
Fuente: elaboración propia.

4.3. Visualizaciones

Uno de los impactos más relevantes de lograr cumplir con los objetivos planteados, es permitir la construcción de visualizaciones que acerquen estos datos a la ciudadanía. Para ello, no sólo se debe dar soporte para la obtención de conocimiento de la integración realizada, si no también, de cada una de las fuentes de forma independiente.

A continuación, se mostraran visualizaciones de datos no triviales obtenidos a través de consultas sobre la ontología propuesta, con el fin de evidenciar que esta estructura permite la entrega de conocimiento que no es de fácil obtención por el gran volumen de datos.

Para todas las consultas de esta sección se utilizaron los prefijos indicados en la Figura 26 a continuación:

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cc: <http://www.imfd.cl/ontologia/convencion#>
PREFIX cpo: <http://www.constituteproject.org/ontology/>
PREFIX cpc: <http://www.constituteproject.org/constitution/>
    
```

Figura 26: Prefijos utilizados para las consultas de validación.
Fuente: elaboración propia.

4.3.1. Convención Constituyente

La ontología nos permite obtener los datos personales de cada candidato y constituyente, y así realizar un análisis general en base a estos atributos. A continuación, en la Figura 27 se presenta la consulta SPARQL que permite obtener la edad de todos los candidatos y constituyentes que participaron en el proceso:

```

SELECT ?edad
WHERE {
  ?candidato rdf:type cc:Candidato .
  # para Constituyentes:
  # ?candidato rdf:type cc:Constituyente .

  ?candidato rdf:type ?persona .
  ?persona cc:age ?edad .
}
    
```

Figura 27: Consulta SPARQL para obtener las edades de todo los candidatos a la CC.
Fuente: elaboración propia.

Estos datos nos permiten hacer un análisis comparativo de la distribución de edades entre estas dos categorías:

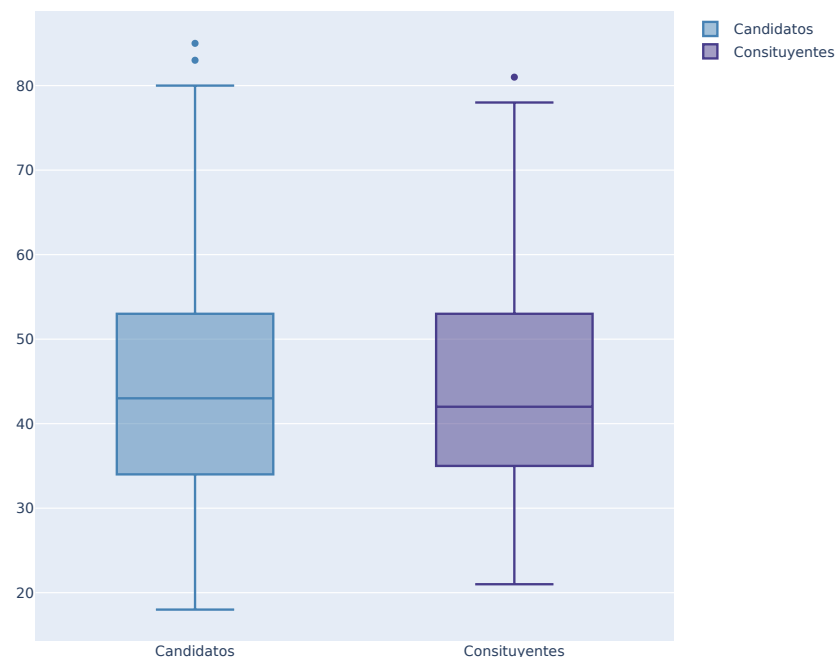


Figura 28: Distribución de edades para Candidatos y Constituyentes de la CC.
Fuente: elaboración propia.

En la Figura 28, se puede evidenciar que no hay grandes diferencias en la distribución de las edades entre estas categorías. Entre las diferencias, se encuentra la edad mínima y la distribución del 50% de valores más bajos, ya que en el conjunto de los candidatos estos se concentran entre los 18 y 43 años, en cambio, en el conjunto de los Constituyentes estos se concentran entre los 21 y 42 años. En general, esto nos indica que la edad de los candidatos no fue un factor con gran impacto en las elecciones, pues los constituyentes presentan similares valores de distribución.

Por otra parte, este conjunto de datos, y estructura, nos permite realizar un análisis interno

a la CC. Por ejemplo, si quisiéramos saber qué temas fueron los más recurrentes en los discursos de cada sesión, podríamos usar la consulta SPARQL presente en la Figura 29 a continuación (se presentan sólo los 20 más recurrentes):

```
SELECT ?nombreTopico (COUNT(?discurso) as ?discursos)
WHERE {
  ?tema rdf:type cc:Topico ;
        cc:nombreTopico ?nombreTopico .
  ?discurso cc:tieneTopico ?tema .
}
GROUP BY (?nombreTopico)
ORDER BY DESC (?discursos)
LIMIT 20
```

Figura 29: Consulta SPARQL para obtener el Top 20 de tópicos hablados en los discursos de la CC.

Fuente: elaboración propia.

Para visualizar estos datos, se propone un árbol de jerarquía presente en la Figura 30:

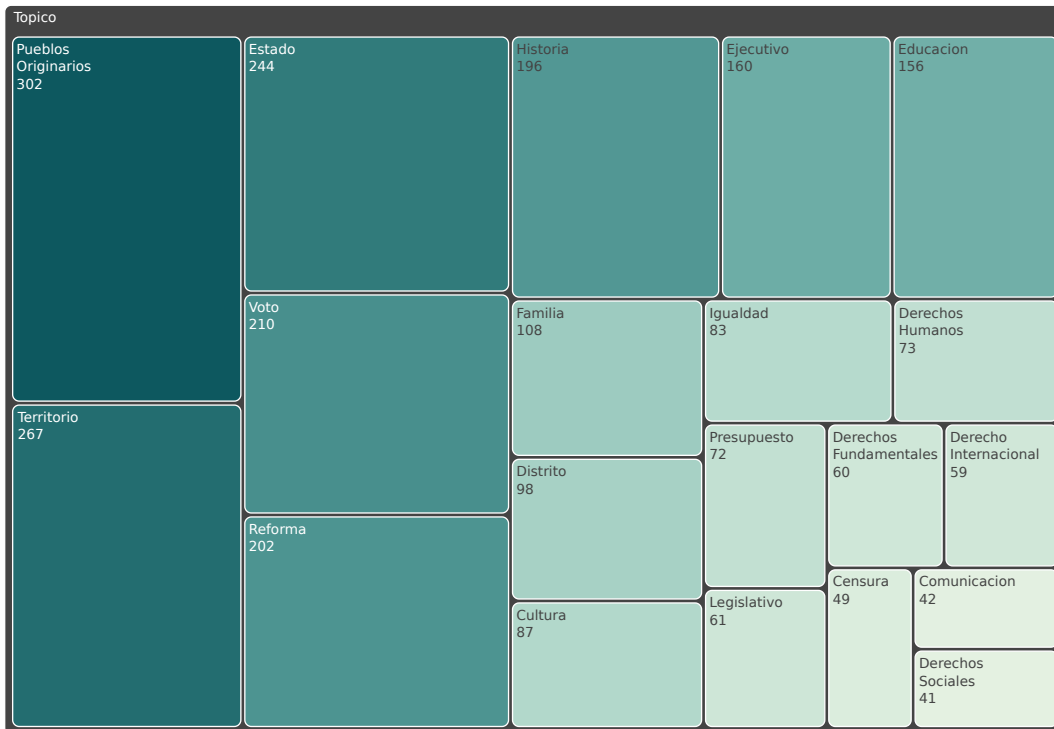


Figura 30: Cantidad de apariciones de tópicos en discursos de la CC.

Fuente: elaboración propia.

En la Figura 30 se destaca que el tema más recurrente fue *Pueblos Originarios* con 302 apariciones, seguido de *Territorio* con 267 y *Estado* con 244. Dando cuenta así, que las discusiones dentro de la CC se centraron en estos tópicos, y se les dio prioridad frente a otros.

Un análisis más detallado sería visualizar cómo se comportó la aparición de un tema a través del tiempo que duró el proceso. Para saber la cantidad de discursos según tópico y fecha, se utiliza la consulta a la ontología presente en la Figura 31 a continuación:

```
SELECT ?fecha ?nombreTopico (COUNT(?discurso) as ?cantidadDiscursos)
WHERE {
  ?discurso rdf:type cc:Discurso ;
    cc:fecha ?fecha ;
    cc:tieneTopico ?topico .

  ?topico cc:nombreTopico ?nombreTopico .
}
group by ?fecha ?nombreTopico
```

Figura 31: Consulta SPARQL para obtener la cantidad de discursos según tópico y fecha en la CC.

Fuente: elaboración propia.

Estos datos, nos permiten visualizar como se comportó en general la cantidad de discursos, y compararlas con los discursos de un tópico en específico, como por ejemplo *Pueblos Originarios* que es el más recurrente. En la Figura 32 a continuación, se presenta una visualización para este propósito:

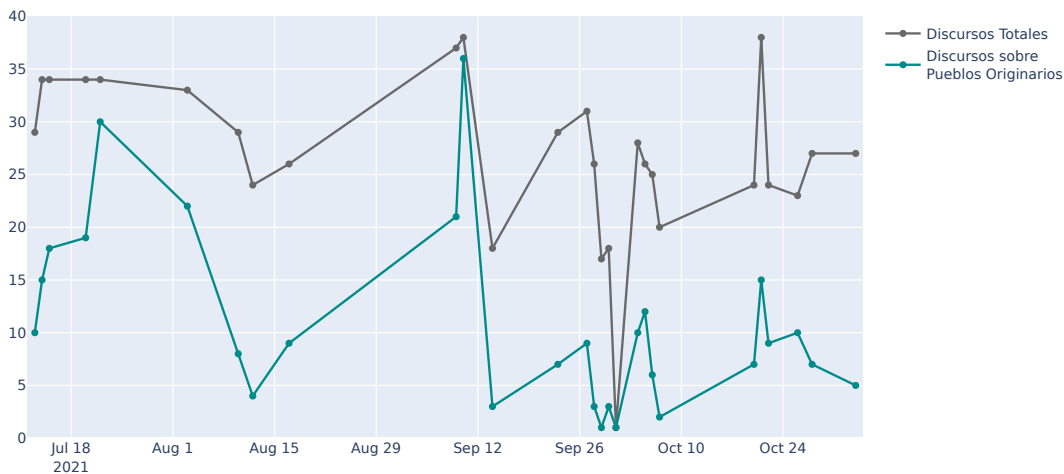


Figura 32: Cantidad de discursos totales y con tópico *Pueblos Originarios* a lo largo del tiempo en al CC.

Fuente: elaboración propia.

Gracias a la Figura 32 podemos evidenciar que en general, el tema más recurrente de la CC, acompañó proporcionalmente a la cantidad de discursos a lo largo del proceso. Esto indica que, la ocurrencia de este tópico fue a lo largo de todo el proceso. Se destaca incluso, que la única sesión con 1 discurso, incluía este tópico, o que una de las sesiones con más discursos (cerca del 12 de septiembre), también fue una de las sesiones con más discursos que incluían a *Pueblos Originarios*.

4.3.2. Comparative Constitution Project

Las constituciones, a pesar de ser la carta fundamental de un país, no son estáticas, pues se pueden realizar Reformas o Enmiendas (siempre y cuando la misma constitución lo permita) para agregar, quitar o editar pasajes que el nuevo contexto requiere. Al contar con esta fuente de datos internacional, podemos realizar una comparación de la fecha de la última nueva constitución de cada país, con la última reforma a esta, si es que se han realizado. Para esto, se utiliza la siguiente consulta sobre la ontología, presente en la Figura 33:

```
SELECT (?nombrePais as ?Pais) ?iso3 ?codeUN (MAX(?año) as ?Año)
WHERE {
  ?const cpc:isConstitutionOf ?pais .
  ?pais cpo:nameListES ?nombrePais;
        cpo:codeISO3 ?iso3;
        cpo:codeUN ?codeUN;
        cpo:isInGroup cpo:World .

  ?const rdf:event_year ?año ;
        rdf:event_type ?tipo .

  FILTER (?tipo IN ('New', 'Interim')) .
  # para considerar enmiendas:
  # FILTER (?tipo IN ('New', 'Interim', 'Amendment')) .
}
Group by ?pais ?nombrePais ?iso3 ?codeUN
ORDER BY ?Año
```

Figura 33: Consulta SPARQL para obtener el año de la última constitución de cada país. Se incluye un comentario en caso de que se quiera considerar las enmiendas y reformas.

Fuente: elaboración propia.

Para realizar un mejor análisis, en las Figuras 34 y 35 se presentan ambos casos: el año de la última constitución sin considerar enmiendas, y considerando enmiendas, respectivamente. En estas Figuras, se destacan sectores como África, que en su mayoría de países se presenta con una nueva constitución en los últimos 5 años. Por el contrario, América del Norte presenta constituciones creadas antes de 1940. Por otra parte, la parte de Europa Oeste su última nueva constitución ronda entre los años 1950 y 1970, no así América del Sur que ronda entre 1975 y 2000. Al comparar con los años considerando las enmiendas y reformas (Figura 35), podemos evidenciar que en todo el mundo predominan los años mayores a 1990, indicando que a pesar que algunas constituciones fueron creadas hace décadas, estas se mantienen actualizadas. Aunque hay excepciones, como Japón, que no ha realizado actualizaciones a su constitución.

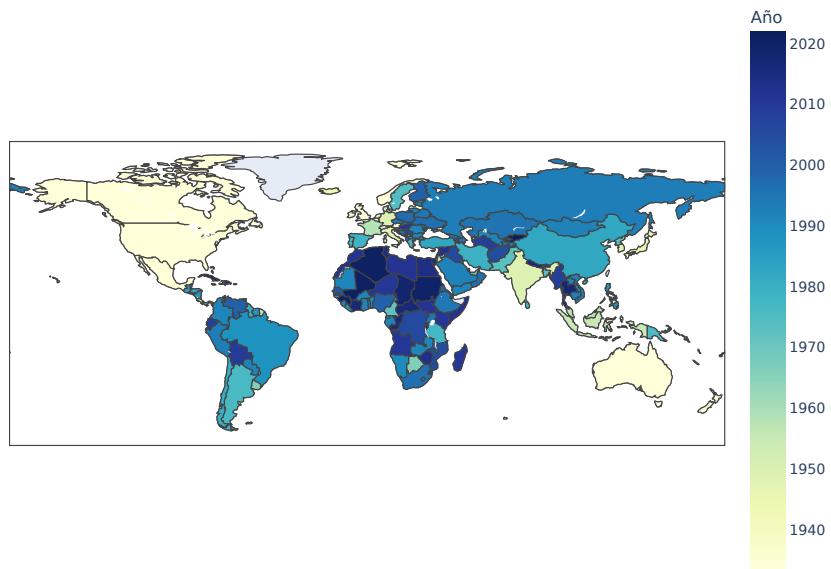


Figura 34: Año de la creación de la última constitución a nivel mundial.
Fuente: elaboración propia.

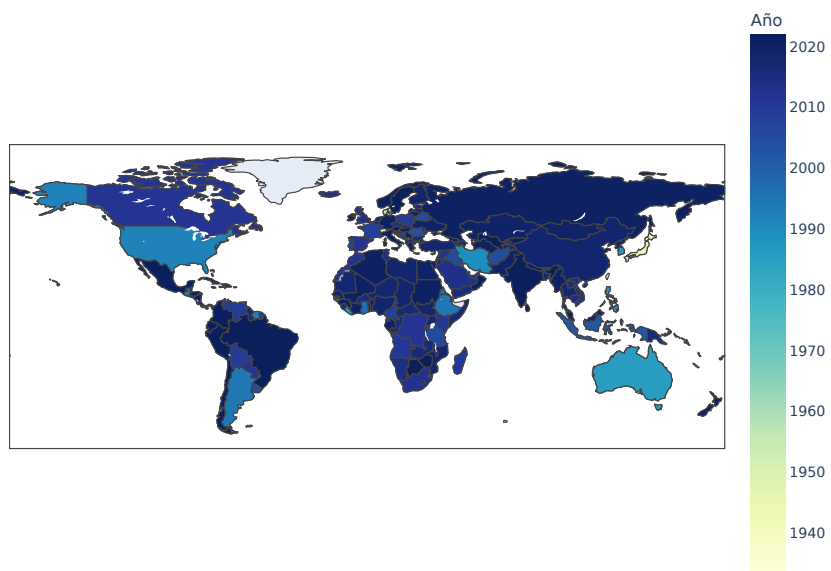


Figura 35: Año de la última actualización de constitución a nivel mundial.
Fuente: elaboración propia.

La ontología, también permite filtrar según distintos parámetros, lo que facilita la comparación de datos en base a estos. Por ejemplo, si quisiéramos comparar qué tópicos son los más recurrentes en las constituciones de América y Europa, podemos utilizar la consulta presente en la Figura 36 a continuación (se consideran sólo los 10 tópicos más recurrentes):

```

Select ?nombreTopico (COUNT(?Evento) as ?constituciones)
WHERE {
  ?Evento rdf:type ?topico .
  ?topico rdfs:subClassOf* cpo:Topics ;
         rdfs:label ?nombreTopico .

  FILTER(LANG(?nombreTopico) = "es")
  {
    SELECT (MAX(?evento) as ?Evento) ?nombrePais (MAX(?año) as ?Año)
    WHERE {
      ?evento rdf:type rdf:event ;
             rdf:event_year ?año;
             rdf:event_type ?tipo;
             cpc:isConstitutionOf ?pais.

      ?pais cpo:nameListES ?nombrePais ;
           cpo:isInGroup ?continente.

      FILTER(?tipo IN ('New', 'Amendment','Interim'))
      FILTER(?continente = cpo:Americas)
      # para Europa:
      # FILTER(?continente = cpo:Europe)
    }
    GROUP BY ?nombrePais
    ORDER BY (?nombrePais)
  }
}
Group By (?nombreTopico)
ORDER BY DESC (?constituciones)
LIMIT 10

```

Figura 36: Consulta SPARQL para obtener los 10 tópicos más recurrentes según región del mundo. Se considera la última actualización de constitución de cada país.

Fuente: elaboración propia.

Para la visualización y análisis de estos datos, se proponen los arboles de jerarquía de la Figura 37 para América y la Figura 38 para Europa.

En ambos Continentes predomina el tópico *Requisitos de los representantes de la primera cámara*, pero ya desde el el segundo nivel de jerarquía hay diferencias en el orden. En América es más predominante el tópico de *Estructura de las cámaras legislativas* que el tópico *Derecho Internacional*, que es predominante en Europa, incluso en Amperica este último tópico no está dentro de los 10 más recurrentes. Sucede lo mismo con los tópicos *Ratificación de Tratados* e *Inmunidad de legisladores*. Por el contrario, los tópicos que son recurrentes en América que no aparecen dentro de los 10 más recurrentes de Europa son: *Libertad religiosa*, *Aprobación de legislación general* y *Libertad de asociación*. Esto pude dar paso a un análisis cultural y social de ambos continentes, y a investigar por qué se dan estas diferencias y en estos tópicos en específico. Así mismo, del porqué de las similitudes de tópicos relevantes en ambas regiones, que en su mayoría describen estructura y procedimientos respecto al poder Legislativo.

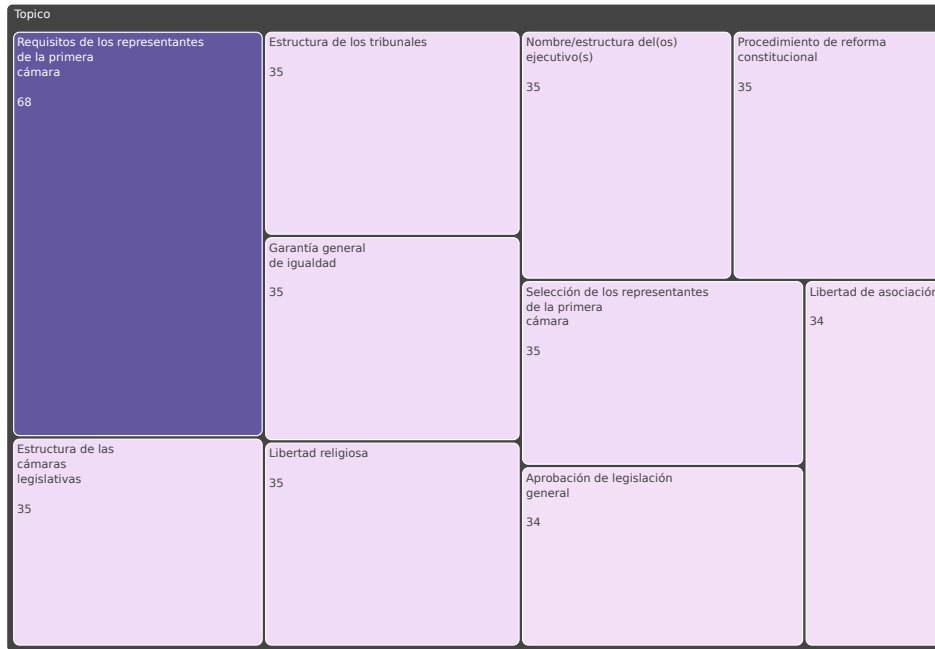


Figura 37: Top 10 tópicos más recurrentes en las constituciones actuales de América.
Fuente: elaboración propia.

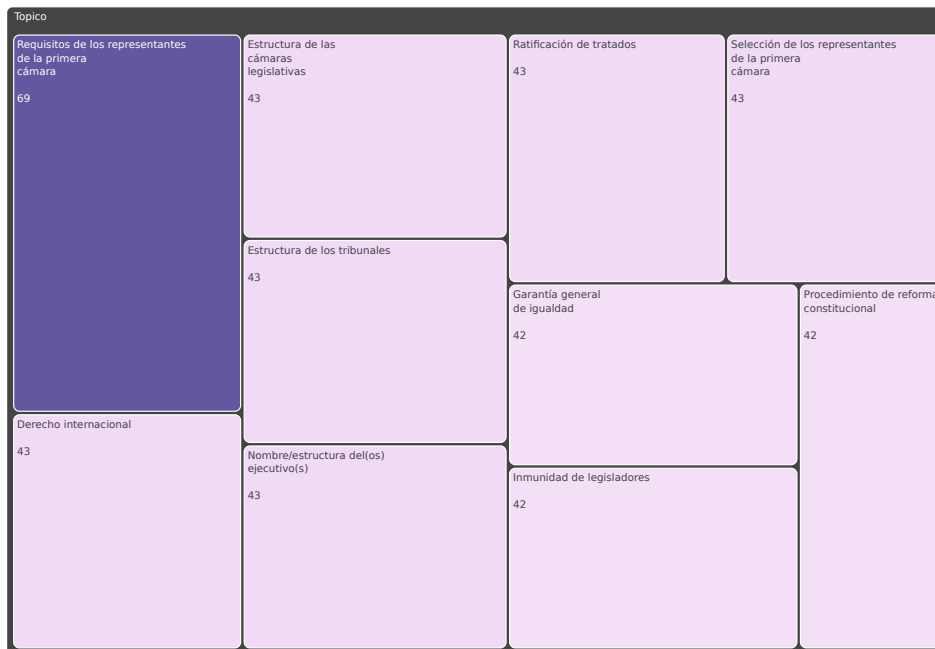


Figura 38: Top 10 tópicos más recurrentes en las constituciones actuales de Europa.
Fuente: elaboración propia.

Así cómo se pueden realizar comparaciones internacionales con otras regiones, también es posible evaluar datos en un ámbito más cercano. En la Figura 39, se presenta la consulta a la ontología para obtener la cantidad de eventos constitucionales (constituciones nuevas, internas o enmiendas) en América del Sur desde el año 2000.

```

SELECT ?nombrePais ?codeIS03 (COUNT(?constitucion) as ?constituciones)
WHERE{
  ?constitucion rdf:type rdf:event ;
                rdf:event_type ?tipo ;
                rdf:event_year ?año ;
                cpc:isConstitutionOf ?pais.

  ?pais cpo:nameListES ?nombrePais ;
        cpo:codeIS03 ?codeIS03 ;
        cpo:isInGroup ?continente .

  FILTER(?tipo IN ("New","Interim","Amendment"))
  FILTER(?año >= "2000")
  FILTER(?continente = cpo:southern_America)
}
GROUP BY ?pais ?nombrePais ?codeIS03
    
```

Figura 39: Consulta SPARQL para obtener los eventos constitucionales en América del Sur desde el año 2000.

Fuente: elaboración propia.

Para el análisis de los datos, se propone la visualización de la Figura 40 a continuación:

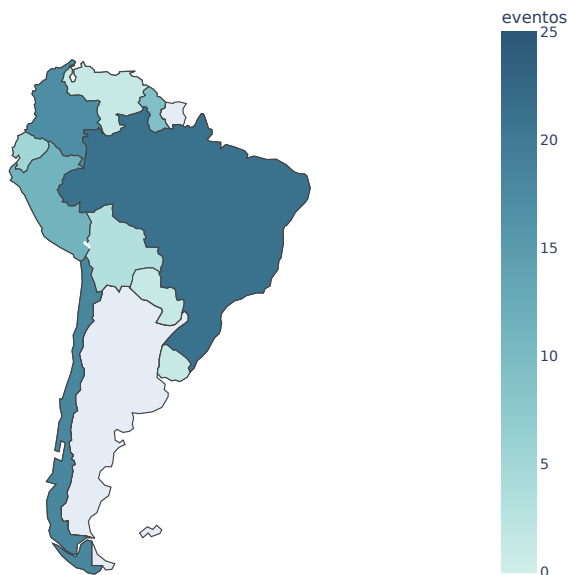


Figura 40: Cantidad de eventos constitucionales en América del Sur desde el año 2000.

Fuente: elaboración propia.

Al observar la Figura 40, se evidencia que en los últimos 20 años han existido más de 40 eventos constitucionales entre Chile y Brasil, que presentan una gran diferencia entre sus

vecinos. Por ejemplo Argentina no presenta datos, es decir, no presenta eventos constitucionales después del año 2000. La mayoría de países está por debajo de los 15 eventos. Este dato nos ayuda a buscar relaciones entre los eventos constitucionales y la situación social, cultural y económica de cada país, al compararse con sus pares de la región sur del continente.

4.3.3. Integración

Al integrar ambas fuentes de datos, permite la obtención de datos de puntos de vistas más amplios, y así realizar un análisis del proceso constitucional más completo, en este caso, sobre los tópicos tratados dentro de la CC.

Ya se observó, en la Figura 30, que el tema *Pueblos Originarios* fue el más recurrente dentro de la CC. La integración realizada, nos permite revisar qué países actualmente contienen este tópico dentro de sus constituciones. Esto, a través de la consulta mostrada en la Figura 41 a continuación:

```

SELECT ?codeIS03 ?nombrePais ?Año
WHERE{
  ?topico cc:incluidoEn ?Evento;
         cc:nombreTopico ?nombreTopico .

  FILTER(?nombreTopico = "Pueblos Originarios")
  {
    SELECT (MAX(?evento) as ?Evento) ?nombrePais ?codeIS03 (MAX(?año) as ?Año)
    WHERE {
      ?evento rdf:type rdf:event ;
              rdf:event_year ?año;
              rdf:event_type ?tipo;
              cpc:isConstitutionOf ?pais.

      ?pais cpo:nameListES ?nombrePais;
            cpo:codeIS03 ?codeIS03.
      FILTER(?tipo IN ("New", "Amendment", "Interim"))
    }
  }
GROUP BY ?nombrePais ?codeIS03
ORDER BY (?nombrePais)
}
}

```

Figura 41: Consulta SPARQL para obtener los países que actualmente implementan el tópico *Pueblos Originarios* en sus constituciones.

Fuente: elaboración propia.

Para visualizar estos datos, se presenta la Figura 42 a continuación:

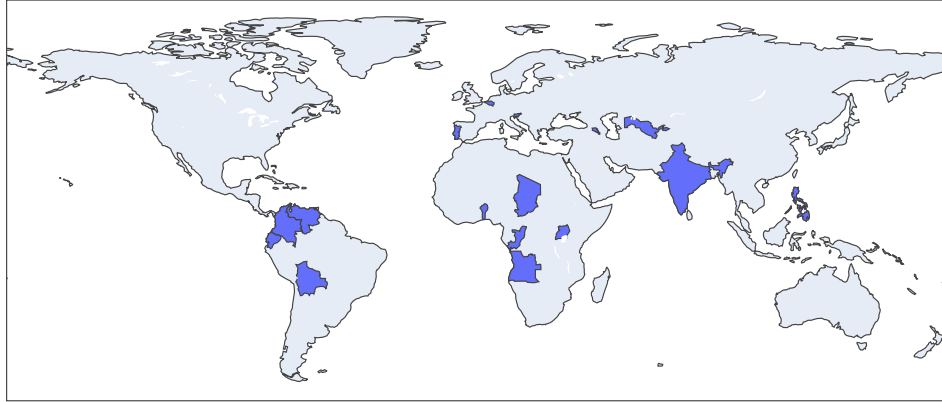


Figura 42: Países que actualmente implementan el tópico *Pueblos Originarios* en sus constituciones.

Fuente: elaboración propia.

Según lo observado en las Figuras 30 y 42, a pesar de que el tópico *Pueblos Originarios* es el más recurrente dentro de la CC, este no está presente en muchos países a nivel mundial. Lo que puede implicar tres situaciones: que es un tópico relativamente nuevo y por eso sólo ha sido considerado por unos pocos; que es un tópico que depende del contexto interno del país, ya sea por historia u otras razones, y sólo se considera en aquellos países donde es realmente relevante; o que el tópico fue considerado hace años a nivel mundial, pero en las actualizaciones más recientes a la constitución este tópico se ha descartado, lo que indica que los países que lo incluyen no están actualizados. En cualquiera de los casos, es interesante investigar el porqué de esta diferencia entre los considerados en la CC y los considerados en el resto del mundo respecto a este tópico.

En base a esto último, también es interesante saber qué temas de la CC son los más recurrentes en las constituciones actuales a nivel mundial. Para consultar esto a la ontología, se utiliza el código mostrado en la Figura 43, y para realizar un mejor análisis se propone la visualización de la Figura 44.

Se evidencia así, que los tópicos relacionados a Economía, Cultura y Estado son los más recurrentes a nivel mundial. Destaca también, que en su mayoría, los tópicos más recurrentes están relacionados a los poderes del estado, que por definición, son principalmente definidos en las constituciones. Estos datos, pueden guiar la discusión del porqué estos son los más recurrentes, o iniciar el análisis de qué tan recurrentes son estos tópicos dentro de la CC, y así evaluar si los tópicos de la CC deben corresponder a los más relevantes a nivel internacional, o es necesario discutir sobre aquellos que menos apariciones tienen en el resto del mundo.

```

SELECT ?nombreTopico (COUNT(?Evento) as ?constituciones)
WHERE{
  {
    ?topico rdf:type cc:Topico ;
      cc:nombreTopico ?nombreTopico .

    ?Evento cc:incluyeParteDe ?topico.
  }
  UNION
  {
    ?topico rdf:type cc:Topico ;
      cc:nombreTopico ?nombreTopico .

    ?topico cc:incluidoEn ?Evento .
  }
  {
    SELECT (MAX(?evento) as ?Evento) ?nombrePais (MAX(?año) as ?Año)
    WHERE {
      ?evento rdf:type rdf:event ;
        rdf:event_year ?año;
        rdf:event_type ?tipo;
        cpc:isConstitutionOf ?pais.

      ?pais cpo:nameListES ?nombrePais;
        FILTER(?tipo IN ("New", "Amendment", "Interim"))
    }
    GROUP BY ?nombrePais
    ORDER BY (?nombrePais)
  }
}
GROUP BY ?nombreTopico
ORDER BY DESC(?constituciones)
LIMIT 20

```

Figura 43: Consulta SPARQL para obtener la cantidad de apariciones de los tópicos de la CC en las constituciones actuales del mundo.

Fuente: elaboración propia.

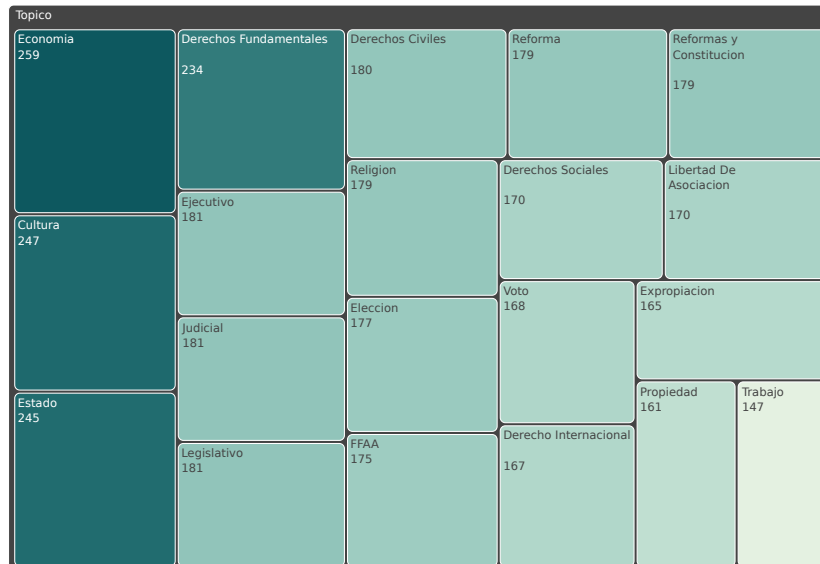


Figura 44: Cantidad de apariciones de los tópicos de la CC en las constituciones actuales del mundo.

Fuente: elaboración propia.

Otro punto de vista que ayuda a la discusión, es la cantidad de tópicos de la CC que son considerados por distintas regiones, y así comparar las diferencias y similitudes entre estas regiones bajo este contexto. Para consultar la cantidad de tópicos de la CC relacionados a la constitución actual de cada país en una región en específico, se utiliza el código presente en la Figura 45 a continuación:

```

SELECT (COUNT(?topico) as ?cantidadTopicos)
WHERE{
  {
    ?topico rdf:type cc:Topico .
    ?Evento cc:incluyeParteDe ?topico.
  }
  UNION
  {
    ?topico rdf:type cc:Topico .
    ?topico cc:incluidoEn ?Evento .
  }
  {
    SELECT (MAX(?evento) as ?Evento) ?nombrePais (MAX(?año) as ?Año)
    WHERE {
      ?evento rdf:type rdf:event ;
      rdf:event_year ?año;
      rdf:event_type ?tipo;
      cpc:isConstitutionOf ?pais.

      ?pais cpo:nameListES ?nombrePais;
      cpo:isInGroup ?grupo .
      FILTER(?tipo IN ("New", "Amendment", "Interim")).
      FILTER(?grupo = cpo:Europe).
      # para America:
      # FILTER(?grupo = cpo:Americas).
    }
    GROUP BY ?nombrePais ?codeISO3
    ORDER BY (?nombrePais)
  }
}
GROUP BY ?nombrePais
ORDER BY DESC(?cantidadTopicos)

```

Figura 45: Consulta SPARQL para obtener la cantidad de tópicos de la CC relacionados a la constitución actual de cada país en una región en específico.

Fuente: elaboración propia.

Para realizar una comparación de distribución de las cantidades de tópicos relacionados a las constituciones actuales de América y Europa, se propone la visualización de la Figura 46. En esta, se presenta una distribución más amplia en América, concentrando el 50 % de sus datos entre 26 y 36 con una mediana de 32, en cambio los datos de Europa concentran su 50 % entre 26 y 34 con una mediana de 33. Indicando así, que en Europa hay una concentración entre 34 y 35 tópicos de la CC abordados por sus constituciones, esto de demuestra dado el rango del cuarto cuartil, en Europa es de 32 a 34, en cambio en América este rango es desde 31 a 36. Interesante es investigar qué tópicos de la CC son los más recurrentes en Europa y América, y así complementar la información obtenida.

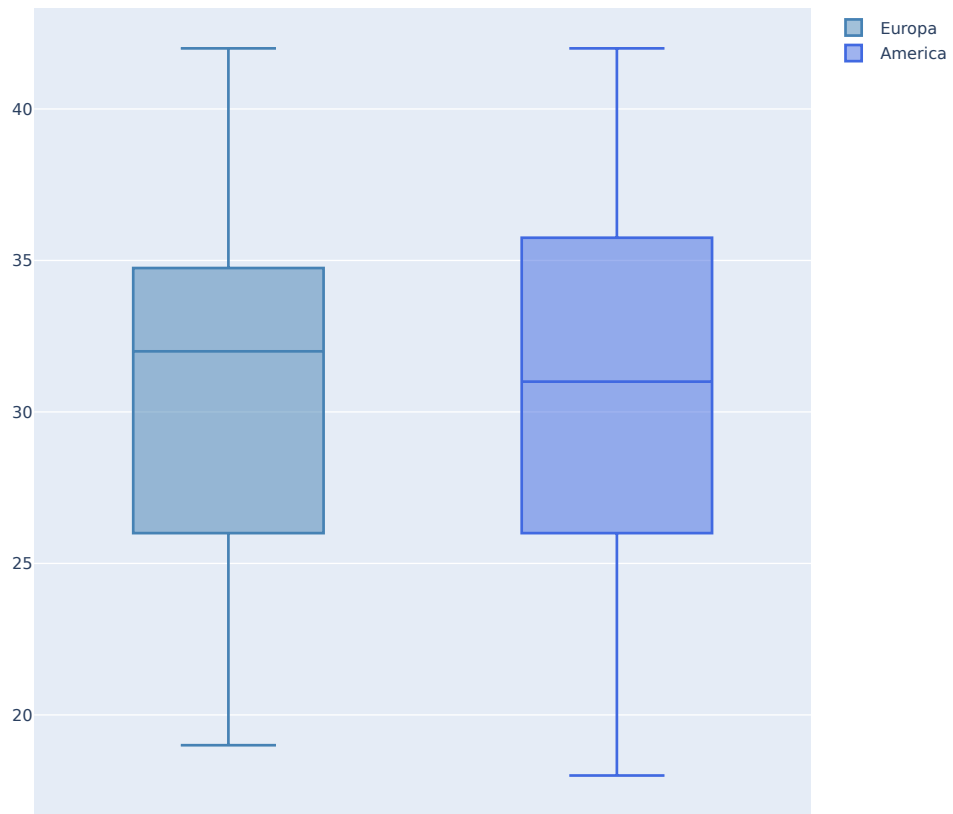


Figura 46: Distribución de la cantidad de tópicos de la CC relacionados a la constitución actual de cada país en Europa y América.
Fuente: elaboración propia.

CAPÍTULO 5

CONCLUSIONES

Cada minuto que pasa se crean grandes volúmenes de datos alrededor del mundo, y esta cantidad aumenta de forma exponencial cada año. Datos de toda índole, desde interacciones en redes sociales, compras en un supermercado, hasta metadatos de videojuegos. Esto conlleva no sólo problemas de almacenamiento, sino también problemas de análisis en la toma de decisiones, aún más cuando estos datos pueden entregar información de procesos políticos, sociales o culturales dentro de un país.

Las dificultades del análisis se ven directamente influenciadas por la gran cantidad de datos a procesar y la falta de mecanismos y estructuras estandarizadas que permitan realizar esta tarea, y además integrarlas con otras fuentes.

Un caso de esto, fue el proceso constitucional en Chile entre los años 2021 y 2022, en el que por primera vez en la historia del país nació una propuesta de constitución creada por ciudadanos escogidos por votación popular. Proceso, que a pesar de contar con plataformas oficiales de información, como las creadas por la Biblioteca del Congreso Nacional, o entidades privadas que analizaron los datos que rodeaban este proceso, como lo realizado por el Instituto Milenio Fundamento de los Datos, todas estas trabajaban con datos independientes y no hubo una integración de datos que permita realizar un análisis del proceso en mayor profundidad. Esto se dio principalmente por la falta de mecanismos que permitan la integración de datos actuales y futuros de la Convención Constitucional.

Tal como se comentó al principio, las principales razones de esta problemática son el gran volumen de datos que día a día se generaban acerca del proceso y la falta de una estructura en las plataformas creadas que permita la integración con fuentes externas. A esto, se le suma la falta de mecanismos claros establecidos, dado que el proceso es primera vez que ocurría en Chile, y las plataformas no estaban preparadas para esto.

La falta de estos mecanismos y estructuras, impacta directamente en el ciudadano, pues al no poder integrar los datos, no es posible realizar comparaciones con fuentes externas o internacionales, provocando un escasez de información al alcance de la población, y la imposibilidad de creación de visualizaciones que permitan un análisis visual en profundidad del proceso.

En este trabajo se definió, diseñó e implementó la integración de datos de la Convención Constitucional con los datos de *Comparative Constitute Project*, proyecto dedicado a la recolección y clasificación de eventos constitucionales de todos los países del mundo. Con el objetivo de permitir futuras integraciones, y la reducción de complejidad en la obtención de datos, esta estructura se basó en una ontología RDF bajo el concepto de Web Semántica, siguiendo los lineamientos internacionales recomendados por la W3C e implementado técnicas de búsqueda actuales para la obtención de datos, y posterior construcción de visua-

lizaciones que aporten nuevo conocimiento.

Para esta implementación, se utilizó como base el trabajo realizado por Carina Flores, estudiante de la Universidad Técnica Federico Santa María, que en conjunto con el Instituto Milenio Fundamentos de los Datos, construyeron una ontología RDF para dar soporte a los datos de la Convención Constitucional, que incluyen datos relevantes de los candidatos, los constitucionales e incluyen la clasificación en tópicos de los discursos realizados en cada una de las sesiones de este proceso, gracias al previo análisis mediante procesamiento de texto. A esta base, se le sumó la definición y diseño de una estructura para dar soporte a los datos provenientes de *Comparative Constitute Project*. Esta estructura, se basó en el análisis de estructura de los datos provenientes de esta fuente, que posteriormente, fueron procesados para poblar la nueva estructura planteada.

La integración en sí, se basó en la homologación de tópicos definidos en ambos conjuntos, es decir, los tópicos de la Convención Constitucional para clasificar discursos fueron mapeados a su respectivo tópico en *Comparative Constitute Project* para clasificar constituciones. En este contexto se definieron dos casos: el caso en que el tópico en *Comparative Constitute Project* es igual o más general que el tópico en la Convención Constitucional, y el caso contrario, en que el tópico en la Convención Constitucional es más general que el tópico en *Comparative Constitute Project*. En la ontología propuesta, se dio soporte a estos casos mediante dos propiedades de relación entre los eventos constitucionales y los tópicos de la convención.

Para el diseño de la ontología, y proceso de almacenamiento y consulta, se utilizaron herramientas avaladas por la comunidad y empleadas por organizaciones internacionales, que al igual que este trabajo, siguen los lineamientos internacionales propuestos por la W3C. Estas herramientas, además, facilitan la extracción de las respuestas a las consultas realizadas, para su posterior procesamiento y visualización.

La validación de la ontología propuesta se centró en la aplicación al caso concreto planteado, y en la demostración empírica de los procesos de consulta y visualización de datos. Para esto, se desarrollaron consultas que permiten obtener conocimiento no trivial en tres categorías: consultas sólo sobre la Convención Constitucional, consultas sólo el repositorio histórico de *Comparative Constitute Project* y consultas que impliquen la obtención de datos de ambas fuentes en conjunto. Además, se consideró la demostración que estos datos pueden obtenerse utilizando la técnica de búsqueda facetada, que es directamente soportada por los grafos RDF. En lo presentado en esta validación, es posible observar que la ontología permite la comparación local e internacional respecto a varios atributos, ya sean temporales o estáticos, además de permitir la evaluación de lo trabajado por la convención respecto a cómo se relacionan los tópicos hablados, con los tópicos presentes en las constituciones en el resto del mundo. Así mismo, permite iniciar debates interesantes, e investigaciones que permitan explicar el porqué de ciertos resultados visualizados.

Se espera que este trabajo sirva como base para toda iniciativa que busque la obtención de nuevos conocimientos a partir de este punto de vista implementado. Siempre es bueno tener en cuenta fuentes históricas y actuales internacionales que nos permitan evaluar y

comparar los procesos locales de cada país, más aún, cuando este proceso es para definir una propuesta de los fundamentos legales, culturales y sociales que rigen al estado de cada nación. También se espera, que este trabajo sea la base para futuras integraciones con otras fuentes de datos tanto locales como internacionales, y así sumar nuevos puntos de vista relevantes a los procesos actuales y futuros.

5.1. Trabajo Futuro

A día de hoy, se está realizando un nuevo proceso constitucional en Chile, con nuevos representantes, pero bajo la misma estructura. Se plantea la necesidad de actualizar los datos de la BCN con los nuevos que se han generado y se generarán en este nuevo proceso. Así, los ciudadanos contarán con esta fuente de información a la hora de evaluar los resultados, y así votar en las próximas elecciones más informados.

Por otro lado, a pesar de trabajar con herramientas y lenguajes estandarizados, para consultar sobre la ontología se requieren ciertos conocimientos previos, no sólo sobre SPARQL, sino también sobre la ontología en sí. Esto, dificulta el acceso y la utilización de lo construido por un usuario común. Para evitar esto, se propone la implementación de un sistema de Preguntas y Respuestas automatizado, que sea capaz de traducir preguntas en lenguaje natural a SPARQL, además del diseño de una interfaz capaz de indicar todos los filtros y valores de búsqueda disponibles.

Por último, esta memoria se centró en la construcción de una ontología que permita la integración de ambas fuentes evaluadas, pero es interesante evaluar el alcance que se puede tener en las consultas sobre la ontología y el conocimiento adicional que se puede obtener analizando en profundidad las visualizaciones construidas, o investigar qué otras visualizaciones se pueden construir con los datos disponibles.

5.2. Reflexiones

Trabajar sobre grandes volúmenes de datos siempre es un desafío, más aún cuando no hay estructuras definidas para dar soporte al análisis de estos y a la integración con otras fuentes. Durante el desarrollo de esta memoria se evidenció la importancia de la implementación de lineamientos que permitan la publicación de datos facilitando su accesibilidad, reutilización y sobre todo su integración. No basta con disponer los datos, si estos no siguen los estándares internacionales y no están preparados para incorporación de nuevas fuentes. En este contexto, se destaca que a pesar de las dificultades anteriormente planteadas, se logró implementar una ontología robusta que permite la obtención de nuevo conocimiento. En su mayoría, gracias a la estandarización y estructuración de los datos en ambas fuentes integradas. Recalcar que, los datos existen, en todo ámbito imaginable, y no podemos permitir que las faltas de metodologías y estructuras sólidas nos limiten el nuevo conocimiento.

ANEXOS

A. Datos de la Biblioteca del Congreso Nacional

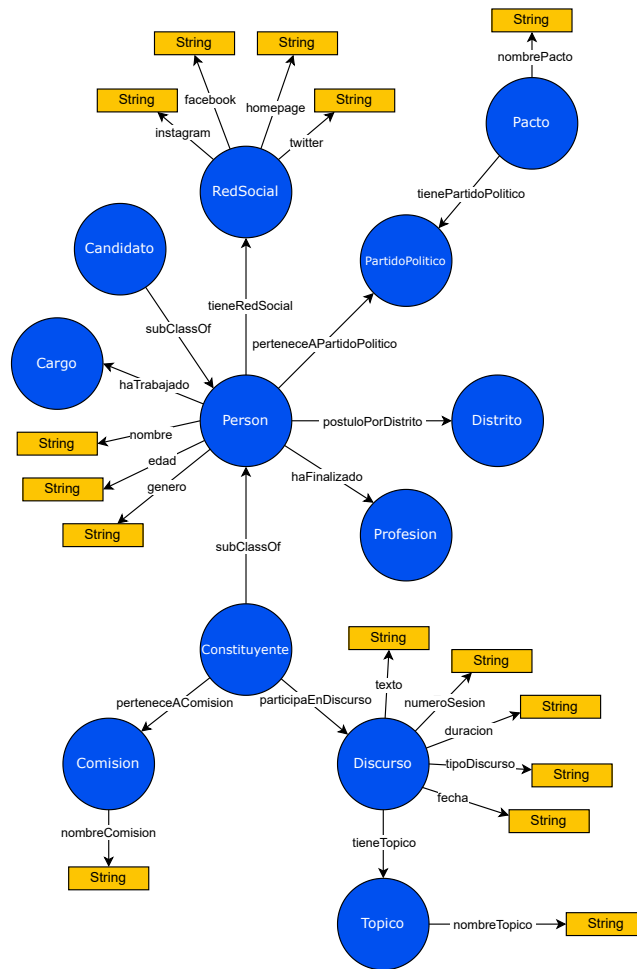


Figura 47: Representación gráfica de la ontología utilizada para almacenar los datos de la BCN.

Fuente: elaboración propia

Tabla 1: Taxonomía de los tópicos identificados en los discursos de la CC. Fuente: elaboración propia.

| Nombre de Tópicos | |
|--------------------------|---|
| Comisiones | Anti Corrupción Derechos Humanos Telecomunicación Verdad |
| Cultura | Ciencia Historia Símbolos |
| Derecho Internacional | |
| Derechos Fundamentales | Derechos Civiles Familia Libertad de Asociación |
| Derechos Sociales | Adulto Mayor Agua Discapacidad Educación MedioAmbiente Menor de Edad Salud Trabajo |
| Economía | Banco Expropiación Mercado Ocupación Presupuesto Propiedad |
| Estado | Ejecutivo Judicial Legislativo Territorio |
| FFAA | |
| Igualdad | Color de Piel Discapacidad Edad Estatus Social Familia |

Tabla 1 Continuación: Taxonomía de los tópicos identificados en los discursos de la CC. Fuente: elaboración propia.

| Nombre de Tópicos | |
|--------------------------|--|
| | Finanza Marital Orientación Sexual |
| Medios de Comunicación | Censura Comunicación Reputación |
| Origen | Deportación Frontera Nacionalidad Pueblo Originario Religión |
| Partidos Políticos | Derecho a Partido Partido Regulación a Partido |
| Reformas y Constitución | Deber Reforma |
| Sistema Electoral | Censo Distrito Eleccion Voto |

B. Datos de Comparative Constitution Project

Tabla 2: Taxonomía de los tópicos de CCP con dos niveles de profundidad. Fuente: elaboración propia.

| | Nombre de Tópicos |
|------------------------|--|
| Enmienda | Disposiciones constitucionales no sujetas a reforma Procedimiento de reforma constitucional |
| Cultura e Identidad | Grupos Indígenas Idioma Nacionalidad Raza y Etnicidad Religión |
| Elecciones | Normas y Procedimientos Electorales Partidos Políticos Referendos e Iniciativas Sufragio y Participación Electoral Supervisión Electoral |
| Ejecutivo | Estructura del Ejecutivo Fuerzas Armadas Gabinete Independencia y Facultades del Ejecutivo Jefe de Estado Jefe de Gobierno |
| Federalismo | Estructura del Estado Facultad Legislativa Secesión y Anexión |
| Normas Internacionales | Política Exterior Referencias Explícitas al Derecho Internacional Tratados |
| Judicial | Autonomía y Facultades del Judicial Corte Suprema Estructura del Poder Judicial Revisión Judicial Tribunal Constitucional Tribunal Electoral Tribunales administrativos Tribunales ordinarios |
| Legislativo | Destitución y Sustitución |

Tabla 2 Continuación: Taxonomía de los tópicos de CCP con dos niveles de profundidad. Fuente: elaboración propia.

| Nombre de Tópicos | |
|--------------------------|--|
| | Estructura de las cámaras legislativas Independencia y Facultades del Legislativo Legislación Legislación Especial Primera Cámara Reglas y Restricciones Legislativas Segunda Cámara |
| Principios y Símbolos | Definición del Estado y Símbolos Principios Básicos |
| Regulación y Supervisión | Elecciones Medios de comunicación Social Issues Órganos Autónomos y Comisiones |
| Derechos y Deberes | Deberes Ciudadanos Deberes Generales Derechos Civiles y Políticos Derechos Económicos Derechos Procesales Derechos Sociales Derechos a la Integridad Física Derechos de Igualdad, Género y Minoría Ejecución |
| Secciones Especiales | Disposiciones transitorias Preámbulo |

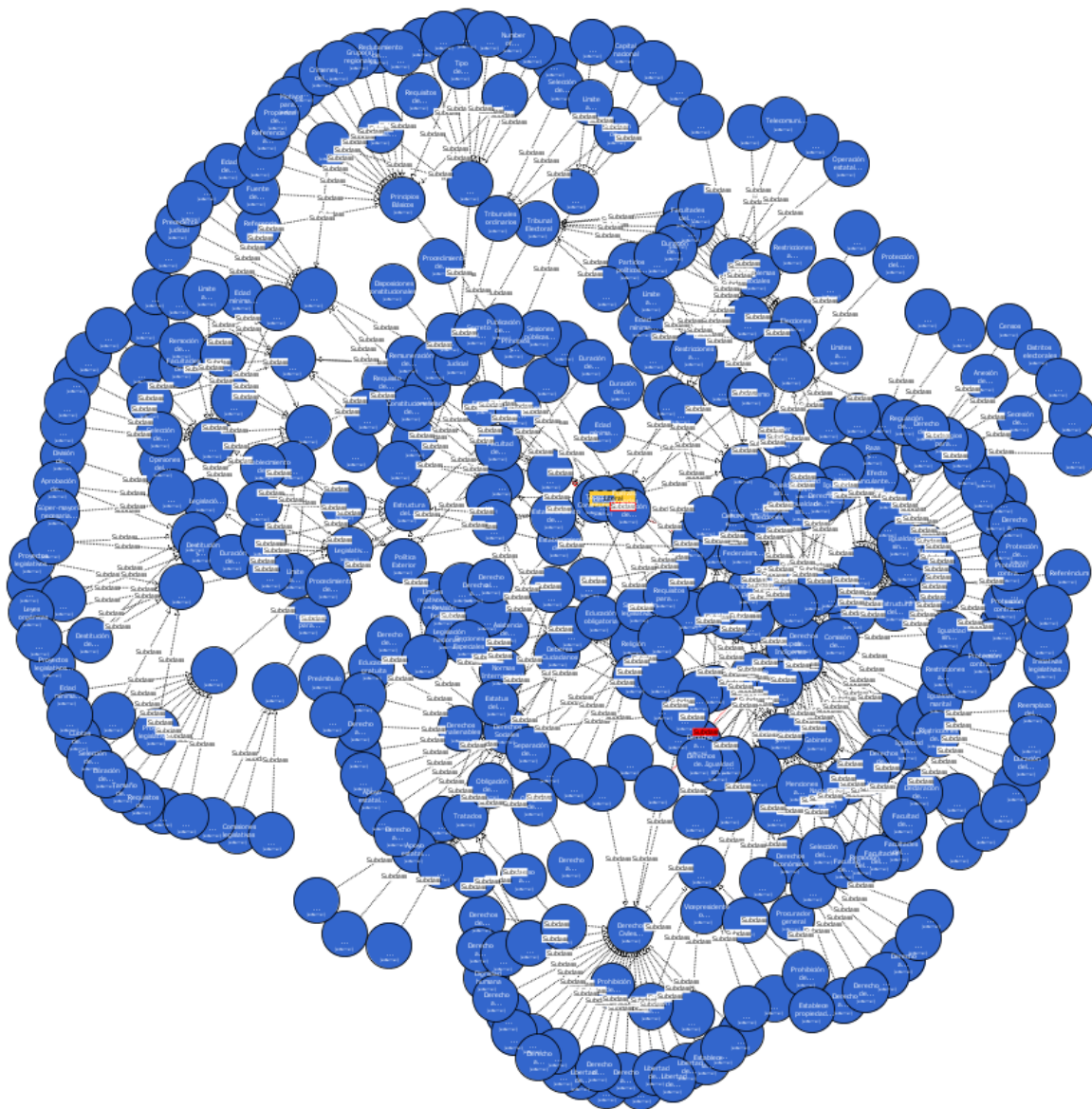


Figura 48: Representación gráfica de la estructura del subgrafo de los tópicos de CCP.
Fuente: elaboración propia a través de [Musen, 2015]

C. Estructura de integración

Tabla 3: Homologación de tópicos en el que el tópico CCP es igual o más general que el tópico BCN. Fuente: elaboración propia.

| | Nombre Tópico BCN | Nombre clase CCP |
|------------------------|---|---|
| Comisiones | Anti Corrupción Derechos Humanos Telecomunicación Verdad | Comisión anti-corrupción Comisión de derechos humanos Comisión de telecomunicaciones Comisión de verdad y reconciliación |
| Cultura | Ciencia Historia | Referencia a la ciencia Referencia a la historia del país |
| Derecho Internacional | | Derecho Internacional |
| Derechos Fundamentales | Derechos Civiles Libertad de Asociación Agua Salud | Derechos Civiles y Políticos Libertad de asociación Derecho al agua Derecho a la salud |
| Economía | Banco | Banco central |
| Estado | Ejecutivo Judicial Legislativo | Ejecutivo Judicial Legislativo |
| FFAA | | Fuerzas Armadas |
| Igualdad | Color de Piel Discapacidad Edad Estatus Social Familia Finanza Marital Orientación Sexual Pueblo Originario Religión | Igualdad sin distinción de color de piel Igualdad de personas con discapacidades Igualdad sin distinción de edad Igualdad sin distinción de estatus social Igualdad sin distinción de origen familiar Igualdad sin distinción de estatus financiero Igualdad marital Igualdad sin distinción de orientación sexual Grupos Indígenas Religión |
| Partidos Políticos | Regulación a Partido | Regulación de partidos políticos |
| Sistema Electoral | Censo Distrito Elección Voto | Censos Distritos electorales Elecciones Sufragio y Participación Electoral |

Tabla 4: Homologación de tópicos en el que el tópico BCN es más general que el tópico CCP. Fuente: elaboración propia.

| | Nombre Tópico BCN | Nombre clase CCP |
|------------------------|------------------------|---|
| Cultura | Símbolos | Himno Nacional Bandera Nacional |
| Derechos Fundamentales | Familia | Derecho a formar una familia |
| Derechos Sociales | Adulto Mayor | Apoyo estatal para adultos mayores |
| | Discapacidad | Apoyo estatal para personas con discapacidades |
| | Educación | Acceso a la educación superior Educación gratuita Educación obligatoria |
| | Medio Ambiente | Protección del medio ambiente |
| | Menor de Edad | Apoyo estatal para menores de edad Límites relativos al empleo de menores de edad Derechos de los menores de edad |
| Economía | Trabajo | Derecho al trabajo Derecho a un ambiente laboral seguro Derecho a huelga Derecho a asociación de sindicatos |
| | Expropiación | Protección contra la expropiación |
| | Mercado | Derecho a un mercado competitivo |
| | Ocupación | Derecho a escoger una ocupación |
| | Presupuesto | Presupuesto equilibrado Proyectos legislativos presupuestarios |
| Estado | Propiedad | Derecho a contar con propiedad Derecho a transferencia de propiedad Establece propiedad intelectual |
| | Territorio | Secesión y Anexión |
| | Medios de Comunicación | Censura Comunicacion Reputacion |
| | Origen | Comisión de telecomunicaciones Telecomunicaciones |
| | | Deportación |
| | Frontera | Restricciones de entrada y salida |

Tabla 4 Continuación: Homologación de tópicos en el que el tópico BCN es más general que el tópico CCP. Fuente: elaboración propia.

| Nombre Tópico BCN | Nombre clase CCP |
|-------------------------|--|
| Nacionalidad | Requisitos para nacionalización Requisitos para obtener nacionalidad por nacimiento |
| Partidos Políticos | Derecho a formar partidos políticos Restricciones a partidos políticos |
| Reformas y Constitución | Deber de obedecer la constitución Enmienda |

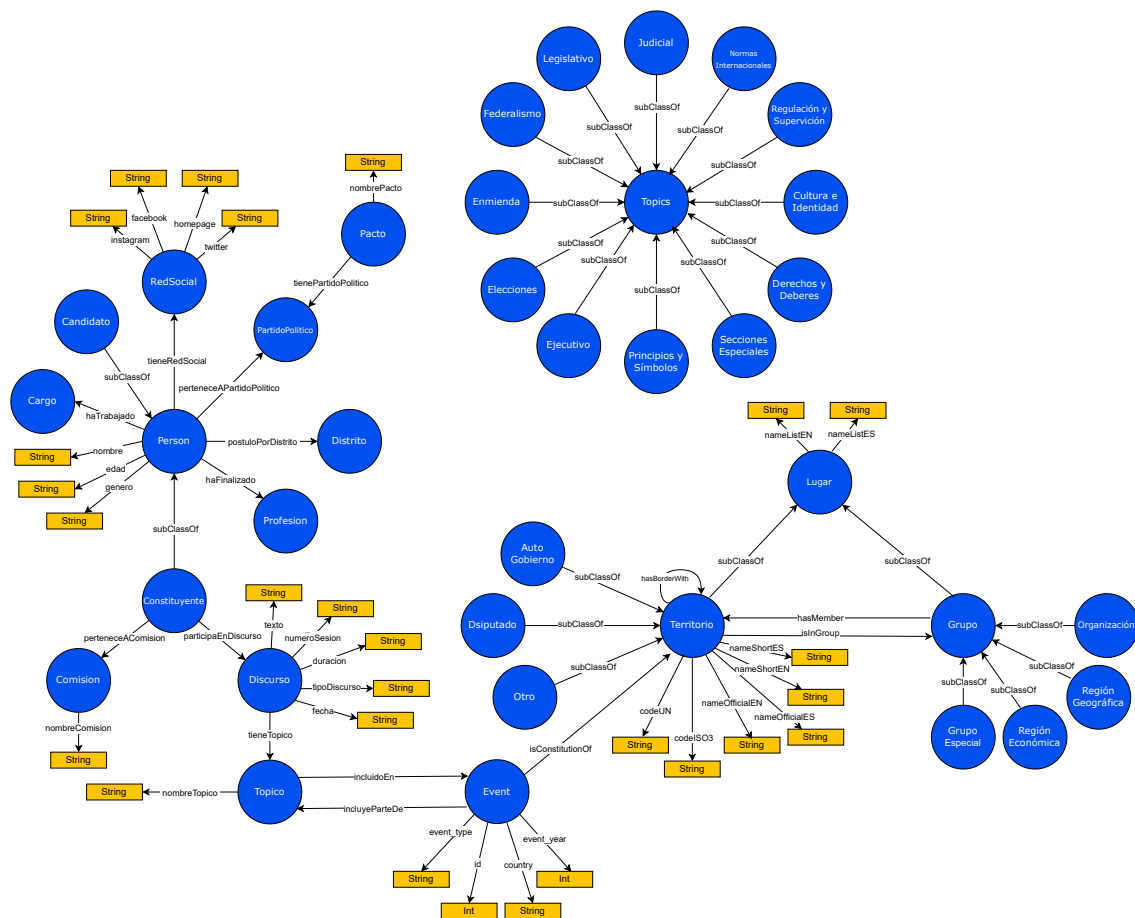


Figura 49: Representación gráfica de la ontología propuesta. Se consideró sólo un nivel de profundidad para los tópicos CCP. Fuente: elaboración propia.

D. Procesamiento de datos

```

Inicio
  definir formato de triples
  definir predicados
  iniciar variables sujeto, objeto y predicado

  abrir archivo de origen en modo lectura
  abrir archivo de destino en modo escritura

  por cada línea en el archivo de origen
    si la línea define la IRI del sujeto, entonces
      asignar IRI a variable sujeto
    si no, si la línea incluye la IRI de una propiedad y su valor, entonces
      asignar valor a variable objeto
      asignar IRI de la propiedad definida a la variable predicado
      escribir en archivo de destino sujeto, predicado y objeto según formato
    fin si entonces
  fin por cada

  cerrar archivo de origen
  cerrar archivo de destino
Fin

```

Figura 50: Pseudocódigo utilizado para la extracción de datos de los archivos *countries.xml* y *chronology.xml*.

Fuente: elaboración propia.

```

Inicio
  definir formato de triples
  iniciar variables sujeto, objeto y predicado
  asignar IRI de la propiedad type a variable predicado

  cargar tabla desde el archivo de origen
  abrir archivo de destino en modo escritura

  por cada fila en la tabla
    asignar IRI del evento a variable sujeto según id de evento
    si la respuesta a la pregunta x1 es "Sí", entonces
      asignar IRI de la clase a variable objeto
      escribir en archivo de destino sujeto, predicado y objeto según formato
    fin si entonces
  fin por cada

  cerrar archivo de destino
Fin

```

Figura 51: Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 1 en el archivo *ccpcnc_v4.csv*.

Fuente: elaboración propia.

```

Inicio
  definir formato de triples
  iniciar variables sujeto, objeto y predicado
  asignar IRI de la propiedad type a variable predicado

  cargar tabla desde el archivo de origen
  abrir archivo de destino en modo escritura

  por cada fila en la tabla
    asignar IRI del evento a variable sujeto según id de evento
    si la respuesta a la pregunta x2 es del tipo afirmativa, entonces
      asignar IRI de la clase a variable objeto
      escribir en archivo de destino sujeto, predicado y objeto según formato
    fin si entonces
  fin por cada

  cerrar archivo de destino
Fin

```

Figura 52: Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 2 en el archivo *ccpcnc_v4.csv*.
Fuente: elaboración propia.

```

Inicio
  definir formato de triples
  iniciar variables sujeto, objeto y predicado
  asignar IRI de la propiedad type a variable predicado

  cargar tabla desde el archivo de origen
  abrir archivo de destino en modo escritura

  por cada fila en la tabla
    asignar IRI del evento a variable sujeto según id de evento
    por cada respuesta en columnas de respuesta para pregunta x3
      si respuesta es afirmativa, entonces
        asignar IRI de la clase a variable objeto
        escribir en archivo de destino sujeto, predicado y objeto según formato
      fin si entonces
    fin por cada
  fin por cada

  cerrar archivo de destino
Fin

```

Figura 53: Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 3 caso 1 en el archivo *ccpcnc_v4.csv*.
Fuente: elaboración propia.

```

Inicio
  definir formato de triples
  iniciar variables sujeto, objeto, predicado y verificacion
  asignar IRI de la propiedad type a variable predicado

  cargar tabla desde el archivo de origen
  abrir archivo de destino en modo escritura

  por cada fila en la tabla
    asignar IRI del evento a variable sujeto según id de evento
    asignar a variable verificación el valor Falso
    por cada respuesta en columnas de respuesta para pregunta x4
      si respuesta es afirmativa, entonces
        asignar a variable verificación el valor Verdadero
      fin si entonces
    fin por cada

    si verificacion tiene el valor Verdadero, entonces
      asignar IRI de la clase a variable objeto
      escribir en archivo de destino sujeto, predicado y objeto según formato
    fin si entonces
  fin por cada

  cerrar archivo de destino
Fin

```

Figura 54: Pseudocódigo utilizado para la extracción de datos de respuestas Tipo 3 caso 2 en el archivo *ccpcnc_v4.csv*.
Fuente: elaboración propia.

```

Inicio
  definir formato de triples
  iniciar variables sujeto, objeto y predicado

  abrir archivo de origen en modo lectura
  abrir archivo de destino en modo escritura

  definir clases CCP y Tópicos BCN de tipo A
  definir clases CCP y Tópicos BCN de tipo B

  por cada declaracion en el archivo de origen
    si el objeto de la declaracion es una clase de tipo A, entonces
      asignar IRI de evento a variable objeto
      asignar IRI de propiedad "IncluidoEn" a variable predicado
      por cada tópico en Tópicos BCN de tipo A para la clase CCP respectiva
        asignar IRI de tópico a variable sujeto
        escribir en archivo de destino sujeto, predicado y objeto según formato
      fin por cada
    fin si entonces

    si el objeto de la declaracion es una clase de tipo B, entonces
      asignar IRI de evento a variable sujeto
      asignar IRI de propiedad "IncluyeParteDe" a variable predicado
      por cada tópico en Tópicos BCN de tipo B para la clase CCP respectiva
        asignar IRI de tópico a variable objeto
        escribir en archivo de destino sujeto, predicado y objeto según formato
      fin por cada
    fin si entonces
  fin por cada

  cerrar archivo de origen
  cerrar archivo de destino
Fin

```

Figura 55: Pseudocódigo utilizado para la creación de declaraciones de integración a partir del archivo *datosDuros.ttl*.

Fuente: elaboración propia.

REFERENCIAS BIBLIOGRÁFICAS

- [Allemang y Hendler, 2008] Allemang, D. y Hendler, J. A. (2008). *Semantic web for the working ontologist: modeling in RDF, RDFS and OWL*. Morgan Kaufmann Publishers/Elsevier, Amsterdam ; Boston. OCLC: ocn184925396.
- [Anderson y Nicholson, 2022] Anderson, B. y Nicholson, B. (2022). SQL vs. NoSQL databases: What's the difference?
- [Arenas et al., 2019] Arenas, M., Gutierrez, C., y Pérez, J. (2019). On the semantics of SPARQL. En De Virgilio, R., Giunchiglia, F., y Tanca, L., editores, *Semantic Web Information Management*, pp. 281–307. Springer Berlin Heidelberg.
- [Bartholomew, 2010] Bartholomew, D. (2010). Sql vs. nosql. *Linux Journal*, 2010(195):4.
- [BCN, 2011] BCN (2011). Acerca de este proyecto — sitio datos abiertos enlazados. Publisher: Biblioteca del Congreso Nacional de Chile.
- [BCN, 2014] BCN (2014). Biblioteca del congreso nacional de chile. Publisher: Biblioteca del Congreso Nacional de Chile.
- [BCN, 2020] BCN (2020). Historia política. Publisher: BCN. Biblioteca del Congreso Nacional de Chile.
- [Beckett et al., 2014] Beckett, D., Berners-Lee, T., Prud'hommeaux, E., y Carothers, G. (2014). RDF 1.1 turtle.
- [Berners-Lee et al., 1998] Berners-Lee, T., Fielding, R., y Masinter, L. (1998). Uniform resource identifiers (URI): Generic syntax. (RFC2396):RFC2396.
- [Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., y Lassila, O. (2001). The semantic web. *Scientific american*, 284(5):34–43.
- [Berners-Lee et al., 1994] Berners-Lee, T., Masinter, L., y McCahill, M. (1994). Uniform resource locators (URL). (RFC1738):RFC1738.
- [Borja y Pérez, 2019] Borja, M. E. y Pérez, M. M. (2019). Big data: un análisis documental de su uso y aplicación en el contexto de la era digital. *Rev. Prop. Inmaterial*, 28:273.
- [Carroll et al., 2004] Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., y Wilkinson, K. (2004). Jena: implementing the semantic web recommendations. En *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, WWW Alt. '04*, pp. 74–83. Association for Computing Machinery.
- [Chen et al., 2014] Chen, Y., Xu, Z., Ni, Y., Cao, G., y Zhang, S. (2014). A RIF Based Mapping of RDB2RDF. *International Journal of Database Theory and Application*, 7:29–44.

- [ChileAtiende, 2022] ChileAtiende (2022). Chileatiende - proceso constituyente.
- [Cyganiak et al., 2014] Cyganiak, R., Wood, D., y Lanthaler, M. (2014). Resource description framework (RDF): Concepts and abstract syntax.
- [DuCharme, 2013] DuCharme, B. (2013). *Learning SPARQL: Querying and Updating with SPARQL 1.1*. .o'Reilly Media, Inc.", 2 edición. Google-Books-ID: s2RgGxd572QC.
- [Dürst y Suignard, 2005] Dürst, M. J. y Suignard, M. (2005). Internationalized resource identifiers (IRIs). (RFC 3987). Num Pages: 46.
- [Elkins y Ginsburg, 2022] Elkins, Z. y Ginsburg, T. (2022). *Characteristics of National Constitutions, Version 4.0*. Comparative constitutions project.
- [EuropeanCommission, 2021] EuropeanCommission (2021). Data europa EU.
- [Fagan, 2010] Fagan, J. C. (2010). Usability studies of faceted browsing: A literature review. 29(2):58.
- [Gandon y Schreiber, 2014] Gandon, F. y Schreiber, G. (2014). RDF 1.1 XML syntax.
- [Gvishiani et al., 2021] Gvishiani, A., Dobrovolsky, M., y Rybkina, A. (2021). *Chapter 6 Big Data and FAIR Data for Data Science*, pp. 105–117. Springer International Publishing, Cham.
- [Haase et al., 2004] Haase, P., Broekstra, J., Eberhart, A., y Volz, R. (2004). A comparison of rdf query languages. pp. 502–517.
- [Hevia, 2020] Hevia, S. (2020). Instituto milenio fundamentos de los datos, IMFD.
- [Inc., 2015] Inc., P. T. (2015). Collaborative data science.
- [León, 2015] León, P. V. (2015). ¿Ontología u Ontologías? *Disputatio. Philosophical Research Bulletin*, 4(5):41.
- [McCarthy, 2005] McCarthy, P. (2005). Archived | search RDF data with SPARQL.
- [Miller et al., 2010] Miller, F. P., Vandome, A. F., y McBrewster, J. (2010). *Apache Maven*. Alpha Press.
- [Moreno-Vega y Hogan, 2018] Moreno-Vega, J. y Hogan, A. (2018). GraFa: Scalable faceted browsing for RDF graphs. En Vrandečić, D., Bontcheva, K., Suárez-Figueroa, M. C., Presutti, V., Celino, I., Sabou, M., Kaffee, L.-A., y Simperl, E., editores, *The Semantic Web – ISWC 2018*, volumen 11136, pp. 301–317. Springer International Publishing. Series Title: Lecture Notes in Computer Science.
- [Musen, 2015] Musen, M. A. (2015). The protégé project: A look back and a look forward. 1(4):4–12.

- [Nations, 2021] Nations, U. (2021). *Macrodatos para el desarrollo sostenible | naciones unidas*. Publisher: United Nations.
- [Oren *et al.*, 2006] Oren, E., Delbru, R., y Decker, S. (2006). Extending faceted navigation for RDF data. En Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., y Aroyo, L. M., editores, *The Semantic Web - ISWC 2006*, volumen 4273, pp. 559–572. Springer Berlin Heidelberg. Series Title: Lecture Notes in Computer Science.
- [Prud'hommeaux y Seaborne, 2008] Prud'hommeaux, E. y Seaborne, A. (2008). SPARQL query language for RDF.
- [PUC, 2021] PUC (2021). UC - IMFD presenta plataforma telar: investigación basada en datos al servicio de la ciudadanía.
- [Sahatqija *et al.*, 2018] Sahatqija, K., Ajdari, J., Zenuni, X., Raufi, B., e Ismaili, F. (2018). Comparison between relational and nosql databases. En *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MI-PRO)*, pp. 0216–0221.
- [Schewe y Thalheim, 2011] Schewe, K.-D. y Thalheim, B., editores (2011). *Semantics in Data and Knowledge Bases*, volumen 6834 de *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Sequeda y Miranker, 2015] Sequeda, J. y Miranker, D. P. (2015). Ultrawrap mapper: A semi-automatic relational database to rdf (rdb2rdf) mapping tool. En *SEMWEB*.
- [Sequeda *et al.*, 2012a] Sequeda, J., Priyatna, F., y Villazón-Terrazas, B. (2012a). Relational database to rdf mapping patterns. En *WOP*.
- [Sequeda *et al.*, 2012b] Sequeda, J. F., Arenas, M., y Miranker, D. P. (2012b). On directly mapping relational databases to RDF and OWL. *Proceedings of the 21st international conference on World Wide Web - WWW '12*.
- [Simonini y Zhu, 2015] Simonini, G. y Zhu, S. (2015). Big data exploration with faceted browsing. En *2015 International Conference on High Performance Computing & Simulation (HPCS)*, pp. 541–544. IEEE.
- [Stadler *et al.*, 2015] Stadler, C., Unbehauen, J., Westphal, P., Sherif, M., y Lehmann, J. (2015). Simplified rdb2rdf mapping. En *Proceedings of the 8th Workshop on Linked Data on the Web*, volumen 1409.
- [Tascón, 2013] Tascón, M. (2013). Introducción: Big data. pasado, presente y futuro. *Telos: Cuadernos de comunicación e innovación*, (95):47–50.
- [Thapa y Giese, 2021] Thapa, R. B. y Giese, M. (2021). A Source-to-Target Constraint Rewriting for Direct Mapping. En Hotho, A., Blomqvist, E., Dietze, S., Fokoue, A., Ding, Y., Barnaghi, P., Haller, A., Dragoni, M., y Alani, H., editores, *The Semantic Web – ISWC 2021*, Lecture Notes in Computer Science, pp. 21–38, Cham. Springer International Publishing.

- [Thuy *et al.*, 2014] Thuy, P. T. T., Thuan, N. D., Han, Y., Park, K., y Lee, Y.-K. (2014). RDB2RDF: completed transformation from relational database into RDF ontology. En *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication - ICUIMC '14*, pp. 1-7, Siem Reap, Cambodia. ACM Press.
- [van der Ham *et al.*, 2006] van der Ham, J. J., Dijkstra, F., Travostino, F., Andree, H. M., y de Laat, C. T. (2006). Using RDF to describe networks. *Future Generation Computer Systems*, 22(8):862-867.
- [Van Rossum y Drake, 2009] Van Rossum, G. y Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- [W3C, 1994] W3C (1994). Semantic Web Standards.
- [W3C, 2012] W3C (2012). A Direct Mapping of Relational Data to RDF.
- [Wardani y Kiing, 2014] Wardani, D. W. y Kiing, J. (2014). Semantic mapping relational to graph model. En *2014 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, pp. 160-165.
- [Zou, 2018] Zou, L. (2018). SPARQL. En Liu, L. y Özsu, M. T., editores, *Encyclopedia of Database Systems*, pp. 3554-3558. Springer New York.